



Saúde em Debate

ISSN: 0103-1104

ISSN: 2358-2898

Centro Brasileiro de Estudos de Saúde

Moita, Galba Freire; Raposo, Vítor Manuel dos Reis; Barbosa, Allan Cláudius Queiroz
Validação colaborativa de macrodimensões e indicadores-chave
para avaliação de performance de serviços de saúde no Brasil
Saúde em Debate, vol. 43, núm. 5, Esp., 2019, pp. 232-247
Centro Brasileiro de Estudos de Saúde

DOI: <https://doi.org/10.1590/0103-11042019S519>

Disponível em: <https://www.redalyc.org/articulo.oa?id=406369096020>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em [redalyc.org](https://www.redalyc.org)

Validação colaborativa de macrodimensões e indicadores-chave para avaliação de performance de serviços de saúde no Brasil

Collaborative validation of macrodimensions and key indicators for health services performance evaluation in Brazil

Galba Freire Moita^{1,2,3}, Vítor Manuel dos Reis Raposo⁴, Allan Claudius Queiroz Barbosa⁵

DOI: 10.1590/0103-110420195519

RESUMO A criação do Sistema Único de Saúde (SUS) no Brasil, em 1988, representou avanços na organização sistêmica e descentralização da gestão única; entretanto, passados 30 anos a governança de resultados parece frágil. A nova gestão pública tem exigido esforços de monitoramento de resultados, controladoria e responsabilização dos gastos (*accountability*). Este estudo explora a translação de conhecimentos de uma amostra de gestores e profissionais (*stakeholders*), para validação de um painel de indicadores do SUS. A aplicação dos instrumentos de captação e validação das percepções obteve resultados das três fases iniciais ($n=108$) que consolidaram um instrumento aplicado para validação de campo ($n=112$), cuja análise descritiva validou cinco dimensões e 24 indicadores-chave para gestão de resultados em organizações de saúde. A análise inferencial gerou um modelo final que garantiu confiabilidade e validade das cinco dimensões (macrodomínios), mas apenas de 17 indicadores (domínios) de desempenho propostos pelos decisores a partir de seus conhecimentos prévios.

¹Ministério da Saúde (MS)
- Brasília (DF), Brasil.
pргalba@gmail.com

²Fundação Oswaldo Cruz
(Fiocruz) - Fortaleza (CE),
Brasil.

³Universidade Nova de
Lisboa, Instituto de Higiene
e Medicina Tropical -
Lisboa, Portugal.

⁴Universidade de Coimbra,
Faculdade de Economia,
Centre for Business and
Economics Research
(CeBER) e Centro de
Estudos e Investigação em
Saúde da Universidade
de Coimbra (Ceisuc) -
Coimbra, Portugal.

⁵Universidade Federal de
Minas Gerais (UFMG),
Faculdade de Ciências
Econômicas, Observatório
de Recursos Humanos
em Saúde (Observa-RH)
- Belo Horizonte (MG),
Brasil.

PALAVRAS-CHAVE Governança. Avaliação de processos e resultados (cuidado de saúde). Avaliação em saúde. Indicadores básicos de saúde. Monitoramento.

ABSTRACT The creation of the Unified Health System (SUS) in Brazil, in 1988, represented advances in the systemic organization and decentralization of the unified management; however, after 30 years the governance of results seems fragile. The new public management has demanded efforts to monitor results, controllership and accountability. This study explores the translation of knowledge from a sample of managers and professionals (*stakeholders*), for validation of a panel of SUS indicators. The application of perceptual capture and validation instruments yielded results from the three initial phases ($n=108$), which consolidated an instrument validated for field validation ($n=112$), whose descriptive analysis validated five dimensions and 24 key indicators for management of results in health organizations. Inferential analysis generated a final model that guaranteed reliability and validity of the five dimensions (macrodomains), but only of 17 performance indicators (domains) proposed by the decision makers based on their previous knowledge.

KEYWORDS Governance. Process and results evaluation (health care). Health assessment. Basic health indicators. Monitoring.

Introdução

As demandas da sociedade em relação à saúde têm sido crescentes, apesar do sistema de saúde do Brasil ter apresentado uma evolução importante em estrutura e organização. Os principais avanços estão relacionados às garantias legais¹⁻³ de atendimento, da organização sistêmica e da descentralização da gestão da saúde para os níveis estadual e municipal; porém, poucos avanços são observados na governança das organizações e unidades de saúde. Por sua vez, a importância do monitoramento e da avaliação se reafirma pelos fatores econômicos, de necessidade de acesso, da exigência da qualidade da atenção à saúde, em busca da eficiência, efetividade e satisfação dos usuários do Sistema Único de Saúde (SUS).

Em busca dos primórdios do processo de avaliação em saúde Dos Reis et al.⁴ sintetizaram que a avaliação de resultados na área de saúde, talvez por sua complexidade, é uma iniciativa relativamente recente. Lembram do relatório de Flexner⁵, que sob o patrocínio da Fundação Carnegie propôs uma avaliação do ensino e sistematização da prática médica, e, do relatório Codman⁶ nos quais identificam os primeiros sinais da preocupação da qualidade e da *performance* das organizações de saúde e a criação do embrião da Joint Commission of Accreditation of Hospitals (JCAH), em 1928, uma das principais agências mundiais de fomento à qualidade no setor de saúde.

Por seu turno, o aumento indiscriminado de custos da área médica e a necessidade do controle da qualidade da assistência médica pressionam os sistemas de saúde para uma adequada gestão dos recursos, conforme Perez Arias⁷⁽⁵¹⁾. Estes autores ainda argumentam que

estas são razões suficientes para se buscar o controle dos custos e da qualidade a atenção médica, e que a tendência ao aumento dos custos do setor saúde é, na verdade, universal.

Em linha similar, Alkin⁸ apud Samico et al.⁹ estabelece as origens do campo da avaliação a

partir da necessidade de prestação de contas (*accountability*) e controle dos programas.

Segundo Samico et al.⁹ o campo de avaliação em saúde, engloba uma diversidade de termo, conceitos e métodos conforme sua heterogeneidade, complexidade e subjetividade das intervenções, sejam elas ações, serviços, programas ou políticas públicas, e por isso é um dilema a seleção de um modelo ou abordagem de avaliação¹⁰.

A partir da proposta de Avendis Donabedian¹¹ – 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. O modelo de Brook & Lohr^{4,12} propõe a avaliação das dimensões: eficácia, efetividade, e níveis de qualidade do cuidado, e as variações das características populacionais. Uchimura e Bosi¹³ referem autores e listas das possíveis dimensões de avaliação: Gattinara et al.¹⁴ sinalizam fatores que determinam a qualidade dos serviços de saúde: eficácia; [...] eficiência [...]. Vuori¹⁵ citado por Acurcio et al.¹⁶, Akerman & Nadanovsky¹⁷, e, Santos,¹⁸ apontam: [...] efetividade; [...] eficácia; [...] eficiência.

