

A construção de um modelo de avaliação por profissionais e gestores de saúde: validação e consolidação de dimensões e indicadores de performance em serviços e unidades de saúde do SUS

Galba Freire Moita

Universidade de Coimbra – CeBER, Coimbra, Portugal.

Vítor Manuel dos Reis Raposo

Universidade de Coimbra, CEISUC - Coimbra, Portugal.

Allan Claudius Queiroz Barbosa

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte – MG, Brasil

Após 30 anos de SUS, ainda não se registra nenhum sistema nacional institucionalizado e contínuo de avaliação de resultados e desempenho em saúde pública, apesar do Estado brasileiro ter implementado ações estruturais no bojo da Nova Gestão Pública ou Gerencialismo, que exigem dos gestores esforços para monitorar resultados, controle e prestação de contas (*accountability*). Objetivo: estudaram-se alguns relevantes quadros teóricos (*framework*) e propostas de dimensões e indicadores de *performance* em serviços e unidades de saúde, visando construir um instrumento inovador a partir do conhecimento prévio de uma amostra de especialistas e gestores no contexto do SUS. Métodos: através de painéis colaborativos com 220 gestores e profissionais de saúde (em fases sucessivas – painel de especialistas, grupos *Delphi* e pesquisa de campo), captou-se valorações de itens de avaliação no contexto do SUS. Os resultados obtidos de 108 especialistas e 112 gestores de saúde foram objetos de análise descritiva e inferencial. Resultados: os itens de avaliação de *performance* que surgiram e tiveram valorações médias acima do nível 3 (relevante), classificados como subcategorias (20 e 24 domínios de avaliação) e como categorias (cinco e seis macrodomínios de avaliação), foram usados para análise inferencial, com rigor metodológico. As análises de confiabilidade e validade resultaram em um modelo final de 17 domínios (indicadores) e quatro macrodomínios de avaliação que validaram um *proxy* de painel de indicadores e dimensões para unidades do SUS.

Palavras-chave: avaliação de processos e resultados, avaliação em saúde, indicadores básicos de saúde

DOI: <https://doi.org/10.21874/rsp.v7i13.3955> | E- ISSN: 2357-8017

[Artigo recebido em 11 de julho de 2019 e aceito em 09 de março de 2020.]



La construcción de un modelo de evaluación por parte de profesionales y gestores de salud: validación y consolidación de dimensiones e indicadores de desempeño en los servicios y unidades de salud del SUS

Después de 30 años del SUS aún no se registra ningún sistema nacional institucionalizado y continuo de evaluación de resultados y desempeño en salud pública, a pesar del Estado brasileño haber implementado acciones estructurales en el seno de la Nueva Gestión Pública o Gerencialismo, que exigen de los gestores esfuerzos para monitorear los resultados, el control y la rendición de cuentas (*accountability*). Objetivo: se estudiaron algunos relevantes cuadros teóricos (*framework*) y propuestas de dimensiones e indicadores de desempeño en servicios y unidades de salud, buscando construir un instrumento innovador a partir del conocimiento previo de una muestra de especialistas y gestores en el contexto del SUS. Métodos: a través de paneles colaborativos con 220 gestores y profesionales de salud (en fases sucesivas - panel de especialistas, grupos Delphi e investigación de campo) se captó valoraciones de ítems de evaluación en el contexto del SUS. Los resultados obtenidos de 108 especialistas y 112 gestores de salud fueron objetos de análisis descriptivo e inferencial. Resultados : los ítems de evaluación de desempeño que surgieron y tuvieron valoraciones medias por encima del nivel 3 (relevantes), clasificados como subcategorías (20 y 24 dominios de evaluación) y como categorías (cinco y seis macro dominios de evaluación), fueron utilizados para análisis inferencial con rigor metodológico. Los análisis de confiabilidad y validez resultaron en un modelo final de 17 dominios (indicadores) y cuatro macro dominios de evaluación que validaron un proxy de panel de indicadores y dimensiones para unidades del SUS.

Palabras clave: evaluación de procesos y resultados, evaluación de salud, indicadores básicos de salud

The construction of an evaluation model by health professionals and managers: validation and consolidation of dimensions and performance indicators in SUS health services and units

After 30 years of SUS, there is still no institutionalized and continuous national system of evaluation of results and performance in public health, although the Brazilian State has implemented structural actions in the area of New Public Management or Managerialism, which require managers to monitor results, control and accountability. Objective: We studied some relevant theoretical frameworks and proposals of dimensions and performance indicators in health services and units, aiming at constructing an innovative instrument based on the prior knowledge of a sample of specialists and managers in the SUS context. Methods: Collaborative panels with 220 managers and health professionals (in successive phases - panel of experts, Delphi groups and field research) were collected valuations of evaluated items in the SUS context. The results obtained from 108 specialists and 112 health managers were the object of descriptive and inferential analysis. Results: The performance evaluation items that emerged and had mean scores above level 3 (relevant), classified as subcategories (20 and 24 evaluation domains) and as categories (five and six evaluation domains), were used for inferential analysis, with methodological rigor. Reliability and validity analyses resulted in a final model of 17 domains (indicators) and four evaluation domains that validated a panel proxy for indicators and dimensions for SUS units.

Keywords: process and results evaluation, health assessment, basic health indicators

1 Introdução

A criação do Sistema Único de Saúde (SUS) em 1988, no Brasil, representou avanços do ponto de vista de garantias de atendimento, da organização sistêmica e da descentralização da gestão única; porém, com frágil governança de desempenho. Por sua vez, a Nova Gestão Pública (NGP), ou Gerencialismo, tem exigido dos gestores públicos de saúde esforços de monitoramento de resultados, controladoria e responsabilização (*accountability*). No caso da saúde, o elevado volume de recursos aplicados, por si, exige avanços no monitoramento de resultados. Dados relativos a 2016 apontam que 35 países da OCDE gastam em torno de 9,0% do PIB no sistema de saúde, nos Estados Unidos cerca de 17,2%, em Portugal 8,9% e no Brasil 6,2% (OECD, 2017). Além disso, os gastos com a saúde têm crescido nos últimos anos (2003-2016) em média 3,6% nesses países (OECD, 2017). Nesta perspectiva, este estudo optou por uma abordagem integrada do monitoramento e avaliação, que aborda o processo de análises aprofundadas de dados e situações, pertinentes para a tomada de decisão e gestão baseada em resultados (UNDP, 2009). Assume-se, ainda, o escopo instrumental do monitoramento de resultados definido no âmbito da OECD, que trata o monitoramento como a função contínua que usa a recolha sistemática de dados, para fornecer à gestão e aos decisores (*stakeholders*) parâmetros decisores sobre uma intervenção em curso, por meio de indicadores específicos sobre a extensão do progresso, do atingimento dos objetivos e dos progressos alcançados na utilização de recursos alocados (HURST, 2000; MORRA IMAS; RIST, 2009; OECD, 2001).

Em busca de desenvolver um quadro teórico-conceitual de avaliação de *performance* para organizações de saúde, aplicável às unidades do SUS, este estudo analisou algumas metodologias de mensuração de resultados em saúde no cenário mundial, além de uma pesquisa integrativa dos principais modelos aplicados no Brasil, especialmente quanto aos quadros de dimensões e indicadores prevalentes na literatura.

Diante desse contexto, surge a problematização central desta investigação: em um ambiente de escassez de metodologias e instrumentos validados de mensuração de performance e resultados em saúde, qual a importância relativa entre os itens de avaliação mais relevantes e seus possíveis fatores mediadores na validação e consolidação de um *proxy* de painel de indicadores multidimensionais aplicados em saúde pública no Brasil?

Assim, surge o objetivo principal de, a partir de alguns modelos relevantes na literatura mundial, validar e consolidar um grupo de indicadores multidimensionais distribuídos em macrodimensões de avaliação (qualitativos, quantitativos, de efeitos, satisfação e sistêmico/estratégico) por uma amostra de profissionais e gestores de saúde pública, através de um processo avaliativo e análises de confiabilidade e validade (por meio de Correlação e Análise Fatorial Exploratória) que permitam futuramente a estruturação de instrumento aplicado à gestão de performance de unidades do SUS.

No âmbito do SUS, justifica-se a importância deste estudo avaliativo, pois a avaliação é vista não só como instrumento para verificar a efetividade de programas, mas utilizada também com o intuito de melhorar a qualidade dos serviços por um processo de julgamento de valor orientado para a tomada de decisão e governança (RAPOSO, 2007).

A proposta fundamenta-se no campo da translação de conhecimentos, no qual as teorias consolidadas são alvos de transliteração à luz de experiências dos atores sociais envolvidos (BARBOSA; NETO, 2017 apud CLAVIER *et al.*, 2011; HARTZ *et al.*, 2008). Assim, pautou-se em um processo colaborativo visando sintetizar e aplicar conhecimentos com vista à melhoria de serviços em prol de preenchimentos de lacunas de saberes e instrumentos para transformar políticas e práticas (BARBOSA; NETO, 2017 apud CIHR, 2004, p.2; WHO, 2006, p.1) para a gestão de serviços do SUS, que possam ser aplicados nos serviços públicos de saúde.

Dessa forma, abordou-se o campo da avaliação e o modelo lógico, estabeleceu-se uma metodologia, procedeu-se à análise e à validação de resultados e, por fim, propôs-se as considerações finais.

2. Avaliação em saúde e o modelo teórico-lógico

Os primórdios da avaliação em saúde são citados em (DOS REIS; DOS SANTOS; DE CAMPOS; ACÚRCIO *et al.*, 1990), nos trabalhos de Flexner (FLEXNER, 1910) e no relatório Codman (PORTERFIELD, 1976), com uma proposta de metodologia de avaliação de rotina do estado de saúde dos pacientes hospitalares, para estabelecer os resultados finais das intervenções médicas.

A partir da proposta inicial de Avendis Donabedian (DONABEDIAN, 1980) – Avaliação de Estrutura, Processos e Resultados –, há uma busca por modelos integrativos, em que se avaliam as relações entre estado de saúde, qualidade do cuidado e gastos de recursos. Também se pode citar o modelo de Brook & Lohr (1985) e de (DOS REIS; DOS SANTOS; DE CAMPOS; ACÚRCIO *et al.*, 1990), que propõe a avaliação das dimensões: eficácia do cuidado, efetividade do cuidado, as variações das características populacionais e os níveis de qualidade de cuidado.

Alkin (2004 APUD SAMICO; FELISBEERTO; FIGUEIRÓ; FRIAS, 2010; SAMICO; FELISBERTO; FIGUEIRÓ; FRIAS, 2010) estabelece as origens do campo da avaliação a partir de duas necessidades: a prestação de contas (*accountability*) e o controle dos programas.

Por sua vez, Chianca e Youker (2004) registram um crescente interesse no campo de avaliação de sistemas de saúde da América Latina e Caribe. No entanto, Bosi e Mercado (2006) elaboram uma extensa revisão de literatura quanto às experiências de avaliação nos países da América Latina e concluem que a maioria dos trabalhos é de índole teórica, sendo poucos os estudos empíricos ou de caso. Citam ainda que quase sempre os enfoques avaliativos usados se inspiram em certos autores e obras eurocêntricas (FALS BORDA; MORA-OSEJO, 2004), do mundo anglo-saxão, em especial, provenientes das teorias e modelos emergentes da península ibérica (BOSI; MERCADO, 2006). Assim, o campo de avaliação em saúde, por suas complexidades, subjetividades e múltiplas perspectivas, engloba uma diversidade de termos, conceitos e métodos, o que está coerente com a multiplicidade de questões consideradas como pertinentes na área da saúde, na heterogeneidade e complexidade das intervenções, sejam elas ações, serviços, programas ou políticas públicas.

Neste estudo, escolheu-se uma abordagem avaliativa pautada na teoria dos *stakeholders* organizacionais por intermédio do envolvimento de especialistas e gestores para a captação de informação relevante para a previsão do comportamento organizacional, na perspectiva de um grupo de decisores-chave do SUS.