Não há na literatura um consenso sobre escopo e amplitude da avaliação e do monitoramento em saúde, mas muitos possíveis quadros teóricos conceituais¹⁹⁻²². Arah et al.²⁰ apresentam uma comparação global das diversas abordagens de quadro teórico-conceitual (*framework*) dos sistemas de medição de *performance* do Reino Unido (NHS), do Canadá, da Austrália, dos EUA, do relatório World Health Report 2000²³, além das publicações ‘Health at a Glance e OECD Health Data’ da Organisation for Economic Co-operation and Development (OECD)^{24,25}. No Brasil, no escopo deste trabalhou voltado à uma visão integrada, defendida pela United Nations Development Programme (UNDP), para a tomada de decisão e gestão baseada em resultados²⁶ destacam-se algumas propostas como por exemplo os projetos: Projeto Avaliação do

Desempenho do Sistema Saúde (Proadess) e Índice de Desempenho do Sistema Único de Saúde (IDSUS)^{27,28}.

Em busca de superar tais desafios e construir consensos, Guba e Lincoln²⁹ propuseram a avaliação de quarta geração, pautada em um processo construtivista de negociação entre interessados na avaliação. Neste escopo, considerou-se como decisores-chave (*key stakeholders*) os gestores de saúde dos níveis governamentais/sistêmicos, organizacionais e chefias setoriais, envolvidos em um processo colaborativo, a fim de mapear os seus propósitos em torno da adequabilidade à realidade e utilidade percebida, para a validação de uma matriz de dimensões e de indicadores para avaliação em unidades de saúde.

Diante da necessidade de prospectar e validar as macrodimensões e indicadores-chave de mensuração de desempenho de organizações de saúde, pautou-se na teoria dos *stakeholders*, em especial, nas suas perspectivas 'descritiva e instrumental' destacadas pela literatura^{30,31} em que os grupos de *stakeholders* podem efetivamente exercer um impacto considerável nos resultados organizacionais e sobre o seu desempenho e competitividade. Dessa forma, este estudo considerou o aspecto instrumental do monitoramento de resultados definido no âmbito da OECD, que tem o monitoramento para prover aos decisores (*stakeholders*) parâmetros para uma intervenção, por meio de indicadores específicos da extensão do progresso, do alcance dos objetivos e dos resultados obtidos na utilização de recursos alocados^{32,33}. Assim, definiu-se como objetivo elaborar uma proposta metodológica para envolver gestores e profissionais na seleção e validação de dimensões e indicadores-chave para a mensuração de desempenho de organizações e serviços de saúde no contexto do SUS. Diante disso, surge a questão central deste estudo: qual a relevância/importância relativa atribuída aos indicadores e dimensões de desempenho, em uma amostra multicêntrica de profissionais e gestores de saúde pública, através de análises descritiva, confiabilidade e

validade, que permita a validação de um *proxy* de painel de indicadores e dimensões para a governança de unidades do SUS?

No presente estudo, apropriou-se de um processo colaborativo visando prospectar, sintetizar e aplicar conhecimentos visando à melhoria de serviços em prol de preenchimentos de lacunas de saberes e instrumentos para transformar políticas e práticas³⁴⁻³⁶ para a gestão de serviços do SUS. O processo aplicado pode ser expandido para outros serviços públicos, visando estruturar e validar painéis de indicadores específicos.

Métodos

A pesquisa é observacional, transversal, por método misto de avaliação. Apoiou-se em estudo de casos múltiplos³⁷ selecionados por conveniência, em uma amostra de informantes (gestores e profissionais) no âmbito do Ph.D. da Universidade de Coimbra (CAAE: 54972816.9.0000.5051). O universo da investigação são as organizações de saúde, unidades primárias, policlínicas, hospitais e demais unidades da rede do SUS. Estão incluídas na amostra as unidades em que os gestores de topo firmaram o termo de autorização de participação.

Diante da consulta à literatura^{14-18,27,38} 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-Silva³⁹ e de macrodimensões de avaliação de desempenho prevalentes na literatura.

Neste estudo, nos aspectos de avaliação e monitoramento, este modelo lógico³⁹ propõe algumas categoriais de fatores para a avaliação em saúde, no qual se destacam: eficácia, eficiência, efetividade, impacto, qualidade técnico-científica, percepção do usuário (satisfação e aceitabilidade), dentre outros. Tornou-se base de consulta da percepção dos decisores-chave para a estruturação participativa, em fases incrementais: visão global (fase 1, n=11), visão específica (fase 2, n=42), além da visão

específica ajustada (fase 3, n=55) somando 108 respondentes, consolidadas pela aplicação de campo (fase 4, n=112), com análise estatística a partir da ponderação de itens por uma escala Likert de 5 níveis. Ressalta-se que cada item poderia ter diferentes quantidades de respostas, pois os informantes poderiam optar por não responder a totalidade dos itens, mas a comparabilidade foi garantida pela análise das médias ponderadas das respostas obtidas.

A primeira etapa do estudo (validação incremental do instrumento) incluiu as fases 1 a 3, a partir do conhecimento dos gestores e profissionais de saúde, que consolidou um instrumento que foi aplicado na segunda etapa, fase 4 (*figura 1*). Com base nos resultados obtidos, foi realizada a análise descritiva integrada das respostas das fases 1 a 3, seguida pela análise descritiva e inferencial das respostas da fase 4. Em virtude do desalinhamento de termos propostos pelos decisores manteve-se nomenclaturas genéricas para as categoriais de avaliação (macrodomínios = dimensão) e para as subcategorias de avaliação (domínios = indicadores). Os resultados foram equalizados pela literatura vigente e geraram instrumentos contendo indicadores e dimensões de *performance* para serem utilizados em fases posteriores.

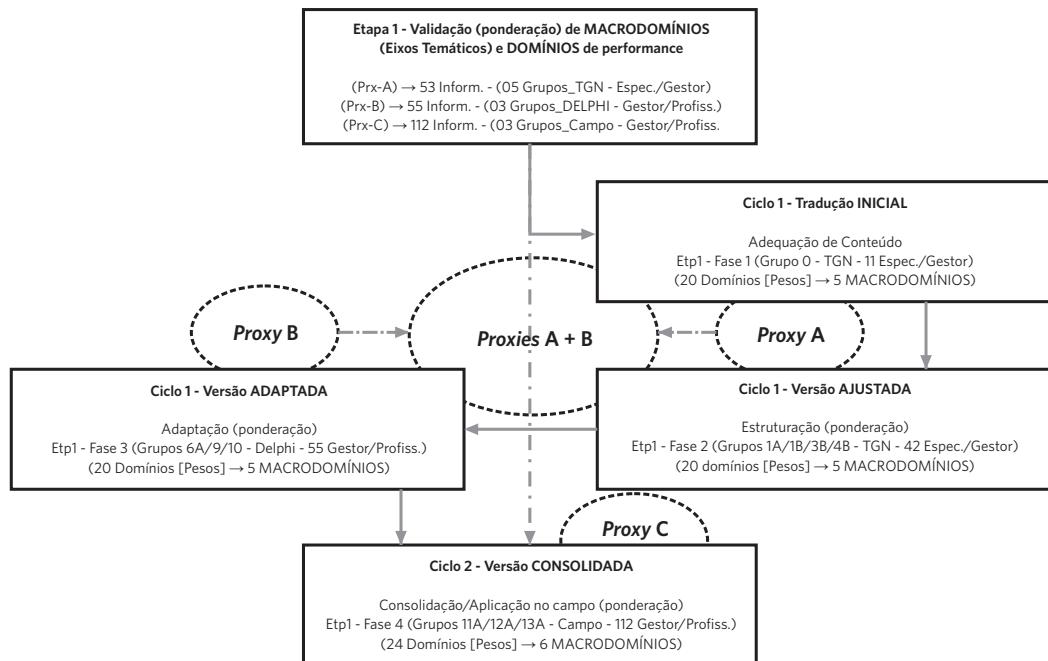
De forma pragmática, mapearam-se quais os decisores-chaves (*key stakeholders*) interessados na avaliação para a tomada de decisão, nos níveis Governamentais/Sistêmicos, Organizacionais e Gerenciais/Setoriais dos serviços de saúde. Na etapa seguinte, foi realizada consulta às preferências dos decisores-chave em duas rodadas: de validação de face e; de valoração de itens. Para efetivar o processo de consultas na estruturação do problema e ordenação das preferências de metodologias e indicadores, utilizaram-se as técnicas qualitativas de painéis de especialistas para proposição e validação de face do conteúdo; seguiu-se a valoração de itens por grupos Delphi de entrevistas semiestruturadas, que

consistiu em buscar o consenso por sucessivas consultas aos grupos de decisores de forma escrita e estruturada³⁷.

Os informantes-chave selecionados foram questionados sobre o grau de relevância e preferência dos macrodomínios (macrodimensões) e domínios (indicadores) de avaliação de desempenho em unidades e serviços de saúde, inclusive pela proposição e ponderação de novos itens de avaliação, segundo o seu grau de conhecimento pessoal, baseados em modelos que emergiram da revisão da literatura. Também foram solicitados a classificar o grau de importância/relevância das macrodimensões e indicadores de *performance*, posteriormente analisados segundo os potenciais variáveis mediadoras selecionadas.

Dessa forma, a aplicação inicial da validação de face abordou um grupo focal de 11 especialistas em avaliação e gestão de saúde (fase 1). A validação de face posterior, pela Técnica de Grupo Nominal (TGN), envolveu quatro grupos pilotos com 42 gestores e profissionais (fase 2) e identificou os domínios (indicadores) e macrodomínios de avaliação (dimensões de desempenho) considerados como itens mais relevantes na mensuração de *performance*, na *proxy A*. 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 elevada (superior ou cerca de 3, relevante/importante) nas fases anteriores (1 e 2), tendo abordado três grupos Delphi com 55 gestores e profissionais (fase 3), na *proxy B*. Por fim, aplicou-se o instrumento ajustado, que emergiu dos resultados obtidos da fase 3 (com domínios e macrodomínios de avaliação de desempenho que receberam valoração elevada, superior ou cerca de 3, relevante/importante), e captou-se respostas de outros três grupos Delphi com 112 gestores e profissionais (fase 4), na *proxy C*. A *figura 1* sintetiza o fluxograma das quatro etapas de aplicação da pesquisa.

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



Nas quatro fases era possível agregar novos itens de avaliação (domínios e/ou macrodomínios) que julgassem relevantes, a fim de consolidar os níveis de conhecimentos sobre os domínios e macrodomínios de avaliação em sistemas e serviços de saúde. Enfatize-se que os resultados que obtiveram níveis de importância elevados (média superior ou cerca de 3,0) nas respostas das fases 1 a 3, pautados exclusivamente nos conhecimentos prévios dos participantes, foram avaliados, categorizados e readequados conforme as macrodimensões de *performance* definidos pela literatura mais relevante^{14-18,27,38} para subsidiar o ajuste de um questionário semiestruturado a ser aplicado e validado na fase de pesquisa de campo (fase 4).

Em cada fase, as respostas foram consolidadas com o *software* Excel 2016, traçados gráficos de valores e aplicadas as análises descritivas e inferenciais com o *software* R (versão 3.3.2), ambos com versões de licenças livres.

Análise estatística descritiva e fatorial de validação do modelo

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). De forma complementar, uma análise para mensurar os níveis de ‘significância estatística’ e a influência potencial de três variáveis mediadoras destacadas (nível de atenção à saúde, nível de complexidade da assistência e nível de cargo/função ocupado) quanto a valoração dos quatro macrodomínios de avaliação (Quantitativos, Qualitativos, Efeitos e Satisfação), tendo ainda a adição do Sistêmico/Estratégico, na fase 4.

A análise final e a modelagem do instrumento definitivo deste estudo partiram das valorações atribuídas aos domínios

e macrodomínios de avaliação por uma amostra de 112 gestores e profissionais, estruturada sobre os resultados obtidos na aplicação de campo (fase 4: *proxy C*). Após a análise descritiva, iniciou-se a análise inferencial, aplicada exclusivamente nas respostas da fase 4 (n=112), que se dividiu em quatro análises de resultados: Estudo das cargas fatoriais dos domínios *versus* macrodomínios de avaliação; Validação por critérios de qualidade e validade dos macrodomínios de avaliação; Análise da qualidade (ajustamento) do modelo final resultante e; Análise das correlações entre as variáveis e suas influências na Performance Global Percebida (PGP).

Para validar os macrodomínios de avaliação estudados, procedentes da *proxy C*, foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al.)⁴⁰, a fim de mensurar as cargas fatoriais de cada um dos 24 domínios de avaliação e suas distribuições sobre os cinco macrodomínios para, posteriormente, construir e mensurar suas contribuições para a formação da variável latente PGP.