Segundo Raposo (2007), a avaliação é vista não só como instrumento para verificar a efetividade de programas, mas utilizada também com o intuito de melhorar a qualidade dos serviços. Sendo assim, a avaliação pode ser entendida como um processo de julgamento de valor orientado para a tomada de decisão e governança.

Em outro prisma, Hartz e Vieira-da-Silva (2005) abordam a avaliação enquanto desenhos de modelos teóricos-lógicos para a compreensão das relações de lógica entre os objetivos, atividades e resultados do programa. Assim, o objeto da investigação científica é um constructo da realidade passível de modelagem, instrumentalizado em dois momentos: a idealização do objeto-modelo ou modelo conceitual e a construção da teoria do objeto-modelo ou, em outros termos, a elaboração do modelo teórico (BUNGE, 1974). O modelo lógico (ML) é entendido enquanto um esquema visual que apresenta como um programa deve ser implementado e que resultados são esperados (ROWAN, 2000).

Em uma breve análise dos aspectos filosóficos da avaliação, escolheu-se trabalhar sob as bases da ontologia idealista, na qual os modelos e representações que formam o nível empírico são contingentes e socialmente produzidos - neste caso, pelos agentes de decisão chaves (BOSI; MERCADO, 2006).

Enquanto isso, nos aspectos metodológicos, assume-se nesta investigação um processo integrativo e participativo da avaliação, em busca de construir consensos quanto a visão global, metodologias, macrodimensões e indicadores de mensuração de performance organizacional, em especial no âmbito de agentes de decisão chaves de unidades de saúde pública e/ou vinculadas ao SUS.

Este trabalho aborda aspectos de aplicabilidade da avaliação para a tomada de decisão em saúde no Brasil, como nos exemplos dos relatos de Tanaka e Tamaki (2012), de Contandriopoulos (2006), de Fernandes, Ribeiro e Moreira (2011), de Campos e Furtado (2011) e alguns outros estudos iniciais aplicados no sistema de saúde do Brasil. De acordo com Travassos, Noronha e Martins. (1999), é necessário destacar a importância do desenvolvimento de estratégias para permitir que os resultados das avaliações melhorem a efetividade dos cuidados hospitalares. Assim, o monitoramento de indicadores pode promover a efetividade da assistência, bem como da busca pela eficiência organizacional em saúde.

A monitorização dos indicadores constitui também uma boa prática de governança hospitalar (RAPOSO, 2007), uma vez que reforçam a transparência e a prestação de contas/responsabilização (*accountability*). Entretanto, a multiplicidade

de perspectivas e de complexidades envolvidas, além das diversidades de escopos e definições, demanda a análise dos principais quadros teóricos e metodológicos de avaliação relevantes da literatura.

Na literatura há diversas propostas de quadro teórico-conceitual para avaliação de performance e a seleção de indicadores, por exemplo, painéis de dimensões de avaliação e indicadores de performance usados para a medição da qualidade de alguns países e organizações multilaterais (Reino Unido, Canadá, Austrália, EUA, OMS e OCDE) citados como relevantes em outro importante estudo (ARAH; KLAZINGA; DELNOIJ; ASBROEK *et al.*, 2003). Outro destaque é o quadro teórico-conceitual proposto pela Organização Mundial de Saúde (OMS), chamado projeto PATH, que está baseado em competências profissionais na aplicação do conhecimento atual, tecnologias e recursos disponíveis; eficiência no uso de recursos; risco mínimo para o paciente; responsividade ao paciente; contribuição ótima para os resultados de saúde (VEILLARD; CHAMPAGNE; KLAZINGA; KAZANDJIAN *et al.*, 2005).

Por sua vez, Papanicolas & Smith (2010) comparam dez abordagens destacadas na literatura de quadro teórico-conceitual (*framework*) de sistemas de saúde e as classificam segundo estas três categorias, além de expor as limitações de cada um dos dez modelos estudados, sendo nove deles assinalados também como quadro teórico-conceitual (*framework*) de avaliação de desempenho de sistemas de saúde, exceto o modelo *behavioural health care* (Quadro 1).

Quadro 1 | Quadro de objetivos versus quadro teórico-conceitual (framework)

Framework (nome)	Objetivo Intermediário	Objetivo Final
<i>Framework for Assessing Behavioural Healthcare</i> (1998)	Eficácia; Eficiência; Equidade	Saúde e bem-estar
EGIPSS model (1998)	Produtividade; Volume de cuidados e serviços; Qualidade de cuidados e serviços	Melhoria da saúde; Eficácia; Eficiência; Equidade
WHO Performance Framework (2000)	Acesso ; Cobertura; Qualidade; Segurança	Nível e distribuição de saúde; Nível e distribuição de capacidade de resposta; Justiça no financiamento; Eficiência
OECD Performance Framework (2001)		- Nível e distribuição de saúde; Nível e distribuição de capacidade de resposta e acesso; Equidade; Eficiência macroeconômica e microeconômica
<i>Control Knobs Framework</i> (2003)	Eficiência; Qualidade; Acesso	Estado de saúde; Satisfação do usuário; Proteção contra riscos
<i>Commonwealth Fund Framework</i> (2006)	Alta qualidade do cuidado; Cuidado eficiente; Acesso; Inovação e melhoria do sistema e da força de trabalho	Vidas longas, saudáveis e produtivas
OECD HCQI Framework (2006)		Melhoria da saúde; Eficiência macroeconômica (sustentabilidade) e eficiência microeconômica (valor para o dinheiro); Equidade
<i>WHO Building Blocks Framework</i> (2007)	Acesso; Cobertura; Qualidade; Segurança	Nível e distribuição de saúde; Nível e distribuição de capacidade de resposta; Justiça no financiamento; Eficiência
<i>Systems Framework</i> (2008)	Equidade; Escolha; Eficiência; Eficácia	Saúde; Proteção contra risco financeiro; Satisfação do usuário
<i>IHP Common Evaluation Framework</i> (2008)	N/A	N/A

Fonte: adaptado de Papanicolas e Smith (2010: 33-35)

De forma mais instrumental, pode-se citar ainda alguns quadros teóricos-conceituais como os projetos *Quality Indicator Project* (QIP)¹ da Agência de Pesquisa e Qualidade nos Serviços de Saúde – *U.S. Agency for Healthcare Research and Quality* (AHRQ)², com destaque para o projeto da Fundação Europeia para Gerenciamento da Qualidade – *European Foundation for Quality Management* (EFQM)³, além de sistema de indicadores do *Project Assessment Tool of Quality Improvement in Hospitals* (PATH)⁴ da Organização Mundial de Saúde (OMS)⁵. No Reino Unido, são duas as abordagens, a primeira baseada na definição de indicadores do Serviço Nacional de Saúde/NHS⁶, e a segunda interessada na concretização de uma metodologia para estabelecer *hospital ratings*⁷; em Portugal, o *Sinas@Hospitais*⁷ apresenta-se sob a coordenação da Entidade Reguladora de Saúde (ERS)⁸, que classifica os hospitais avaliados (*ranking*) e é utilizada pela Administração Central do Sistema de Saúde (ACSS)⁹, entre outros.

Além disso, estudaram-se os projetos brasileiros como o Projeto Desenvolvimento de Metodologia de Avaliação do Desempenho do Sistema de Saúde Brasileiro (Proadess)¹⁰, indicadores do Índice de Desempenho do SUS (IDSUS)¹¹, Indicadores e Dados Básicos da Saúde (IDB)¹² da Rede Interagencial de Informações para a Saúde (RIPSA) e indicadores do novo Programa Nacional de Avaliação de Serviços de Saúde (PNASS). No âmbito da Agência Nacional de Saúde Suplementar (ANS), foram estudados Índices de Desempenho da Saúde Suplementar (IDSS) e indicadores Qualiss do Programa de Qualificação dos Prestadores de Serviços de Saúde Suplementar

¹ QIP – *Quality Indicator Project* (Projeto de Indicadores de Qualidade).

² AHRQ – *U.S. Agency for Healthcare Research and Quality* (Agência de Pesquisa e Qualidade nos Serviços de Saúde).

³ EFQM – *European Foundation for Quality Management* (Fundação Europeia para Gerenciamento da Qualidade).

⁴ PATH – *Performance Assessment Tool for Quality Improvement in Hospitals* (ferramenta de avaliação de performance para a melhoria da qualidade hospitalar).

⁵ OMS – Organização Mundial de Saúde (WHO), representada no Brasil, América Latina, e Caribe pela OPAS-Organização Pan-Americana de Saúde.

⁶ NHS – *National Health System* (Serviço Nacional de Saúde do Reino Unido).

⁷ *Sinas@Hospitais* – Sistema de Indicadores de Performance de Hospitais do Serviço Nacional de Saúde de Portugal.

⁸ ERS – Entidade Reguladora de Saúde que operacionaliza a regulação do sistema público e privado do SNS de Portugal.

⁹ ACSS – Administração Central do Sistema Nacional de Saúde (SNS) de Portugal.

¹⁰ PROADESS – Projeto Desenvolvimento de Metodologia de Avaliação do Desempenho do Sistema de Saúde Brasileiro.

¹¹ IDSUS – Matriz de indicadores em 18 macrodimensões gerenciada pelo DATASUS e o Ministério da Saúde do Brasil.

¹² IDB – Indicadores e Dados Básicos para a Saúde no Brasil gerenciados pela RIPSA e o Ministério da Saúde do Brasil.

(Qualiss), em fase de estruturação; nas organizações privadas: indicadores Prohasa/CQH¹³ e indicadores Sipageh¹⁴.

No contexto brasileiro, uma revisão sistemática, voltada para avaliação de qualidade hospitalar, mostra que os primeiros estudos surgiram entre 1991 e 2003, tendo encontrado apenas 48 publicações (MACHADO; MARTINS; MARTINS, 2013). Em outro estudo sobre a avaliação em saúde no Brasil, acharam-se apenas 44 estudos (FERNANDES; RIBEIRO; MOREIRA, 2011).

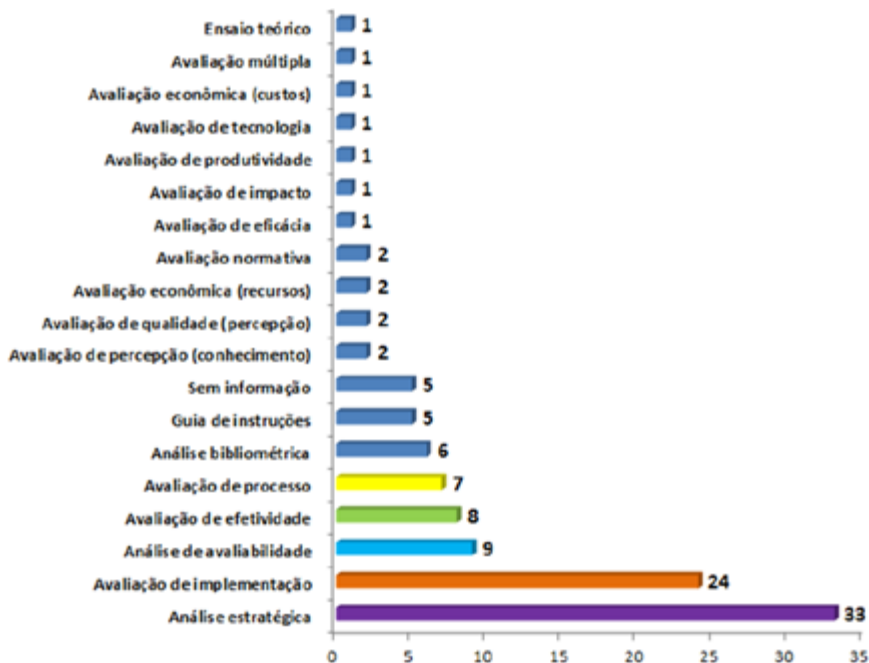
Neste propósito, procedeu-se a uma revisão sistemática ampliada nas principais bases publicações de artigos científicos do Brasil (Scielo e Lilacs e Medline, via Biblioteca Virtual em Saúde - BVS), tendo localizado 112 publicações no período de 1990 a 2018, sendo apenas 12 produções na linha de avaliação de performance em saúde e somente dois artigos no campo da avaliação de qualidade percebida caracterizados, comprovando-se a incipiência de estudos e/ou publicações sobre o tema. Dessa forma, na Figura 1, sintetizou-se a seguinte classificação das publicações que evidenciam a importância desta investigação, conforme dados da pesquisa original (MOITA, 2019).