A AFC supõe que as variáveis latentes apresentem distribuição normal. Por definição, as variáveis em estudo não apresentaram distribuição normal, mas uma escala ordinal, discreta e limitada (escala Likert). Logo, foram utilizados estimadores robustos para estrutura de covariância na AFC com a estatística de teste reescalada pelo método de Satorra e Bentler⁴¹.

De acordo com Hair et al.⁴⁰ os componentes com cargas fatoriais menores que 0,5 devem ser eliminados, pois ao não contribuírem de forma relevante para formação da variável latente, podem levar a não alcançar as suposições básicas para a validade e qualidade do modelo final, que envolvem a ‘dimensionalidade’, ‘confiabilidade’ e ‘validade convergente’. Também foram excluídas, do modelo em análise, quaisquer variáveis que suas permanências impiediam a convergência da AFC.

Análise de critérios de confiabilidade e validade dos macrodomínios de avaliação

Para analisar a qualidade e validade dos macrodomínios de avaliação, foi verificada a dimensionalidade, confiabilidade e validade convergente. A dimensionalidade foi verificada pelo critério da Kaiser, que retorna a quantidade de dimensões de constructos, baseado na quantidade mínima de variância explicada por cada dimensão. Para a validade discriminante foi utilizado o critério de Fornell e Larcker⁴², que garante a validade discriminante quando a Variância Média Extraída (AVE) de uma variável latente for maior que a Variância Compartilhada (VCM) com as demais. A confiabilidade foi mensurada pelo Alfa de Cronbach (A.C.) e Confiabilidade Composta (C.C.). De acordo com Tenenhaus et al.⁴³ os indicadores A.C. e C.C. devem ser maiores que 0,7 para garantir a confiabilidade, mas em pesquisas exploratórias valores acima de 0,6 são aceitos. Para verificar a validade convergente foi utilizado o critério proposto por de Fornell e Larcker⁴². Ele garante tal validade caso a AVE sejam superiores a 50%⁴⁴, ou 40% no caso de pesquisas exploratórias^{45,46}.

Análise da qualidade (ajustamento) do modelo final da Performance Global Percebida

Após propor um modelo final resultante das análises, verificou-se as medidas de validade e qualidade da PGP, por meio da validação convergente, Alfa de Cronbach ou Confiabilidade Composta, a unidimensionalidade pelo critério de Kaiser e a validação discriminante pelo critério de Fornell e Larcker⁴². Também se aplicou testes para medir a qualidade (ajustamento)

do modelo construído para os macrodomínios e domínios de avaliação da *proxy C*, com base em índices selecionados na literatura relevante ($\chi^2/G.L.$, CFI, TLI e RMSEA).

De forma complementar, procedeu-se a análise das correlações entre as variáveis estudadas e suas possíveis influências na variável latente PGP. Dessa forma, concluiu-se a análise inferencial das respostas da fase 4, *proxy C* (112 informantes).

Explorou-se, ainda, uma análise descritiva da variável PGP, inclusive quanto às possíveis influências das variáveis nível de atenção à saúde e nível de cargo/função.

Resultados

Nas fases 1 e 2 destacam-se os resultados quanto ao conhecimento global dos respondentes sobre possíveis macrodimensões e indicadores de desempenho em unidades de saúde, que por terem nível reduzido de alinhamento com as nomenclaturas da literatura corrente preferiu-se denominar de domínios e macrodomínios de avaliação, mantendo-se a contribuição dos informantes-chave. Em análise prévia, notou-se um razoável alinhamento nos conteúdos e nos níveis de valoração nas respostas obtidas nas fases iniciais (1 a 3), que permitiu a análise integrada das respostas obtidas nas fases 1 a 3 ($n=108$) comparativamente à fase 4 ($n=112$).

Quanto aos respondentes tinham níveis elevados em ambas as amostras, pois 36,11% exercia um dos três níveis de gestão (Sistêmico/ Governacional, Organizacional/ Estratégico ou Gerência/ Chefia setorial) nas fases 1 a 3, porém 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); mas, na fase 4, a maior parte (41,96%) exercia cargo Sistêmico/Governacional. Em todas as fases, a ampla maioria dos indivíduos possuía nível superior, no mínimo. 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.

Análise descritiva (valoração média de domínios e macrodomínios)

A *tabela 1* apresenta a valoração média dos domínios de avaliação da amostra fases 1 a 3 e fase 4. A análise global permite afirmar que a ampla maioria dos 20 domínios de avaliação analisada obteve valoração média elevada (superior ou cerca de 4,0). Dessa forma, todos os 20 domínios de avaliação foram considerados relevantes/importantes para análise nas próximas fases de investigação, na estruturação de um novo instrumento com domínios e macrodomínios de avaliação devidamente alinhados pela literatura relevante de dimensões e indicadores de *performance*.

Não houve diferença significativa entre o grau de importância dos macrodomínios de avaliação estudados, uma vez que todos os intervalos de confiança das valorações médias obtidas para os macrodomínios se sobrepõem. Dessa forma, apesar da variabilidade observada, não houve significância estatística para considerar os níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos respondentes enquanto variáveis mediadoras na ampla maioria dos quatro macrodomínios, inclusive no quinto macrodomínio (Sistêmico/Estratégico), acrescido na fase 4.

Ressalta-se que os respondentes acrescentaram quatro novos domínios de avaliação (D5W_IndICSAP, D5X_AtendVincReg, D5Y_AtendDesVincReg e D5V_ExtrapTetoOrç), adicionados na fase 3 em um novo macrodomínio de avaliação (Sistêmico/Estratégico), que não tiveram respostas individualizadas nesta análise (foram distribuídos nos domínios Outros), mas foram incorporados na nova fase de investigação (fase 4) de consolidação do instrumento.