Destaca-se, ainda, que o período de 1999 a 2005 teve média de apenas uma publicação anual, entre 2006 e 2010 houve 39 produções, e entre 2011 e 2018 a média foi de quatro publicações anuais, sendo que após 2011 foram publicados 18 artigos, cinco monografias e oito teses, que mostram o nível de interesse pela avaliação em saúde.

¹³ PROHASA/CQH – Indicadores do Programa de Estudos Avançados em Adm. Hospitalar e de Sistemas de Saúde da FMUSP.

¹⁴ SIPAGEH – Sistema de Indicadores Padronizados para Gestão Hospitalar, com adesão voluntária de 37 hospitais do Sul do Brasil.

Figura 1 | Bibliometria de estudos de avaliação em saúde no Brasil (1990 – 2018)



Fonte: elaboração própria

Em outra perspectiva, na saúde brasileira existem alguns sistemas informatizados, indicadores e instrumentos gerenciais no campo da saúde pública, tais como: indicadores do Datusus, Projeto Proadess, indicadores IDSUS, indicadores RIPSA e indicadores do novo PNASS. No âmbito da ANS, observam-se os Índices IDSS e indicadores Qualiss; e nas organizações privadas, constam os indicadores Prohasa/CQH e indicadores Sipageh. Infelizmente, nenhum desses quadros de indicadores tem sido usado de forma geral e permanente no SUS, possivelmente pela falta de legitimidade no uso.

Um estudo aplicado da (ROSALEM, 2013) apresentou um quadro integrativo do rol de indicadores das metodologias do CQH, Prohasa, Sipageh e de três outras listas de indicadores de alguns órgãos e agências internacionais (ACHS – Australian Council on Healthcare Standards, CND – Canadá Health Indicators – 2003, HP – Health People 2010 – USA, NHS - National Health Security, RALC - Grupo para Reforma del Sector Salud en América Latina y Caribe – OPAS) (Quadro 4, Anexo 1).

A inclusão dos diversos atores na decisão da modelagem da avaliação se faz necessária também pela variedade de dimensões mensuráveis nos serviços de saúde, e pela necessidade de priorização por seus futuros utilizadores. Neste quesito, Uchimura e Bosi (UCHIMURA; BOSI, 2002) referem-se a autores e listas das possíveis dimensões de avaliação. Gattinara e outros (GATTINARA; IBACACHE; PUENTE; GIACONI *et al.*, 1995) sinalizam vários fatores que determinam a qualidade dos serviços de saúde: competência profissional, satisfação dos usuários-utilizadores, acessibilidade, eficácia, eficiência. Por outro lado, Vuori (VUORI, 1991) aponta efetividade, eficácia, eficiência etc., também com modelos referidos por Acurcio (ACURCIO; CHERCHIGLIA; SANTOS, 1991), por Akerman e Nadanovsky (AKERMAN; NADANOVSKY, 1992) e por Santos (SANTOS, 1995) .

Por sua vez, no enquadramento dos estudos de governança organizacional, deve-se selecionar uma teoria de embasamento. Entre as diversas teorias de governança, pode-se destacar a teoria neoclássica, a teoria da agência, a teoria dos custos de transações, a teoria do direito à propriedade, dentre outras, sintetizadas no estudo de Alves (ALVES, 2012).

Há outras teorias que se destacam no âmbito da gestão pública, que não se distanciam desta investigação, por exemplo: a teoria da escolha pública (*public choice*) e a teoria da provedoria (*stewardship*). Na síntese apresentada por Alves (ALVES, 2012) o autor destaca a teoria dos *stakeholders* com diversas vantagens para a gestão dos sistemas produtivos de bens e serviços públicos que se alinham com os objetivos deste estudo.

Ressalte-se que pesquisadores da teoria dos *stakeholders* sustentam que o conhecimento acerca da natureza dos *stakeholders* organizacionais constitui informação relevante para a previsão do comportamento organizacional (DÉNIZ-DÉNIZ; ZÁRRAGA-OBERTY, 2004; MELLAHI; WOOD, 2003). Segundo Freeman (FREEMAN, 1984), um *stakeholder* é qualquer indivíduo ou grupo que pode afetar ou ser afetado pelo alcance dos objetivos de uma organização.

Do ponto de vista teórico, há pelo menos quatro principais perspectivas com diferentes abordagens da ação dos *stakeholders* nas organizações: as visões instrumental, descritiva, normativa e de responsabilidade social (CARVALHO, 2007).

Diante dos objetivos desta investigação, tomou-se por base na teoria dos *stakeholders*, em especial nas suas perspectivas *descritiva e instrumental* destacadas pela literatura (BRYSON, 2004; HART; SHARMA, 2004), em que os grupos de *stakeholders* podem efetivamente exercer um impacto considerável nos resultados organizacionais e sobre o seu desempenho e competitividade.

Assim, a validação do modelo proposto neste trabalho, a partir da identificação dos *stakeholders* e do mapeamento de seus propósitos, concede uma maior adequação ao uso do modelo com base na interação com o ponto de vista do analista e de uma revisão da literatura, na proposição de modelos aceitos pela comunidade científica, mas com adequação à realidade e utilidade percebida pelos decisores-clientes.

Por sua vez, metodologias e sistemas de indicadores de destaque no campo da saúde pública podem ser analisados e sistematizados de acordo com suas possíveis potencialidades, virtudes, limitações e desafios, para compor uma matriz de dimensões e de indicadores para avaliação e validação pelos agentes de decisão da amostra.

3. Método

A natureza da pesquisa caracterizou-se por uma metodologia avaliativa do tipo **observacional** (controle de possíveis influências intencionais no objeto pesquisado), **transversal**, com abordagem por método misto (qualitativo e quantitativo) de avaliação durante a etapa de validação metodológica e instrumental por agentes-chave da amostra (especialistas, profissionais e gestores da saúde).

Carvalho (2007) afirma que no processo de avaliação importa não só identificar os atores principais (*stakeholders*), mas, também, caracterizá-los e agrupá-los em função da sua importância considerada e do impacto que poderão ter no nível do comportamento e dos resultados alcançados pela organização.

Para o escopo deste trabalho, adotou-se a perspectiva da teoria dos *stakeholders* e considerou-se como decisores-chave (*key stakeholders*) os gestores dos níveis setoriais, organizacionais e governacionais, no processo de estruturação de modelos para

mensuração da performance organizacional. Diante disso, esta investigação firma-se no enquadramento conceitual do campo da avaliação participativa e construtivista de 4ª geração proposta por Guba e Lincoln (2011), pautado em um processo construtivista de negociação entre interessados, que se apresenta como inclusiva e participativa, em busca de promover a participação pluralista dos atores sociais e escutar as vozes múltiplas das partes interessadas (*stakeholders*) na avaliação. Assim, envolveu gestores e profissionais para a captação e validação da percepção desses agentes-chave (*stakeholders*) quanto às dimensões e indicadores-chave para a mensuração de desempenho em organizações e serviços de saúde no contexto do SUS-Brasil.

Sendo assim, nos aspectos metodológicos, assume-se nesta investigação um processo integrativo e participativo da avaliação, em busca de construir consensos quanto a visão global, metodologias, macrodimensões e indicadores de mensuração de performance organizacional, em especial no âmbito de agentes de decisão chaves de unidades de saúde pública e/ou vinculadas ao SUS.

Na exploração inicial dos conhecimentos prévios dos agentes de decisão, seguiu-se os princípios de Sampieri e outros (2006), que sugerem a amostragem por conveniência quando a investigação está centrada em indivíduos ou grupos portadores de um ou vários atributos que possam contribuir para o desenvolvimento de uma teoria.

Dessa forma, nesta etapa partiu-se da pesquisa integrativa realizada (MOITA, 2019; MOITA; RAPOSO; BARBOSA, 2019a) para estudar os modelos de avaliação de performance mais relevantes, mas não se selecionou nenhum modelo específico de avaliação de resultados, evitando confinamentos dos conhecimentos captados.

Assim, o marco teórico norteou os pesquisadores, mas, nesta etapa, buscou-se mapear os conhecimentos prévios dos respondentes, sem fazer vínculos com a literatura (Quadro 4 e Anexo 1), que foram diretrizes para aprofundamentos em etapas posteriores.

Estabeleceu-se um modelo lógico para orientar a proposta inicial de coleta de dados qualitativos para a estruturação da visão global, a partir do esquema proposto por Hartz e Vieira-da-Silva (SOUZA; VIEIRA-DA-SILVA; HARTZ, 2005) e de macrodimensões de avaliação de desempenho prevalentes na literatura. A proposta, do ponto de vista do marco teórico, foi pautada apenas na análise da Figura 2 pelos respondentes.

Figura 2 | Diagrama do modelo lógico – atributos ou características para avaliação em saúde



Fonte: Elaborado a partir do esquema proposto por Hartz e Vieira-da-Silva (2005).

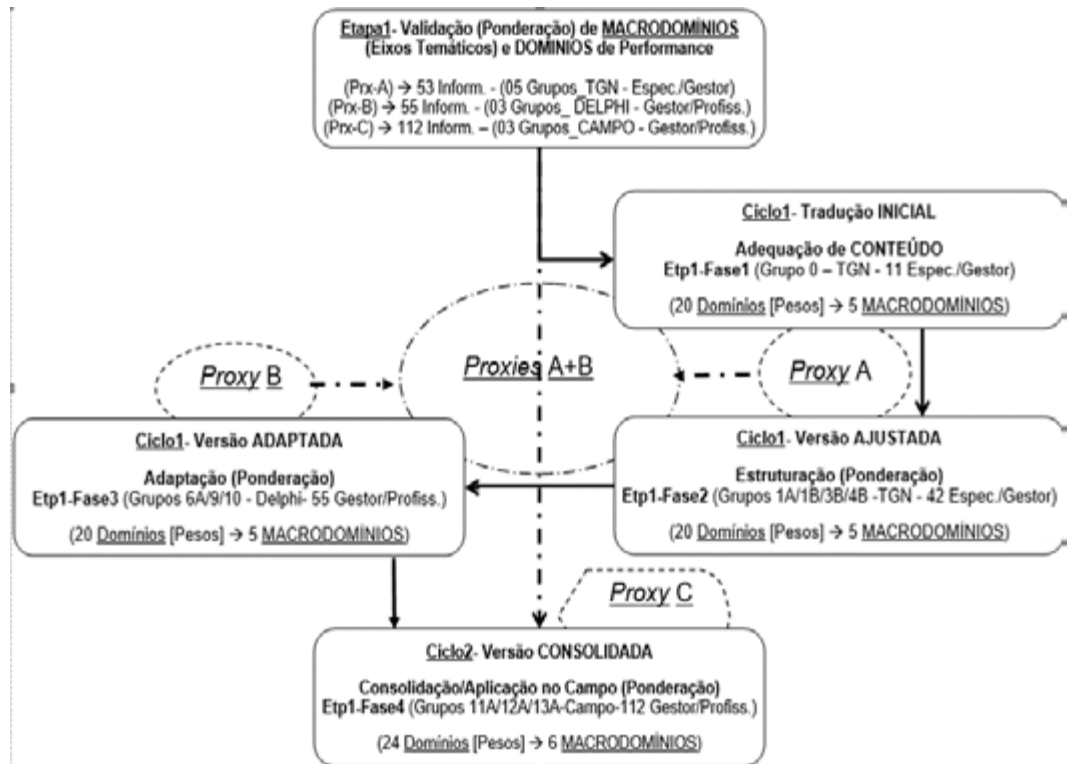
Foi inicialmente selecionado um grupo-piloto de gestores e profissionais para formar um painel de 11 especialistas (fase 1), que fizeram a adequação de conteúdo, devidamente analisado por 42 especialistas e gestores de saúde em quatro grupos-piloto (fase 2) que, em seguida, foi validada por outros três grupos pilotos de 55 profissionais e gestores de saúde (fase 3), conformando um *proxy* inicial. Por fim, aplicou-se uma pesquisa de campo de validação por 112 gestores de saúde que consolidaram um *proxy* final de avaliação de performance em unidades de saúde.