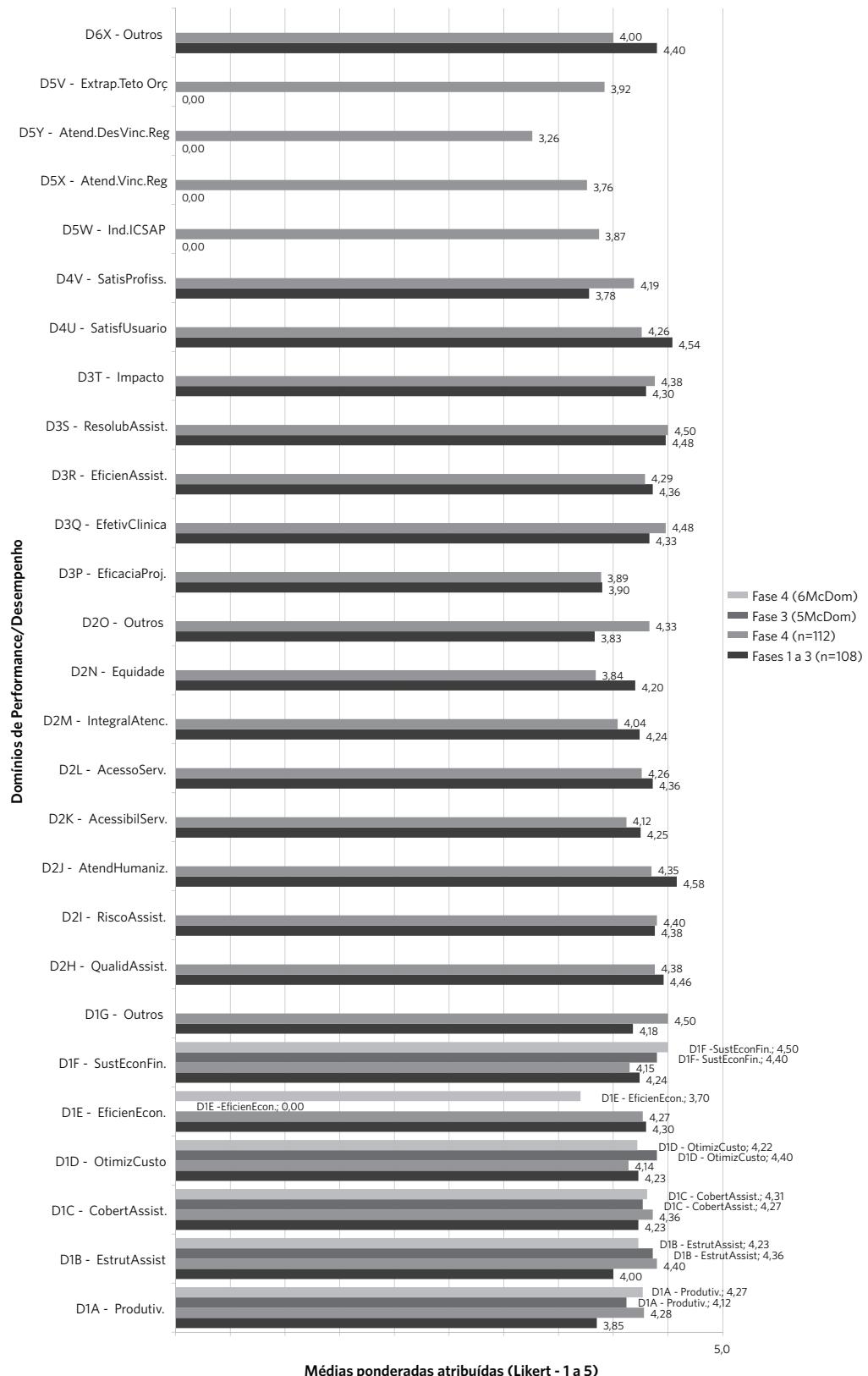
Tabela 1. Análise descritiva dos domínios de avaliação - Proxy A+B (ciclo 1) e proxy C (ciclo 2) (estudo 1) (Visão de especialistas)

Macrodomínios	Domínios	Proxy A+B (fases 1 a 3)				sProxy C (fase 4)			
		N	Média	D.P.	I.C-95%	N	Média	D.P.	I.C-95%
Quantitativos	D1A_Produtiv	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_CobertAssit	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]
Efeitos	D2O_Outros	6	3,83	0,98	[3,17; 4,50]	3	4,33	0,58	[4,00; 5,00]
	D3P_EficaciaProj	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]
Satisfação	D3T_Impacto	94	4,30	0,80	[4,14; 4,45]	111	4,38	0,74	[4,24; 4,52]
	D4U_SatisfUsario	95	4,54	0,70	[4,40; 4,67]	111	4,26	0,84	[4,11; 4,41]
	D4V_SatisfProfiss	37	3,78	1,60	[3,24; 4,24]	112	4,19	0,89	[4,02; 4,35]
Sistêmico/ Estratégico	D5W_IndICSA					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]

Na etapa final (fase 4) destacam-se os resultados de campo quanto ao conhecimento dos respondentes sobre possíveis macrodimensões e indicadores de desempenho em unidades de saúde, mantidas as nomenclaturas de macrodomínios e domínios de avaliação,

respectivamente. A ampla maioria dos 24 domínios obteve valoração média elevada. Dessa forma, todos estes 24 domínios de avaliação foram considerados relevantes/importantes (*tabela 1*) para análise nas próximas fases de investigação (*figura 2*).

Figura 2. Médias de relevância/importância (Likert - 1 a 5)



Em análise complementar, sobre potenciais fatores mediadores das valorações obtidas, pode-se sintetizar que não houve significância estatística ($p<0,05$) para considerar o nível de atenção à saúde enquanto variável mediadora em três macrodomínios de avaliação (Qualitativos, Satisfação e Sistêmico/Estratégico), exceto para os macrodomínios Quantitativos ($p=0,011$) e Efeitos ($p=0,040$). Por sua vez, a variável nível de cargo/função não pode ser considerada variável mediadora na ampla maioria destes macrodomínios de avaliação estudados, exceto para o macrodomínio Quantitativos ($p<0,05$), na fase 4 (*proxy C*). Outras análises revelaram que não houve significância estatística para considerar o nível de complexidade da assistência como variável mediadora em nenhum dos cinco macrodomínios de avaliação, não sendo relevante sua apresentação.

Análise inferencial (confiabilidade e validade de macrodomínios)

Por sua vez, AFC da valoração média obtida para os macrodomínios de avaliação na *proxy C* (fase 4) destaca que no modelo final, após análise e exclusão de sete domínios de avaliação (D1A_Produtiv, D1B_EstrutAssist, D1C_CobertAssit,

D2H_QualidAssist, D2M_IntegralAtenc, D2N_Equidade e D3P_EficaciaProj), todos os demais 17 domínios que permaneceram na análise, apresentaram carga fatorial superior a 0,5 (ou suas permanências não impediram a convergência da AFC) (*figura 3*).

As medidas de confiabilidade, e validade dos macrodomínios de avaliação na *proxy C* (fase 4) demonstraram que a validade dos macrodomínios foram asseguradas, uma vez que todos apresentaram validação convergente (Variância Média Extraída – AVE $> 0,4$), confiabilidade adequada (Alfa de Cronbach - A.C. $> 0,0$ ou Confiabilidade Composta - C.C. $> 0,6$), unidimensionalidade e validação discriminante (Variância Compartilhada Máxima – VCM $< \text{AVE}$) (*tabela 2*).

Quanto à AFC para a variável latente PGP, na *proxy C* (fase 4) apresentou os seguintes resultados: O macrodomínio de maior peso foi Satisfação (27%) e o de menor peso foi Quantitativa (15%), sendo que esse apresentou carga fatorial igual a 0,48. Todos os cinco macrodomínios (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico) permaneceram na análise por não impedir a convergência do modelo.

Tabela 2. Validação por critérios de qualidade e validade dos macrodomínios de avaliação – *Proxy C* (fase 4) (Visão de gestores e profissionais)

Macrodomínios	Domínios	A.C ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
Quantitativos	3	0,76	0,76	1	0,52	0,13
Qualitativos	4	0,71	0,73	1	0,41	0,24
Efeitos	5	0,75	0,75	1	0,44	0,24
Satisfação	2	0,73	0,74	1	0,59	0,24
Sistêmico/Estratégia	4	0,72	0,75	1	0,44	0,14

¹ Alfa de Cronbach; ² Confiabilidade Composta; ³ Dimensionalidade; ⁴ Variância Extraída; ⁵ Variância Máxima Compartilhada.

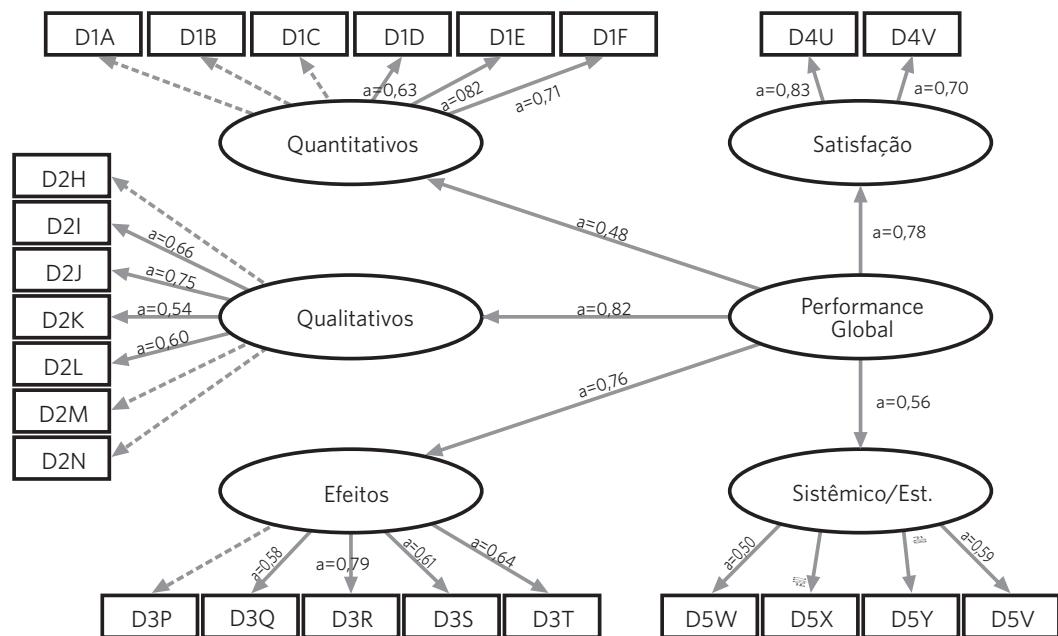
Há uma diversidade de parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais⁴⁸. Os parâmetros da qualidade selecionados do modelo foram: $\chi^2/G.L.$, CFI, TLI e RMSEA, além do valor-p para verificar se o RMSEA foi estatisticamente inferior a 0,05. Para um bom ajuste espera-se que $\chi^2/G.L.$ seja menor que 3^{49,50}, CFI seja maior que 0,8⁵¹, TLI seja maior que 0,8⁵² e que o RMSEA seja menor que 0,1, sendo o ideal abaixo de 0,05⁵³. Destaca-se que o ajuste da AFC foi considerado adequado, 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 0,8 e o RMSEA do foi menor que o limite máximo de 0,1.

Por fim, a comparação da PGP em relação às variáveis de caracterização níveis de atenção à saúde e de cargo/função selecionadas – *Proxy C*, enquanto possíveis variáveis mediadoras da variável latente PGP (média 4,16 e mediana

4,18); 75% das respostas atribuídas resultaram em valoração superior a 3,88 (cerca de 4, muito importante/relevante) para a PGP média; o nível Primário de atenção apresentou a maior variabilidade (0,34) na PGP e a menor valoração média (3,88), enquanto o nível Outros obteve a maior valoração média (4,28). O nível de cargo/função Outros apresentou a maior variabilidade (0,12) na PGP enquanto o nível Organizacional obteve a menor valoração média (4,01) e o nível Sistêmico a maior valoração média (4,25). Não houve diferença significativa (valor-p=0,221) da PGP entre os níveis de atenção à saúde e os níveis do cargo/função (0,332). Não houve diferença significativa (valor-p<0,05) para considerar as variáveis nível de atenção à saúde ou de cargo/função enquanto fatores mediadores da PGP, na *proxy C* (fase 4).

Diante destes resultados, a figura 3 sintetiza o ajuste da AFC e a modelagem para a variável latente PGP.

Figura 3. Análise Fatorial Confirmatória da Performance Global Percebida – *Proxy C* (fase 4) (Visão de gestores e profissionais)



Em análise complementar mostrou-se que os domínios de avaliação apresentaram maiores correlações com seus respectivos macrodomínios, na *proxy* C (fase 4).

Discussão

O processo participativo de validação de indicadores e dimensões de *performance* pelos gestores e profissionais de saúde não encontra paralelo na literatura brasileira, sendo que da análise descritiva da *proxy* C, sintetizou-se que os 24 domínios e cinco macrodomínios da PGP obtiveram médias elevadas (superior ou cerca de 3,0), sendo considerados relevantes/importantes pelos respondentes e que as três possíveis variáveis mediadoras analisadas (níveis de atenção à saúde, de complexidade da assistência e de cargo/função dos gestores) não foram consideradas como fatores mediadores relevantes da maioria dos macrodomínios e da PGP. Por sua vez, a confiabilidade e validade confirmadas dos cinco macrodomínios validados (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico), além dos 17 domínios que permaneceram na análise do modelo final (*figura 3*) guardam similaridades com as propostas do Proadess e IDSUS^{27,28}. Estes achados podem abrir novas perspectivas para a gestão dos serviços de saúde no sentido de se obter informações cientificamente válidas acerca da percepção dos decisores, possivelmente, evitando as dificuldades e desalinhamentos de métodos de aferição sem validação por especialistas e contribuindo para a mensuração da *performance* em serviços de saúde por um método de seleção de itens de avaliação a partir das valorações de especialistas, profissionais e gestores.