Os informantes-chave selecionados foram questionados sobre o grau de relevância e preferência dos macrodônios (macrodimensões) e domínios (indicadores) de avaliação de desempenho em unidades e serviços de saúde pela proposição e ponderação de novos itens de mensuração de desempenho, segundo o seu grau de conhecimento pessoal.

A partir dos conceitos e possíveis categorias analíticas que emergiram dos painéis de especialistas, propôs-se um embasamento teórico-metodológico para estruturar instrumentos (questionários semiestruturados) construídos em etapas e fases sucessivas de colaboração com grupos de 220 especialistas, profissionais, gestores provenientes de 11 estados brasileiros – Bahia, Ceará, Distrito Federal, Goiás, Maranhão, Mato Grosso, Pará,

Piauí, Rio Grande do Norte, Rondônia e Paraná –, divididas em quatro fases sucessivas e incrementais de adaptação, sintetizadas na Figura 3.

Figura 3 | Fluxo do processo de validação de macrodomínios e domínios de avaliação (fases 1 a 4)



Fonte: Elaborado pelos autores a partir de (MOITA, 2019; MOITA; RAPOSO; BARBOSA, 2019a; b).

Esses fluxos têm suas análises complementadas pelo esquema lógico de análise de macrodomínios e domínios de avaliação. Na primeira etapa, no âmbito do estudo 1, procedeu-se à análise estatística de dados das *proxies* da valoração dos domínios de avaliação, além de seus reflexos nos macrodomínios de avaliação, descritos no Quadro 2.

Quadro 2 | Síntese das etapas do processo analítico

Análise Cód.	Etapas/Estudo	Fases	Proxy	Informantes (N) / Categorias
A1) Análises de respostas atribuídas através de 4 <i>proxies</i> (pesos atribuídos para os domínios)				
A1.a)-Inicial	Etp1 – Ciclo1	1+2	A	53 (11+42) Espec./Gestor
A1.b)-Ajuste	Etp1 – Ciclo1	3+4	B	55 (11+42) Gestor/Profiss.
A1.c)-Síntese	Etp1 – Ciclo1	1+2+3+4	A+B	108 (53+55) Gestor/Profiss.
A1.d)-Campo	Etp1 – Ciclo2	5	C	112 Gestor/Profiss.
A.2) Análises de respostas calculadas através de 4 <i>proxies</i> (pesos refletidos para os macrodomínios):				
A2.a)-Inicial	Etp1 – Ciclo1	1+2	A	53 (11+42) Espec./Gestor
A2.b)-Ajuste	Etp1 – Ciclo1	3+4	B	55 (11+42) Gestor/Profiss.
A2.c)-Síntese	Etp1 – Ciclo1	1+2+3+4	A+B	108 (53+55) Gestor/Profiss.
A2.d)-Campo	Etp1 – Ciclo2	5	C	112 Gestor/Profiss.
A3) Análises de respostas calculadas através de um <i>proxy</i> (pesos refletidos para os macrodomínios e para a Performance Global Percebida)				
A3.a)-Campo	Etp1 – Ciclo2	5	C	112 Gestor/Profiss.

Fonte: Elaborado pelos autores a partir de (MOITA, 2019).

A aplicação inicial da pesquisa abordou um grupo focal de 11 especialistas em avaliação e gestão de saúde (fase 1), com validação posterior por quatro grupos-piloto de 42 gestores e profissionais (fase 2), para identificar os principais domínios e macrodomínios de avaliação (dimensões de desempenho) considerados mais relevantes na mensuração de performance em organizações de saúde, pela Técnica de Grupo Nominal (TGN).

A fase 3 manteve os fundamentos do modelo lógico proposto com agregações de todos os domínios e macrodomínios de avaliação de desempenho que receberam valoração acima da média (superior a 3, relevante/importante) nas fases 1 e 2. Nessa nova fase, captou-se respostas de três grupos-piloto (55 gestores e profissionais de gestão de saúde pública), sendo que os novos entrevistados poderiam livremente agregar novos itens de avaliação (domínios e/ ou macrodomínios) que julgassem relevantes, a fim de consolidar os níveis de conhecimentos específicos finais sobre os domínios e macrodomínios de avaliação em sistemas e serviços de saúde.

Em seguida, os resultados obtidos nas respostas das fases 1 a 3 foram avaliados, categorizados e readequados conforme as macrodimensões de performance definidas pela literatura mais relevante (ACURCIO; CHERCHIGLIA; SANTOS, 1991; AKERMAN; NADANOVSKY, 1992; GATTINARA; IBACACHE; PUENTE; GIACONI *et al.*, 1995; SANTOS, 1995; VIACAVA; ALMEIDA; CAETANO; FAUSTO *et al.*, 2004; VIEIRA-DA-SILVA, 2005; VUORI, 1991), subsidiando o ajuste e a consolidação de um novo questionário semiestruturado (fase 4).

Por fim, na fase 4, organizou-se outros três grupos-piloto de informantes-chave, gestores e profissionais de saúde, para a aplicação de um instrumento ajustado, em uma amostra de 112 gestores e profissionais que emergiu dos resultados obtidos da fase 3.

Os dados coletados foram tabulados em planilhas e analisados pelas médias ponderadas das notas atribuídas (valoração de importância/relevância) por meio de uma escala ordinal de *Likert* de cinco níveis (1 = menor nível/grau e 5 = maior nível/grau) de relevância/importância atribuída pelos entrevistados. Em cada fase, as respostas foram consolidadas com o software Excel 2016, traçados gráficos de valores e aplicada análise descritiva e inferencial com o software R (versão 3.3.2), com versões de licenças livres.

A análise descritiva dos resultados obtidos foi aplicada de forma integrada na validação do questionário (fases 1 a 3: *proxy A+B*) e, posteriormente, na pesquisa de campo (fase 4: *proxy C*), e, de forma complementar, uma análise para mensurar os níveis de significância estatística quanto à possível influência de três variáveis mediadoras destacadas (nível de atenção à saúde, nível de complexidade da assistência e nível de cargo/função ocupado) quanto à valoração dos quatro macrodomínios de avaliação (quantitativos, qualitativos, efeitos e satisfação), tendo ainda a adição do sistêmico/estratégico, para a posterior fase 4.

Por fim, aplicaram-se aos resultados obtidos as análises descritivas e significância estatística, além da análise inferencial, e elaborou-se um modelo de validação dos 24 indicadores (domínios de avaliação) e cinco dimensões (macrodomínios de avaliação), no *proxy C*, com eliminação de alguns itens, por intermédio de análises de confiabilidade e validade (por meio de Correlação e Análise Fatorial Exploratória), que permita futuramente estruturar um instrumento de avaliação de qualidade percebida (satisfação e adequação ao uso), com questões adequadas ao SUS.

4. Resultados

Inicialmente, abordou-se uma análise integrada de fases incrementais (fases 1 a 3, *proxies* A e B) que se iniciou pela validação de face (fase 1) e aplicou dois instrumentos diferentes (fases 2 e 3), através de uma amostra de 108 informantes especialistas, gestores e profissionais de saúde. Efetivou-se uma análise integrada das respostas obtidas nestes dois instrumentos (*proxy* A+B), pois apresentaram muitas similaridades (20 domínios e quatro macrodomínios de performance). Em fase posterior (fase 4), abordou-se a análise da *proxy* C – 112 informantes (24 domínios e cinco macrodomínios de performance).

4.1. Validação por análise descritiva

A Tabela 1 apresenta uma análise descritiva das variáveis de caracterização da amostra da *proxy* A+B (108 informantes). Dessa forma, a análise integrada da fase 1 a 3 (n=108), comparativamente à fase 4 (n=112), resultou nos seguintes valores: a amostra foi representada por trabalhadores dos níveis secundário (29,47%; 13,70%), primário (26,32%; 8,22%), terciário (26,32%; 15,07%) de atenção à saúde, e do nível outros (estruturas não assistenciais, como apenas de gestão ou governança – 17,89%; 63,01%). A maior parte dos indivíduos das fases 1 a 3 trabalhava no nível da média complexidade da assistência (33,33%), seguida pela alta complexidade (27,08%), pela baixa complexidade (25,00%) e 14,58% trabalhavam no nível de complexidade outros (estruturas não assistenciais, como apenas de gestão ou governança). Por sua vez, na fase 4, 61,64% trabalhavam no nível de complexidade outros, seguido pelos indivíduos que trabalhavam nos níveis da alta complexidade (15,07%), da média complexidade (15,07%) e da baixa complexidade (8,22%). Menos da metade dos indivíduos (36,11%) exercia um dos três níveis de gestão do estudo (sistêmico/governacional, organizacional/ estratégico ou gerência/chefia setorial) nas fases 1 a 3, sendo que na fase 4 representavam 74,11% da amostra. A maior parte dos indivíduos das fases 1 a 3 (63,89%) exercia o nível de cargo/função outros (por exemplo, assessores de gestão, profissionais em geral e outros); porém, na fase 4, a maior parte (41,96%) era representante do nível de gestão sistêmico/governacional.

Em todas as fases, a maioria dos indivíduos (38,46%, fase 1 a 3; 76,58%, fase 4) possuía pós-graduação e ainda 34,62% (fase 1 a 3) e 12,61% (fase 4) possuíam nível

superior. Respectivamente, nas fases 1 a 3 e na fase 4, a maioria dos indivíduos (49,07%; 70,27%) possuía mais de 60 meses de experiência; boa parte dos indivíduos (20,37%; 17,92%) trabalhava há mais de 60 meses em cargo/função de decisão, enquanto 42,59% e 13,21% não possuía experiência em cargo/função de decisão.

Tabela 1 | Análise descritiva das variáveis de caracterização selecionadas (fases 1 a 4)

Variáveis	Fases 1 a 3 (ciclo 1)		Fases 4 (ciclo 2)		
	N	%	N	%	
Nível de Atenção à Saúde	Atenção Primária	25	26,32%	6	8,22%
	Atenção Secundária	28	29,47%	10	13,70%
	Atenção Terciária	25	26,32%	11	15,07%
	Outros	17	17,89%	46	63,01%
Nível de complexidade Assistência	Baixa	24	25,00%	6	8,22%
	Média	32	33,33%	11	15,07%
	Alta	26	27,08%	11	15,07%
	Outros	14	14,58%	45	61,64%
Nível de cargo/função	Sistêmico	7	6,48%	47	41,96%
	Organizacional	13	12,04%	16	14,29%
	Gerencia de Equipe	19	17,59%	20	17,86%
	Outros	69	63,89%	29	25,89%
Escolaridade	Analfabeto	0	0,00%	0	0,00%
	Ensino Básico	0	0,00%	0	0,00%
	Ensino Fundamental	0	0,00%	0	0,00%
	Ensino Médio	10	9,62%	2	1,80%
	Ensino Superior	36	34,62%	14	12,61%
	Pós-Graduação	40	38,46%	85	76,58%
Tempo de experiência	Mestrado/ PhD	18	17,31%	10	9,01%
	Não tem	9	8,33%	0	0,00%
	Menos de 6 meses	11	10,19%	3	2,70%
	6 - 12 meses	6	5,56%	3	2,70%
	13 - 24 meses	14	12,96%	3	2,70%
	25 - 36 meses	6	5,56%	8	7,21%
	37 - 48 meses	4	3,70%	7	6,31%
	49 - 60 meses	5	4,63%	9	8,11%
> 60 meses	53	49,07%	78	70,27%	
Tempo em cargo de decisão	Não tem	46	42,59%	14	13,21%
	Menos de 6 meses	13	12,04%	19	17,92%
	6 - 12 meses	7	6,48%	16	15,09%
	13 - 24 meses	4	3,70%	4	3,77%
	25 - 36 meses	6	5,56%	14	13,21%
	37 - 48 meses	3	2,78%	8	7,55%
	49 - 60 meses	7	6,48%	12	11,32%
	> 60 meses	22	20,37%	19	17,92%
Formação/Profissão	Enfermeiro	25	23,36%	45	40,18%
	Adm. Hospitalar	24	22,43%	20	17,86%
	Assistente Social	11	10,28%	11	9,82%
	Tecn. Enfermagem	7	6,54%	6	5,36%
	Fisioterapeuta	6	5,61%	5	4,46%
	Outras	34	31,78%	25	22,32%

Fonte: elaboração própria a partir de (MOITA, 2019)

Em termos de variáveis e indicadores (domínios e macrodomínios de avaliação) que emergiram dos conhecimentos prévios dos respondentes, pode-se sintetizar as seguintes categorias:

Quadro 3 | Síntese dos domínios e macrodomínios prevalentes na pesquisa inicial

Macrodomínios	Definições conceituais
I. Quantitativos	Seis domínios referentes a aspectos de produtividade, estrutura assistencial, cobertura e de finanças/economia (D1A_Produtiv – produtividade assistencial, D1B_EstrutAssist estrutura assistencial; D1C_CobertAssit- cobertura de atendimentos, D1D_OtimizCusto- otimização de custos, D1E_EficienEcon – eficiência econômica e D1F_SustEconFin- sustentabilidade econômica e financeira).
II. Qualitativos	Sete domínios referentes a aspectos de qualidade, humanização e integralidade assistencial, acessibilidade e entrega de serviços e equidade (D2H_QualidAssist- qualidade da atenção, D2I_RiscoAssist- risco assistencial, D2J_AtendHumaniz- atendimento humanizado, D2K_AcessibilServ- acessibilidade de serviços, D2L_AcessoServ- acesso aos serviços, D2M_IntegralAtenc- integralidade da atenção e D2N_Equidade- equidade no acesso).
III. Efeitos	Cinco domínios referentes a aspectos de eficácia, efetividade, eficiência, resolutividade e impacto (D3P_EficaciaProj- eficácia projetada, D3Q_EfetivClinica- efetividade clínica, D3R_EficienAssist- eficiência da assistência, D3S_ResolubAssist- resolutividade assistencial e D3T_Impacto- impacto na saúde populacional).
IV. Satisfação	Dois domínios referentes a aspectos de qualidade e satisfação (D4U_SatisfUsuario- satisfação do usuário e D4V_SatisfProfiss- satisfação do profissional).
V. Sistêmico/estratégico	Cinco domínios referentes a aspectos de gestão estratégica do sistema de saúde (D5W_IndICSAP- internações evitáveis por condições sensíveis a APS, D5X_AtendVincReg- atendimento de usuário vinculado a região de saúde, D5Y_AtendDesVincReg- atendimento de usuário vinculado a região de saúde e D5V_ExtrapTetoOrç- extrapolação de teto orçamentário).

Fonte: Elaboração própria a partir de (MOITA, 2019; MOITA; RAPOSO; BARBOSA, 2019a).

Ressalte-se que pelo surgimento de termos similares, mas divergentes da literatura, não houve vinculação desses indicadores (domínio e macrodomínios de avaliação) com termos e categorias prevalentes na literatura, por exemplo, do Anexo I, que foram adaptados e equalizados nas etapas posteriores de pesquisa de campo.

A Tabela 2 apresenta uma análise descritiva relativa à valoração média atribuída aos domínios de avaliação, na análise integrada das *proxies* A e B (108 informantes), no ciclo 1 de validação do instrumento, comparativamente com a *proxy* C (112 informantes), no ciclo 2 de aplicação de campo, na visão de especialistas, profissionais e gestores.

Ressalte-se que os respondentes acrescentaram quatro novos domínios de avaliação (*D5W_IndICSAP*, *D5X_AtendVincReg*, *D5Y_AtendDesVincReg*, *D5V_ExtrapTetoOrç*), adicionados em um novo macrodomínio de avaliação (sistêmico/estratégico), que não tiveram respostas individualizadas nas análises iniciais (fases 1 a 3), mas foram incorporados nos instrumentos (*proxy* C) da nova fase de investigação (fase 4).

Tabela 2 | Valoração dos domínios de avaliação por fase do estudo (fases 1 a 3 vs fase 4)

Macrodómnios	Domínios	Fases 1 e 2				Fase 3			
		N	Média	D.P.	I.C.95%	N	Média	D.P.	I.C.95%
Quantitativos	D1A_Produtivo	108	3,85	0,90	[3,69; 4,03]	111	4,28	0,86	[4,12; 4,42]
	D1B_EstrutAssist	107	4,00	0,90	[3,84; 4,16]	111	4,40	0,79	[4,25; 4,54]
	D1C_CobertAssist	95	4,23	0,80	[4,06; 4,38]	102	4,36	0,81	[4,20; 4,51]
	D1D_OtimizCusto	108	4,23	0,80	[4,07; 4,36]	111	4,14	0,82	[3,99; 4,28]
	D1E_EficienEcon	108	4,30	0,90	[4,15; 4,45]	111	4,27	0,82	[4,13; 4,42]
	D1F_SustEconFin	81	4,24	0,80	[4,06; 4,40]	111	4,15	0,88	[4,00; 4,31]
	D1G_Outros	11	4,18	0,87	[3,64; 4,64]	6	4,50	0,84	[3,83; 5,00]
Qualitativos	D2H_QualidAssist	108	4,46	0,79	[4,30; 4,61]	111	4,38	0,78	[4,23; 4,51]
	D2I_RiscoAssist	107	4,38	0,75	[4,24; 4,52]	110	4,40	0,68	[4,27; 4,53]
	D2J_AtendHumaniz	108	4,58	0,60	[4,46; 4,70]	109	4,35	0,77	[4,20; 4,49]
	D2K_AcessibilServ	88	4,25	0,73	[4,10; 4,41]	112	4,12	0,81	[3,97; 4,27]
	D2L_AcessoServ	106	4,36	0,72	[4,21; 4,50]	112	4,26	0,83	[4,10; 4,42]
	D2M_IntegralAtenc	107	4,24	0,80	[4,09; 4,39]	110	4,04	0,92	[3,85; 4,20]
	D2N_Equidade	96	4,20	0,85	[4,02; 4,35]	55	3,84	1,05	[3,56; 4,11]
D2O_Outros	6	3,83	0,98	[3,17; 4,50]	3	4,33	0,58	[4,00; 5,00]	
Efeitos	D3P_EficienciaPrej	106	3,90	0,85	[3,73; 4,08]	112	3,89	1,00	[3,71; 4,08]
	D3Q_EfetivClinica	107	4,33	0,70	[4,20; 4,45]	112	4,48	0,74	[4,35; 4,61]
	D3R_EficienAssist	105	4,36	0,70	[4,24; 4,50]	112	4,29	0,81	[4,13; 4,44]
	D3S_ResolubAssist	94	4,48	0,60	[4,35; 4,61]	110	4,50	0,75	[4,35; 4,64]
	D3T_Impacto	94	4,30	0,80	[4,14; 4,45]	111	4,38	0,74	[4,24; 4,52]
Satisfação	D4U_SatisUsuario	95	4,54	0,70	[4,40; 4,67]	111	4,26	0,84	[4,11; 4,41]
	D4V_SatisProfiss	37	3,78	1,60	[3,24; 4,24]	112	4,19	0,89	[4,02; 4,35]
Sistêmico / Estratégico	D5W_IndICSAP					111	3,87	1,14	[3,67; 4,07]
	D5X_AtendVincReg					112	3,76	0,87	[3,59; 3,92]
	D5Y_AtendDesVincReg					111	3,26	1,01	[3,08; 3,45]
	D5V_ExtrapTetoOrç					111	3,92	1,00	[3,73; 4,11]
Outros	D6X_Outros	5	4,40	0,90	[3,60; 5,00]	1	4,00	-	[1,50; 3,75]
	D6Z_Outros	1	4,00	-	-	1	5,00	-	[1,80; 4,20]

Fonte: Elaboração própria a partir de (MOITA, 2019)

A análise global permite afirmar que a ampla maioria dos 20 domínios de avaliação nas fases 1 a 3, e dos 24 domínios de avaliação na fase 4, obtiveram valoração média elevada (superior ou cerca de 4, muito importante/relevante), sendo considerados relevantes/importantes para a estruturação de novos instrumentos com domínios e macrodomínios de avaliação devidamente alinhados pela literatura relevante de dimensões e indicadores de performance (ARAH; KLAZINGA; DELNOIJ; ASBROEK *et al.*, 2003) em fases posteriores desta investigação. Por sua vez, a análise dos macrodomínios de avaliação estruturou-se a partir da valoração média atribuída aos domínios de avaliação pelos indivíduos e, em seguida, calculadas para os macrodomínios de avaliação.

A Tabela 3 apresenta uma análise descritiva relativa à valoração média obtida para os macrodomínios de avaliação, na análise integrada das *proxies* A e B (108 informantes), no ciclo 1 de validação do instrumento, comparativamente com a *proxy* C (112 informantes), no ciclo 2 de aplicação de campo, na visão de especialistas, profissionais e gestores.

Tabela 3 | Análise descritiva dos macrodomínios de avaliação – Proxy A+B (ciclo 1) e proxy C (ciclo 2) (estudo 1) (visão de especialistas)

Macrodomínios	Proxy A+B (fases 1 a 3)				Proxy C (fase 4)			
	N	Média	D.P.	I.C-95% ¹	N	Média	D.P.	I.C-95% ¹
Quantitativos	108	4,12	0,54	[4,01; 4,22]	111	4,27	0,53	[4,17; 4,37]
Qualitativos	108	4,36	0,47	[4,27; 4,44]	112	4,23	0,55	[4,13; 4,33]
Efeitos	107	4,27	0,50	[4,18; 4,36]	112	4,31	0,58	[4,20; 4,41]
Satisfação	96	4,40	0,83	[4,23; 4,56]	112	4,22	0,77	[4,09; 4,37]
Outros	5	4,40	0,89	[3,60; 5,00]	112	3,70	0,74	[3,56; 3,83]
					1	4,50	-	[4,50; 4,50]

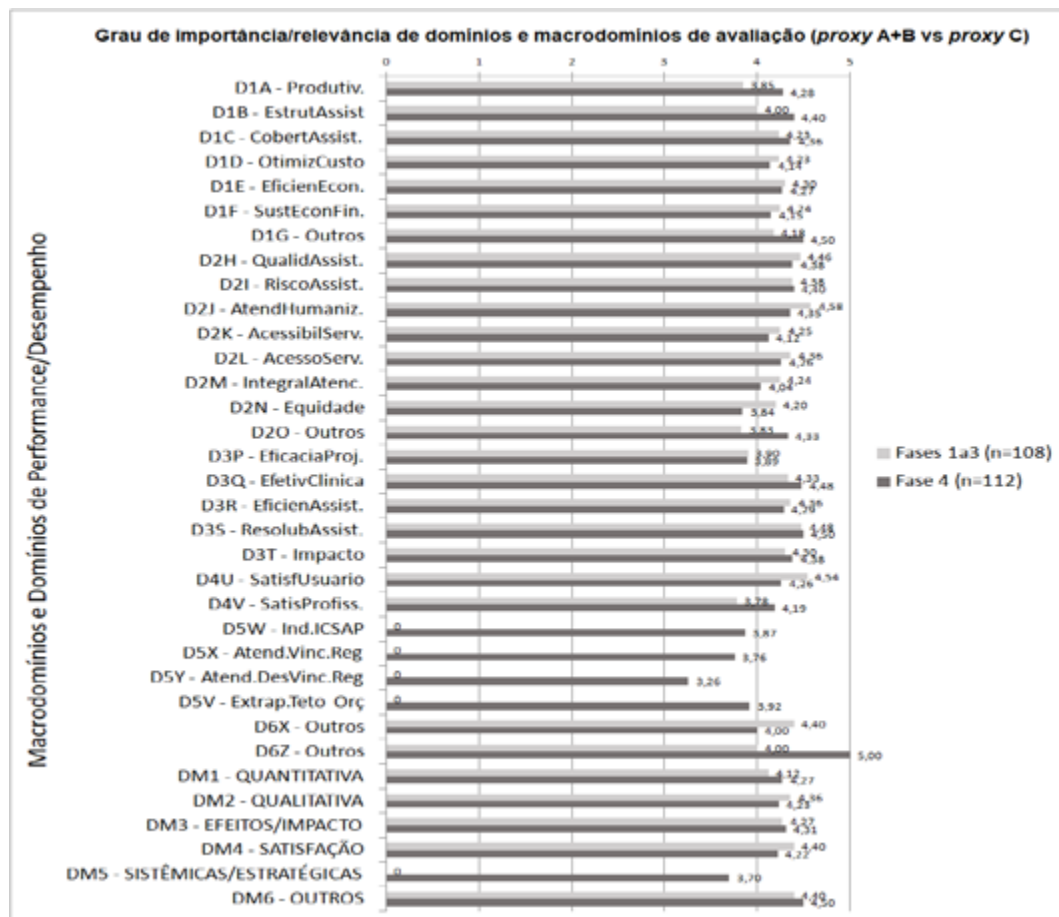
Fonte: Elaboração própria a partir de (MOITA, 2019) – Nota.¹ Intervalo Bootstrap.