Por outro lado a eliminação de sete domínios de avaliação (D1A_Produtiv, D1B_EstrutAssist, D1C_CobertAssit, D2H_QualidAssist, D2M_IntegralAtenc, D2N_Equidade e D3P_EficaciaProj) na validação do modelo final pode ser interpretada como indício da falta de consenso sobre o escopo e a amplitude da

avaliação e do monitoramento em saúde¹⁹⁻²², que se reflete por diferentes abordagens de quadro teórico-conceptual (*framework*) dos sistemas de medição de *performance* do Reino Unido (NHS), do Canadá, da Austrália, dos EUA retratados por Arah et al.²⁰, e, no Brasil, pela multiplicidades de dimensões e indicadores dos projetos Proadess e IDSUS^{27,28} e trabalhos similares.

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 eliminadas no modelo final, revelando ampla concordância entre as propostas do painel de especialistas e as valorações dos grupos pilotos. 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), com impactos na variável latente PGP.

Conclusões

O processo resultou em uma proposta de um *proxy* de painel de indicadores e dimensões (*figura 3*). Destacam-se os aspectos inéditos na literatura de avaliação de *performance*, da colaboração com decisores chaves na valoração da relevância/importância e estruturação de um *proxy* de painel de indicadores e dimensões de *performance* do SUS, além da validação multicêntrica de modelos finais através de análise de cargas fatoriais, validade e confiabilidade dos itens, e, ainda, dos testes de

possíveis fatores mediadores na valoração.

Garantiu-se a confiabilidade e a validade de 17 domínios (indicadores) e de quatro macro-domínios (dimensões) de *performance*, além de modelar os impactos de seus reflexos sobre a PGP no modelo final (*figura 3*).

Os resultados deste estudo, com viés de gerencialismo (*accountability*), além de enfatizar a avaliação dos resultados em saúde⁵⁴ oferece painéis de indicadores^{27,28} para monitoramento dos serviços do SUS⁵⁵, tendo subsidiado a construção de um novo instrumento, alinhado pela literatura relevante de indicadores e dimensões de *performance*²⁰, que será aplicado na fase posterior da investigação, em planejamento de pesquisa de campo.

Colaboradores

Moita GF (0000-0003-4959-9424)* concebeu o estudo e elaborou o planejamento da pesquisa, analisou e interpretou os dados, elaborou o rascunho, revisou criticamente o conteúdo e aprovou a versão final do manuscrito. Raposo VMR (0000-0002-9328-8415)* colaborou substancialmente na revisão crítica do conteúdo, colaborou no planejamento da pesquisa e participou da aprovação da versão final do manuscrito. Barbosa ACQ (0000-0003-1266-5168)* colaborou substancialmente para a análise e interpretação dos dados, contribuiu significativamente na elaboração do rascunho e participou da aprovação da versão final do manuscrito. ■

Referências

1. Brasil. Constituição da República Federativa do Brasil de 1988. Distrito Federal: Brasília, DF: Congresso Nacional; 1988.
2. Brasil. Lei nº 8.080, de 19 de setembro de 1990. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências. Diário Oficial da República Federativa do Brasil. 19 Set1990. [acesso em 2019 nov 28]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/18080.htm
3. Brasil. Lei nº 8.142, de 12 de novembro de 1990. Brasília, DF: Congresso Nacional; 1990. Dispõe sobre a participação da comunidade na gestão do Sistema Único de Saúde (SUS) e sobre as transferências intergovernamentais de recursos financeiros na área da saúde e dá outras providências. Diário Oficial da República Federativa do Brasil. 12 Nov 1990. [acesso em 2019 nov 28]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L8142.htm.
4. Reis EJFB, Santos FP, Campos FE, et al. Avaliação da Qualidade dos Serviços de Saúde: Notas Bibliográficas. Cad. Saúde Pública. 1990; 6(1):50-61.
5. Flexner A. Medical Education in United States and Canada: Report to Carnegie Foundation for Advancement of Teaching. New York: Merrymount Press; 1910.
6. Porterfield JD. Evaluation of Patients: Bull. N. Y. Acad. Med. 1976; 52(1):30-8.
7. Perez AEB, Feller JJ. El Control de los Sistemas de

*Orcid (Open Researcher and Contributor ID).

- Atencion Médica Conceptualizacion y Mecánica Operativa. Medicina y Sociedad. 1983. [acesso em 2019 nov 2018]. Disponível em: www.medicinaysociedad.org.ar.
8. Alkin MC, Christie CA. Evaluation Roots: Tracing Theorist's Views and Influences. Thousand Oaks: Sage; 2004.
 9. Samico I, Felisberto E, Figueiró AC, et al. Avaliação em saúde: bases conceituais e operacionais. Rio de Janeiro: Medbook; 2010.
 10. Worthen BR, Sanders JR, Fitzpatrick JL. Avaliação de programas: Concepção e práticas. São Paulo: Edusp; Gente; 2004.
 11. Donabedian A. Basic approaches to assessment: structure, process and outcome. In: Press HA, editor. Explorations in Quality Assessment and Monitoring. 1. Michigan: Health Administration Press; 1980. p. 77-125.
 12. Brook RH, Lohr K. Efficacy, effectiveness, variations and quality: boundary-crossing research. Med Care. 1985; 23(5):23.
 13. Uchimura KY, Bosi MLM. Qualidade e subjetividade na avaliação de programas e serviços em saúde Cad. Saúde Pública. 2002; 18(6):1561-1569.
 14. Gattinara BC, Ibáñez J, Puente C, et al. Percepción de la comunidad acerca de la calidad de los servicios de salud públicos en los distritos Norte e Ichilo, Bolivia. Cad. Saúde Pública. 1995; 11(3):14.
 15. Vuori H. A qualidade da saúde. Divulgação em saúde para debate. 1991; (3):9.
 16. Acurcio FA, Cherchiglia ML, Santos MA. Avaliação de qualidade de serviços de saúde. Saúde debate. 1991; 38(101):368-378.
 17. Akerman M, Nadanovsky P. Evaluation of Health Services - What to Evaluate? Cad. Saúde Pública. 1992; 8(4):5.
 18. Santos MP. Avaliação da qualidade dos serviços públicos de atenção à saúde da criança sob a ótica do usuário. Revista Brasileira de Enfermagem. 1995; 11: 19. Murray CJL, Frenk J. A framework for assessing the performance of health systems Bulletin of the World Health Organization. 2000; 78(6):15.
 20. Arah OA, Klazinga NS, Delnoij DMJ, et al. Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement International Journal for Quality in Health Care. 2003; 15(5):377-98.
 21. Arah OA, Westert GP, Hurst J, et al. A conceptual framework for the OECD Health Care Quality Indicators Project International Journal for Quality in Health Care. 2006; 18(suppl):5-13.
 22. Arah OA, Custers T, Klazinga NS. Updating the Key Dimensions of Hospital Performance: The Move Towards a Theoretical Framework 3rd Workshop on Hospital Performance Measurement Barcelona: WHO Regional Office for Europe; 2003. [acesso em 2019 nov 18]. Disponível em: https://www.researchgate.net/publication/254902535 Updating_the_key_dimensions_of_hospital_performance_the_move_towards_a_theoretical_framework.
 23. World Health Organization. The World Health Report 2000. Health Systems: Improving Performance. Geneva: WHO; 2000.
 24. OECD Health Data 2001. Paris: Organisation for Economic Cooperation and Development 2001. [acesso em 2018 nov 10]. Disponível em: <http://www.oecd.org/health/health-systems/oecdhealthdata2001a-comparativeanalysisof30countries.htm>.
 25. Unit OHP. Health at a Glance. Paris: Organisation for Economic Co-operation and Development 2001. [acesso em 2019 nov 27]. Disponível em: <https://www.oecd.org/health/health-systems/health-at-a-glance-19991312.htm>.
 26. United Nations Development Programme. Handbook on Planning, Monitoring and Evaluation for Development Results. New York: UNDP; 2009.