A análise global permite afirmar que todos os quatro macrodomínios de avaliação apresentaram valoração média elevada (superior ou cerca de 4, muito importante/relevante), nas *proxies* A e B (n=108) e na *proxy* C (n = 112), além do quinto macrodomínio (sistêmico/estratégico) adicional, composto por agregações de quatro novos domínios de avaliação (D5W_IndICSAP, D5X_AtendVincReg, D5Y_AtendDesVincReg, D5V_ExtrapTetoOrç).

Quanto às análises de significância estatística, com relação às possíveis influências de fatores mediadores das notas obtidas pelos macrodomínios, em síntese, a variável nível de atenção à saúde pode ser considerada como fator mediador apenas dos macrodomínios de avaliação Quantitativos e Efeitos; as variáveis nível de cargo/função e nível de complexidade da assistência não puderam ser consideradas como fator moderador de nenhum dos cinco macrodomínios de avaliação (quantitativos, qualitativos, efeitos, satisfação e sistêmico/estratégico), na *proxy* C (fase 4).

Desta análise descritiva, sintetizou-se que os especialistas e profissionais atribuíram elevados (superior ou cerca de 4) graus de importância/relevância das subcategorias (20 e 24 domínios de avaliação) e das categorias (cinco e seis macrodomínios de avaliação) dos itens de análise, sendo considerados relevantes/importantes para a estruturação de um instrumento de avaliação de performance, conforme a Figura 4. Além disso, em média, as valorações da ampla maioria desses itens de análise foram razoavelmente dentro das mesmas margens de valores, uma vez que a maioria dos intervalos de confiança se sobrepôs.

Figura 4 | Grau de importância/relevância de domínios e macrodomínios de avaliação (proxy A+B versus proxy C) (visão de especialistas, profissionais e gestores de saúde)



Fonte: Elaboração própria a partir de (MOITA, 2019)

Quanto aos resultados do estudo das possíveis influências de três variáveis moderadoras selecionadas, não houve tendência de mediação de valorações médias obtidas na ampla maioria dos macrodomínios de avaliação quanto aos níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos especialistas e profissionais, com raras exceções. Esses resultados subsidiaram a construção de um novo instrumento, devidamente alinhado pela literatura relevante de indicadores e dimensões de performance (ARAH; KLAZINGA; DELNOU; ASBROEK *et al.*, 2003), que será aplicado na fase posterior da investigação, em planejamento de pesquisa de campo.

4.2. Validação e confiabilidade por análise inferencial

Concluídas as análises descritivas, em busca de validação, foi feita uma Análise Fatorial Confirmatória (AFC) para as subdimensões de constructos, na *proxy C* ($n = 112$). Logo, pode-se destacar que, no modelo final, após análise e exclusão de sete domínios de avaliação (D1A_Produtivo, D1B_EstrutAssist, D1C_CobertAssit, D2H_QualidAssist, D2M_IntegralAtenc, D2N_Equidade e D3P_EficaciaProj), todos os demais 17 domínios de avaliação que permaneceram na análise apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy C*.

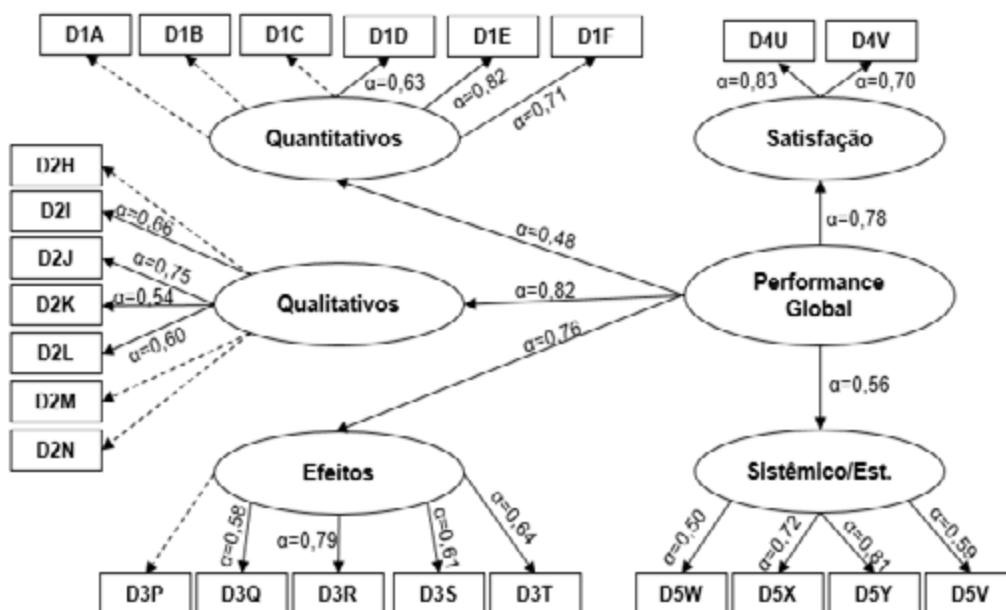
Em outra análise, a qualidade e a validade dos cinco macrodomínios de avaliação analisados (quantitativos, qualitativos, efeitos, satisfação e sistêmico/estratégico), e também da variável latente Performance Global Percebida, foram asseguradas, uma vez que todas apresentaram validação convergente ($AVE > 0,40$), confiabilidade adequada ($A.C. > 0,60$ ou $C.C. > 0,60$), unidimensionalidade e validação discriminante ($VCM < AVE$), na *proxy C* ($n = 112$).

Para validar a variável latente Qualidade Global Percebida, foi feita uma Análise Fatorial Confirmatória (AFC) (HAIR; WILLIAM; BABIN; ANDERSON, 2009) da valoração média obtida para as macrodimensões de qualidade, na *proxy C* ($n = 112$). Logo, pode-se destacar que, no modelo final, todos os cinco macrodomínios (quantitativos, qualitativos, efeitos, satisfação e sistêmico/estratégico) permaneceram na análise e apresentaram carga fatorial superior a 0,50 (ou suas permanências não impediram a validação convergente da AFC), na *proxy C*.

Também se aplicaram testes para medir a qualidade (ajustamento) do modelo construído para os macrodomínios e domínios de avaliação da *proxy C* (112 informantes), por meio de alguns índices selecionados na literatura relevante ($\chi^2/G.L.$, CFI, TLI e RMSEA). Pode-se observar que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste, já que a razão entre a estatística qui-quadrado e os graus de liberdade foi menor que 3, as estatísticas TLI e o CFI foram maiores que 0,80 e o RMSEA foi menor que o limite máximo de 0,10.

Diante desses resultados, a Figura 5 sintetizou o ajuste da Análise Fatorial Confirmatória (AFC) e a modelagem para a variável latente Performance Global Percebida, na *proxy* C (112 informantes, visão direta de gestores e profissionais). Devido ao elevado número de domínios de avaliação, omitiu-se suas nomenclaturas, que ficaram identificados como D1A a D5Y.

Figura 5 | Análise Fatorial Confirmatória da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (visão de gestores e profissionais)



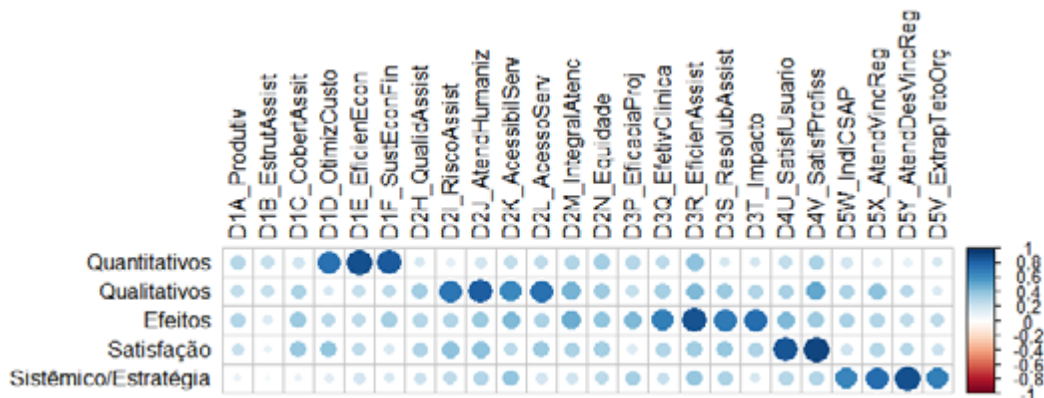
Fonte: Elaboração própria a partir de (MOITA, 2019).

Pode-se destacar que, no modelo final, cinco dos 24 domínios de avaliação (indicadores) propostos pelo painel de especialistas apresentaram cargas fatoriais validadas elevadas (0,76 a 1,00), que impactam fortemente seus respectivos macrodomínios, e outro grupo de 12 domínios com cargas fatoriais medianas (0,50 a 0,75); enquanto sete outros (D1A, D1B, D1C, D2H, D2M, D2N e D3P) foram excluídos do modelo final. Nenhum dos domínios vinculados aos macrodomínios Satisfação e Sistêmico/Estratégico foram eliminados no modelo final, revelando ampla concordância entre as propostas do painel de especialistas e as valorações dos grupos-piloto. Em seguida, a análise de carga fatorial validou no modelo final os cinco macrodomínios de avaliação (quantitativo quase no limite de 0,50), sendo que três macrodomínios (qualitativo, efeitos e satisfação) tiveram

elevado nível de carga fatorial (0,76 a 1,00), tendo os outros dois (quantitativo e sistêmico/estratégico) apenas carga fatorial mediana (0,50 a 0,75), que impactam fortemente a variável latente Performance Global Percebida (PGP). Assim, garantiu-se a confiabilidade e a validade de 17 domínios (indicadores) e de quatro macrodomínios (dimensões) de performance, além de modelar os impactos de seus reflexos sobre a PGP. Após análise e exclusão de sete domínios (3 quantitativos, 3 qualitativos e 1 de efeitos), todos os demais 17 domínios de avaliação que permaneceram na análise apresentaram carga fatorial superior a 0,50; ou seja, nesse modelo final, garantiu-se a confiabilidade e a validade de 17 domínios de avaliação e de todos os cinco macrodomínios de avaliação.

Em outra vertente, a análise de correlação mostrou que os domínios de avaliação apresentaram maiores correlações com seus respectivos macrodomínios, na proxy C (fase 4) (Figura 5).

Figura 5 | Correlações entre domínios e macrodomínios de avaliação - Proxy C (ciclo 2, estudo 1) (visão de gestores e profissionais)



Fonte: Elaboração própria a partir de (MOITA, 2019).

Por fim, a performance global percebida média foi de 4,16 e a mediana de 4,18; 75% das respostas atribuídas resultaram em valoração superior a 3,88 (cerca de 4, muito importante/relevante) para a PGP média.