27. Viacava F, Almeida C, Caetano R, et al. Uma metodologia de avaliação do desempenho do sistema de saúde brasileiro. *Ciênc. Saúde Colet.* 2004; 9(3):711-724.
28. Viacava F, Ugá MAD, Porto S, et al. Avaliação de Desempenho de Sistemas de Saúde: um modelo de análise. *Ciênc. Saúde Colet.* 2012; 17(4):921-934.
29. Guba EG, Lincoln YS. Avaliação de quarta geração. São Paulo: Unicamp; 2011.
30. Bryson J. What to do when stakeholders matter: Stakeholder identification and analysis techniques. *Public Management Review.* 2004; 6(1):33.
31. Hart S, Sharma S. Engaging fringe stakeholders for competitive imagination. *Academy of Management Executive.* 2004; 18(1):7-18.
32. Organização para a Cooperação e Desenvolvimento Económico. *Glossary of Key Terms in Evaluation and Results-Based Management.* Paris: OECD; 2002.
33. Morra Imas LG, Rist RC. *The Road to Results Designing and Conducting Effective Development Evaluations.* Washington, DF: Wolrd Bank; 2009.
34. Barbosa L, Neto AP. Ludwik Fleck (1896-1961) e translação do conhecimento: considerações sobre a genealogia de um conceito. *Saúde debate.* 2017; 41(esp):317-329.
35. CIHR CIoHR. *Knowledge Translation Strategy 2004-2009: Innovation in Action.* Ottawa: CIHR; 2004.
36. World Health Organization. *Bridging the “Know-Do” Gap: Meeting on Knowledge Translation in Global Health.* Geneva: WHO; 2006.
37. Minayo MCS, Assis SG, Sousa ER. Avaliação por triangulação de métodos - Abordagem de programas sociais. Rio de Janeiro: Fiocruz; 2010.
38. Vieira-Da-Silva LM. Conceitos, abordagens e estratégias para a avaliação em saúde. *Avaliação em saúde.* Salvador: EDUFBA; 2005. p. 15-39.
39. Souza LEPE, Vieira-Da-Silva LM, Hartz ZMA. 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: Fiocruz; 2005.
40. Hair JR, JF, William B, Babin B, et al. *Análise multivariada de dados.* 6. ed. Porto Alegre: Bookman; 2009.
41. Satorra A, Bentler PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye A, Clogg CC, editores. *Latent Variables Analysis: Applications for Developmental Research.* Thousand Oaks: Sage; 1994. p. 399-419.
42. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research.* 1981; 18(1):39-50.
43. Tenenhaus M, Esposito Vinzi V, Chatelin Y, et al. PLS path modeling. *Computational Statistics and Data Analysis.* 2005; 48:159-205.
44. Henseler J, Ringle CM, Sinkovics RR. The use of partial least squares path modeling in international marketing. In: Limited EGP, editor. *Inglaterra: New Challenges to International Marketing;* 2009. p. 277-319. (*Advances in International Marketing.* 20).
45. Nunnally JC, Bernstein IH. *The Assessment of Reliability.* *Psychometric Theory.* 1994; 3:248-92.
46. Nunnally JC, Berstein IH. *Psychometric theory.* 3. ed. New York; 1994.
47. Moita GF. *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 [Integrating Evaluation of Multidimensional Performance and Multicriteria Decision: A Panel Proxy of Efficiency, Effectiveness and Quality Indicators for Governance of Hospital Organizations and Health Services in Brazil].* [tese]. Coimbra: Universidade de Coimbra; 2019.

48. Bollen KA, Long JS. Testing Structural Equation Models [internet]. Londres: Newbury Park; 1993. [acesso em 2019 nov 12]. Disponível em: https://books.google.com.br/books?id=FvIxseYDLx4C&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.
49. Arbuckle JL. Amos 17 users' guide. Chicago, IL2008. [acesso em 2019 nov 27] Disponível em: http://www.sussex.ac.uk/its/pdfs/Amos_17.0_Users_Guide.pdf.
50. Wheaton B. Assessment of fit in overidentified models with latent variables. *Sociological Methods & Research*. 1987; 16(1):118-54.
51. Bentler PM. Comparative Fit Indexes in Structural Models. *Psychological Bulletin*. 1990; 107(2):238-46.
52. Bentler PM, Bonett PG. Significance Tests and Goodness-of-Fit in the Analysis of Covariance Structures. *Psychological Bulletin*. 1980; 88(3):588-606.
53. Steiger JH, Shapiro A, Browne MW. On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*. 1985; 50:253-64.
54. Fletcher A, Spiegelhalter D, Staessen J, et al. Implications for trials in progress of publication of positive results. *Lancet* 342. 1993; 5.
55. Silva LMV, Formigli VLA. Avaliação em Saúde: Limites e Perspectivas. *Cad. Saúde Pública*. 1994; 12.

Recebido em 16/04/2019

Aprovado em 11/10/2019

Conflito de interesses: inexistente

Supporte financeiro: não houve