5. Considerações finais

Os resultados obtidos das respostas de 108 especialistas e profissionais de gestão e de 112 gestores de saúde, em quatro fases sucessivas (painel de especialistas, grupos *Delphi* e pesquisa de campo), foram objetos de análise descritiva e de significância estatística, que demonstrou que os informantes atribuíram elevados (superior ou cerca de 4) graus de importância/relevância das subcategorias (20 e 24 domínios de avaliação) e das categorias (cinco e seis macrodomínios de avaliação) dos itens de análise, sendo considerados relevantes/importantes (Figura 4) para a aplicação da análise inferencial. Além disso, em média, as valorações da ampla maioria desses itens de análise foram razoavelmente dentro das mesmas margens de valores, uma vez que a maioria dos intervalos de confiança se sobrepuseram. E, ainda, que não houve tendência de mediação de valorações médias obtidas na ampla maioria dos macrodomínios de avaliação (quantitativos, qualitativos, efeitos, satisfação e estratégico/sistêmico) quanto às variáveis níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos especialistas, profissionais, e gestores, com raras exceções, e também na variável latente PGP, sugerindo-se novos estudos com amostras ampliadas.

Concluídas as análises descritivas, aplicou-se uma Análise Fatorial Confirmatória (AFC) (HAIR; WILLIAM; BABIN; ANDERSON, 2009). Foi garantida a confiabilidade e a validade de 17 domínios (indicadores) e de quatro macrodomínios (dimensões) de mensuração de performance, revelando ampla concordância entre as propostas do painel de especialistas e as valorações dos grupos pilotos quanto à ampla maioria dos indicadores e dimensões. Também se aplicou testes para medir a qualidade (ajustamento) do modelo construído para as macrodimensões de qualidade percebidas através de alguns índices selecionados na literatura relevante ($\chi^2/G.L.$, CFI, TLI e RMSEA). Pode-se observar que os parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais indicaram bom ajuste. No modelo final, foram validados 17 domínios (indicadores) e quatro macrodomínios de avaliação, além de modelar os impactos de seus reflexos sobre a PGP (Figura 5).

Por fim, demonstrou-se que a performance global percebida média foi de 4,16 e a mediana de 4,18; 75% das respostas atribuídas resultaram em valoração superior a 3,88

(cerca de 4, muito importante/relevante) e que as análises de qualidade e validade dos macrodomínios de avaliação foram asseguradas enquanto fatores formadores da PGP.

Desses resultados das análises descritiva e fatorial, demonstrou-se que há diferentes importâncias relativas atribuídas por especialistas e gestores de saúde entre os indicadores e dimensões de mensuração de performance, mas a não validação de sete dos indicadores propostos revela a necessidade de novos estudos, em amostras maiores, tendo sido um processo inovador que estabeleceu subcategorias (indicadores) e categorias (dimensões) que podem subsidiar a construção de um novo instrumento aplicável aos gestores do SUS, já em planejamento para avanços da pesquisa, inclusive por intermédio do desenvolvimento de protótipo de uma plataforma de informática, usado como suporte tecnológico para aplicação do instrumento validado.

Pode-se sintetizar que não houve diferença significativa quanto às variáveis nível de atenção à saúde, nível de complexidade da assistência e nível de escolaridade de usuários, para se considerar enquanto fatores mediadores em todos os macrodomínios e domínios de avaliação (com raras exceções).

Os resultados consolidados abrem novas perspectivas para a gestão dos serviços de saúde, no sentido de se obter informações cientificamente válidas acerca da percepção dos especialistas e gestores da saúde, possivelmente evitando as dificuldades e desalinhamentos de métodos de aferição sem validação por especialistas e contribuindo para a legitimidade e usabilidade dos instrumentos analíticos para comparação entre diferentes unidades avaliadas, enquanto itens de orientação para intervenção de gestores nas unidades de saúde do SUS.

6. Referências bibliográficas

ACURCIO, F. A.; CHERCHIGLIA, M. L.; SANTOS, M. A. Avaliação de qualidade de serviços de saúde. *Saúde em debate*. Rio de Janeiro. v. 33, n. 4, 1991.

AKERMAN, M.; NADANOVSKY, P. Evaluation of Health Services - What to Evaluate? *Cad. Saúde Pública*, Rio de Janeiro, v. 8, n. 5, 1992.

ALKIN, M. C.; CHRISTIE, C. A. Evaluation Roots: tracing theorist's views and influences. *Sage publications*, 2004.

ALVES, A. S. D. *Responsabilidade e Governança na Moderna Gestão Hospitalar*. 2012. Tese (Doutorado). Universidade de Aveiro, Portugal: Aveiro, 2012.

ARAH, O. A. *et al.* Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement. *International Journal for Quality in Health Care*, v. 15, n. 5, p. 22, 2003.

BARBOSA, L.; NETO, A. P. Ludwik Fleck (1896-1961) e translação do conhecimento: considerações sobre a genealogia de um conceito. *Saúde Debate*, Rio de Janeiro, v. 41, n. 13, . 2017.

BOSI, M. L.; MERCADO, F. J. *Pesquisa qualitativa de serviços de saúde*. 2006.

BROOK, R. H.; LOHR, K. Efficacy, effectiveness, variations and quality: boundary-crossing research. *Med Care*, v. 23, n. 5, p. 23, 1985.

BRYSON, J. What to do when stakeholders matter: stakeholder identification and analysis techniques. *Public Management Review*, v.6, n. 1, p. 33, 2004.

BUNGE, M. Os conceitos do modelo: modelo na ciência teórica. In: *Teoria e realidade*. São Paulo: Perspectiva Editora, 1974. p. 11-40.

CAMPOS, R. T. O.; FURTADO, J. P. *Desafios da avaliação de programas e serviços em saúde: novas tendências e questões emergentes*. São Paulo: Editora Unicamp, 2011.

CARVALHO, C. *Organizações, actores envolvidos e partes interessadas: determinantes da saliência dos stakeholders e sua relação com o desempenho organizacional*. 2007. Tese (Doutorado) – Universidade de Coimbra, Coimbra, Portugal, 2007.

CHIANCA, T.; YOUKER, B. *La evaluación em América Latina y El Caribe. Visión general de los desarrollos recientes*. 2004.

CIHR, C. I. O. H. R. *Knowledge Translation Strategy 2004-2009: innovation in action*. CIHR. Ottawa, 2004.

CLAVIER, C.; SÉNÉCHAL, Y.; POTVIN, L. A theory - based model of translation practices in public health participatory research. *Henley-on-Thames: sociology of Health and Illness*, v. 34, n. 16, 2011.

CONTANDRIOPOULOS, A. P. Avaliando a institucionalização da avaliação. *Ciênc. Saúde Coletiva*, v. 10, n. 7, 2006.

DONABEDIAN, A. Basic approaches to assessment: structure, process and outcome. In: PRESS, H. A. (Ed.). *Explorations in quality assessment and monitoring*. Michigan: Health Administration Press, 1980. v. 1, p. 77-125.

DOS REIS, E. J. F. B. *et al.* Avaliação da qualidade dos serviços de saúde: notas bibliográficas. *Cad. Saúde Pública*, Rio de Janeiro v. 6, n. 12, 1990.

DÉNIZ-DÉNIZ, M.; ZÁRRAGA-OBERTY, C. The Assessment of the stakeholders' environment in the new age of knowledge: an empirical study of the influence of the organisational structure. *Business Ethics: A European Review*. v. 13, n. 17, 2004.

FALS BORDA, O.; MORA-OSEJO, L. E. La superación del Eurocentrismo. Enriquecimiento del saber sistémico y endógeno sobre nuestro contexto tropical. . *Polis, Revista de la Universidad Bolivariana*. 2 2004.

FERNANDES, F. M. B.; RIBEIRO, J. M.; MOREIRA, M. R. Reflexões sobre avaliação de políticas de saúde no Brasil. Rio de Janeiro: *Cad. Saúde Pública*. v. 27, n. 11 p. 2011.

FLEXNER, A. Medical Education in United States and Canada: report to carnegie foundation for advancement of teaching. New York: *Merrymount Press* 1910.

FREEMAN, R. *Strategic management: a stakeholder approach*. Boston: Pitman, 1984.

GATTINARA, B. C. *et al.* Percepción de la comunidad acerca de la calidad de los servicios de salud públicos en los distritos Norte e Ichilo, Bolívia. *Cad. Saúde Pública*, v. 11, n. 14, 1995.

GUBA, E. G.; LINCOLN, Y. S. *Avaliação de quarta geração*. São Paulo: Editora Unicamp, 2011. Tradução HONORATO, B.

HAIR, J. F.; WILLIAM, B.; BABIN, B.; ANDERSON, R. E. *Análise multivariada de dados*. 6ª ed. Porto Alegre: 2009.

HART, S.; SHARMA, S. Engaging fringe stakeholders for competitive imagination. *Academy of Management Executive*, v. 18, n. 12, 2004.

HARTZ, S.; JOHN, J. Contribution of economic evaluation to decision making in early phases of product development: a methodological and empirical review *International Journal of Technology Assessment in Health Care*, v.24, n. 4, p. 9, 2008.

HARTZ , Z. M. A.; VIEIRA-DA-SILVA, L. M. *Avaliação em saúde*. EDUFBA. Salvador, 2005.

HURST, J. Challenges for health systems in Member Countries of the Organisation for Economic Co-operation and Development. *Bull World Health Organ*, 78, p. 10, 2000.

MACHADO, J. P.; MARTINS, A. C. M.; MARTINS, M. S. Avaliação da qualidade do cuidado hospitalar no Brasil: uma revisão sistemática. *Cad. Saúde Pública*, Rio de Janeiro, v. 29, n. 20, 2013.

MELLAHI, K.; WOOD, J. The role and potential of stakeholders in “holow participation”: conventional stakeholder theory and institutionalism alternatives. *Business and Society Review*, v.108, n. 2, p. 20, 2003.

MOITA, G. F. *Avaliação integrativa de performance multidimensional e decisão multicritério: um proxy de painel de indicadores de eficiência, efetividade e qualidade para governação de organizações hospitalares e serviços de saúde no Brasil*. 2019. Tese (Doutorado). Orientador: Raposo, V. M. D. R. e Barbosa, A. C. Q – Faculdade de Economia, Universidade de Coimbra, Coimbra, 2019.

MOITA, G. F.; RAPOSO, V. M. R.; BARBOSA, A. C. Q., 2019a, Madrid. *The frontier of the evaluation of results of the Unified Health System: Reliability and validity of the evaluation items for the measurement of health performance in Brazil*. Actas Gobernando el Futuro: agenda 2030. 28.

MOITA, G. F.; RAPOSO, V. M. R.; BARBOSA, A. C. Q. Validação colaborativa de macrodimensões e indicadores-chave para avaliação de performance de serviços de saúde no Brasil. *Rev. Saúde em Debate*, 43, n. 5, p. 232-247, 2019b.

MORRA IMAS, L. G.; RIST, R. C. *The road to results designing and conducting effective development evaluations*. World Bank. Washington, 2009. 611p.

OECD. Performance measurement and performance management in OECD health systems. p. 1-60. 2001. (47).

OECD. *Health at a Glance 2017: OECD Indicators*. Paris, 2017.

PAPANICOLAS, I.; SMITH, P. C. EuroREACH Framework for Health System Performance. 2010.

RAPOSO, V. *Governança hospitalar - uma proposta conceptual e metodológica para o caso português*. Tese (Doutorado). Orientador: Ferreira, P. A. M. L Faculdade de Economia, Universidade de Coimbra, Coimbra, 2007.

ROSALEM, V. *Análise das percepções dos principais atores da cadeia produtiva da saúde sobre a qualidade dos serviços prestados por hospitais no Estado de Goiás - Brasil*. 2013. Tese (Doutorado) – Escola de Administração de Empresas de São Paulo, São Paulo, 2013.

ROWAN, M. S. Logic models in primary care reform navigating the evaluation. *Canadian Journal of Program Evaluation*, v.15, n. 2, p. 12, 2000.

SAMICO, I.; FELISBERTO, E.; FIGUEIRÓ, A. C.; FRIAS, P. G. *Avaliação em saúde: bases conceituais e operacionais*. Rio de Janeiro: Medbook, 2010.

SAMPIERI, R. H.; COLLADO, C. F.; LÚCIO, P. B. *Metodologia de pesquisa*. São Paulo: McGraw Hill, 2006.

SANTOS, M. P. Avaliação da qualidade dos serviços públicos de atenção à saúde da criança sob a ótica do usuário. *Revista Brasileira de Enfermagem*, v. 48, n. 11, 1995.

SOUZA, L. E. P. F.; VIEIRA-DA-SILVA, L. M.; HARTZ, Z. M. A. *Avaliação em saúde: dos modelos teóricos à prática na avaliação de programas e sistemas de saúde*. In: Conferência de consenso sobre a imagem-objetivo da descentralização da atenção à saúde no Brasil. Rio de Janeiro: Editora Fiocruz, 2005.

TANAKA, O. Y.; TAMAKI, E. M. O papel da avaliação para a tomada de decisão na gestão de serviços de saúde. *Ciênc. Saúde Coletiva*, v. 17, n. 8, 2012.

TRAVASSOS, C.; NORONHA, J. C.; MARTINS, M. *Mortalidade hospitalar como indicador de qualidade: uma revisão*. *Ciência & Saúde Coletiva*, v.4, n. 2, 1999.

UCHIMURA, K. Y.; BOSI, M. L. M. Qualidade e subjetividade na avaliação de programas e serviços em saúde. *Cad. Saúde Pública*, v. 18, n. 6, 2002.

UNDP. *Handbook on Planning, Monitoring and Evaluation for Development Results*. 2009.

UNDP. *Handbook on Planning, Monitoring and Evaluation for Development Results*. 2009.

VEILLARD, J. *et al.* A performance assessment framework for hospitals: the WHO regional office for Europe PATH Project. *International Journal for Quality in Health Care*, v. 17, n. 6, p. 9, 2005.

VIACAIVA, F. *et al.* Uma metodologia de avaliação do desempenho do sistema de saúde brasileiro. *Ciência & Saúde Coletiva*, v. 9, n. 3, 2004.

WHO, W. H. O. Bridging the "know-do" gap: meeting on knowledge translation in global health. WHO. Geneva. 2006.

Galba Freire Moita

 ORCID: <https://orcid.org/0000-0003-4959-9424>

Tecnologista em Monitoramento e Avaliação do Ministério da Saúde (MS). Pesquisador convidado da Fundação Oswaldo Cruz (Fiocruz-Ceará). Pesquisador associado do Centre for Business and Economics Research (CeBER) da Faculdade de Economia da Universidade de Coimbra (UC). Mestrado em Saúde Pública (ENSP/Fiocruz). Doutor em Gestão - Ciência Aplicada à Decisão (Universidade de Coimbra). Pós-doutor em Saúde Internacional e Bioestatística (Instituto de Higiene e Medicina Tropical da Universidade Nova de Lisboa).

E-mail: galba.moita@saude.gov.br

Vítor Manuel dos Reis Raposo

 ORCID: <https://orcid.org/0000-0002-9328-8415>

Doutor em Organização e Gestão de Empresas, na especialidade de Ciências dos Sistemas nas Organizações. Professor Auxiliar da Faculdade de Economia, Universidade de Coimbra. Coordenador do Mestrado em Gestão e Economia da Saúde, da Pós-graduação em Economia e Gestão em Organizações de Saúde, e do módulo Health Care Entrepreneurship financiado pelo European Institute of Technology (EIT). Investigador do Centre for Business and Economics Research (CeBER) e do Centro de Estudos e Investigação em Saúde da Universidade de Coimbra (CEISUC). Investigador e colaborador do Observatório Português dos Sistemas de Saúde (OPSS).

E-mail: vraposo@fe.uc.pt

Allan Claudius Queiroz Barbosa

 ORCID: <https://orcid.org/0000-0003-1266-5168>

Professor Titular da Universidade Federal de Minas Gerais (UFMG), desde 1993. Professor Residente do Instituto de Estudos Avançados Transdisciplinares (IEAT) da UFMG (2019/2020). Coordenador do Observatório de Recursos Humanos em Saúde da FACE/UFMG e Membro do Comitê Coordenador da Rede de Pesquisa em Atenção Primária à Saúde, é Bolsista de Produtividade em Pesquisa do CNPq.

E-mail: allan@ufmg.br

Anexo 1 – Quadro 4 | Distribuição dos indicadores versus quadro teórico-conceitual (framework) selecionados (Brasil e exterior)

Indicadores	ACHS	CND	HP	NHS	RALC	PROAHSA/CQH	SPG
% admissões de crianças c/ infecções respiratórias				NHS			
% admissões em relação ao total de casos atendidos				NHS			
% de altas acordadas c/ normas técnicas int. e ext.					RALC		
% de casos em que todos os médicos estão disponíveis					RALC		
% de cirurgias suspensas				NHS		PROAHSA/CQH	
% de gastos c/ medicam. genéricos em relação ao total					RALC		
% de injúrias despercebidas na admissão	ACHS						
% de investimento em atenção primária					RALC		
% obesidade em adultos	ACHS	HP					
% obesidade em criança e adolescentes	ACHS	HP					
% de pacientes que esperam menos de 2h p / admissão				NHS			
% de prof. que são orientados de acordo c/ as metas					RALC		
% dos gastos c/ programas materno-infantil					RALC		
% dos gastos financiados por doação					RALC		

% dos gastos que é destinado à prevenção				RALC		
% screening p/ anomalias do c. do útero em 24 meses	ACHS					
Absenteísmo					PROAHSA/CQH	SPG
Acesso um clínico geral			NHS			
Acidentes de trabalho					PROAHSA/CQH	SPG
Administração dos cuidados primários			NHS			
Avaliação da qualificação dos profissionais de saúde				RALC		
Comparação desempenho obtido e o planeado				RALC		
Concepções em menores de 18 anos			NHS			
Consultas perdidas com pacientes externos			NHS			
Consumo de energia elétrica, água, GLP, oxigênio					PROAHSA/CQH	
Consumo de materiais por grupo					PROAHSA/CQH	
Custo de internação ajustada por Casemix	ACHS					
Custo do leito-dia				RALC		
Custos de pessoal				RALC	PROAHSA/CQH	
Custos gerais e indiretos					PROAHSA/CQH	
Custos globais					PROAHSA/CQH	
Custos mensais de horas extras					PROAHSA/CQH	
Demora no atendimento em casos de trauma	ACHS					
Dentes perdidos e cariados em crianças			NHS			

Dias de espera para admissão de cirurgias eletivas	ACHS					
Doação de órgãos				NHS		
Existência de programas de incentivo para RH					RALC	
Exposição a fumantes	ACHS	CND				
Frequência de consumo de álcool		CND	HP			
Horas de treinamento por funcionário	ACHS					PROAHSA/CQH
Índice de endividamento						PROAHSA/CQH
Índice de infecção hospitalar	ACHS					PROAHSA/CQH
Índice de infecção hospitalar em cirurgias limpas						
Intervalo de substituição						PROAHSA/CQH
Margem Líquida						PROAHSA/CQH
Margem Operacional						PROAHSA/CQH
Média de Permanência	ACHS				RALC	PROAHSA/CQH
Média de permanência em obstetrícia						
Média de permanência pediátrica						
Mortalidade por trauma	ACHS					
Número de admissões emergenciais				NHS		
Número de adolescentes fumantes			HP			
Número de adultos em tratamento mental			HP	NHS		
Número de adultos fumantes			HP			

Número de atendimentos de casos-câncer cervical			NHS		
Número de atendimentos de casos-câncer de mama			NHS		
Número de atendimentos de urgência/emergência					PROAHSA/CQH
Número de atrasos nas altas			NHS		
Número de casos de cirurgias desnecessárias			NHS		
Número de cirurgias realizadas					PROAHSA/CQH
Número de consultas ambulatoriais					PROAHSA/CQH
Número de consultas por enfermeira				RALC	
Número de consultas por médico				RALC	
Número de detecção precoce de câncer			NHS		
Número de enfermeiras por leito				RALC	
Número de enfermeiras por médico				RALC	
Número de exames de diagnóstico por imagem					PROAHSA/CQH
Número de exames laboratoriais					PROAHSA/CQH
Número de internações que geraram eventos adversos	ACHS				
Número de leitos-dia					PROAHSA/CQH
Número de médicos por leito				RALC	

Número de necropsias em óbitos ocorridos no hospital					PROAHSA/CQH
Número de óbitos de pacientes portadores do HIV		CND			
Número de óbitos operatórios			NHS		PROAHSA/CQH
Número de óbitos por injúrias não intencionais		CDN			
Número de pacientes acima de 12 anos c/ depressão		CDN			
Número de pacientes c/ câncer c/ 2 sem. de espera			NHS		
Número de pacientes com pressão alta		CDN	NHS		
Número de pacientes com quatro semanas sem fumar			NHS		
Número de pacientes externos com 13 sem. de espera			NHS		
Número de pacientes-dia			NHS		PROAHSA/CQH
Número de parturientes que realizam o pré-natal			HP		
Número de prescrições de antibiótico	ACHS		NHS		
Número de procedimentos especializados					PROAHSA/CQH
Número de profissionais de nível superior em saúde		CDN	NHS		PROAHSA/CQH
Número de readmissões após fratura de bacia			NHS		
Número de readmissões emergenciais pós-alta	ACHS	CDN	NHS		

Número de readmissões psiquiátricas			NHS	
Número de refeições servidas				PROAHSA/CQH
Número de saídas				PROAHSA/CQH
Número de sobreviventes com câncer de seio			NHS	
Número de sobreviventes de câncer de <i>cólon</i>			NHS	
Número de sobreviventes de câncer de pulmão			NHS	
Número mensal de leitos em funcionamento				PROAHSA/CQH
Óbitos dentro de um período de 30 por AVC			NHS	
Óbitos num período até 30 após cirurgia cardíaca			NHS	
Óbitos num período até trinta 30 pós-cirurgia eletiva			NHS	
Pontualidade no pagamento dos funcionários				RALC
Prescrição de drogas para úlceras cicatrizadas			NHS	
Prescrição de medicam. de conhecido efeito alérgico	ACHS			RALC
Profissionais da saúde com título de especialista				RALC
Proporção de profissionais com qualificações básicas			NHS	PROAHSA/CQH
Qualidade dos dados				
Quilograma (média mensal) de roupa lavada				PROAHSA/CQH
Receita Líquida por funcionário				PROAHSA/CQH

Relação entre o custo da consulta e o do profissional					RALC		
Retorno do investimento						PROAHSA/CQH	
Retorno para a sala cirúrgica na mesma internação	ACHS						
Rotatividade dos leitos						PROAHSA/CQH	
Satisfação do usuário				NHS	RALC		SPG
Soluções de problemas com informações disponíveis					RALC		
Taxa de abandono de tratamento					RALC		
Taxa de cesárea	ACHS	CDN				PROAHSA/CQH	SPG
Taxa de cirurgias de reimplante				NHS			
Taxa de cirurgias eletivas				NHS			
Taxa de cirurgias para doenças coronarianas				NHS			
Taxa de cirurgias para remoção de catarata				NHS			
Taxa de cirurgias para substituição de articulações				NHS			
Taxa de embolia pulmonar pós-operatória	ACHS						
Taxa de internação por asma	ACHS						
Taxa de internação por câncer	ACHS						
Taxa de mortalidade		CND					SPG
Taxa de mortalidade infantil							SPG
Taxa de mortalidade obstétrica							SPG

Taxa de mortalidade por câncer		CND				
Taxa de mortalidade por patologias do ap. circulatório		CND				
Taxa de mortalidade por patologias do ap. respiratório		CND				
Taxa de ocupação				RALC	PROAHSA/CQH	SPG
Tempo de espera para serviços de emergência	ACHS					
<i>Turnover</i>						SPG

Fonte: adaptado wwde (ROSALEM, 2013)