

Confiabilidade e validade dos itens de avaliação no apoio à tomada de decisão e mensuração de performance na saúde – Uma análise do Sistema Único de Saúde

Galba Freire Moita

Ministério da Saúde do Brasil – Fiocruz - Faculdade de Economia da Universidade de Coimbra

Victor Manuel dos Reis Raposo

Faculdade de Economia da Universidade de Coimbra

Allan Claudius Queiroz Barbosa

Observatório de Recursos Humanos da FACE/UFMG

RESUMO

Trata-se de estudo colaborativo com 220 gestores e de saúde em 11 estados brasileiros para 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, que demonstrou elevados (> 4) graus de importâncias/relevâncias das subcategorias (20 e 24 domínios de avaliação) e das categorias (cinco e seis macrodomínios de avaliação). A análise inferencial, comprovou a confiabilidade e a validade do modelo final, validando 17 domínios (indicadores) e quatro macrodomínios de avaliação.

PALAVRAS-CHAVE: Governança em saúde; Avaliação de desempenho em saúde; Indicadores de monitoramento e avaliação em saúde; Avaliação em Saúde, Indicadores de saúde.

Área Temática

4. Políticas e Públicas e Planejamento Regional e Urbano

1. Introdução

Uma tendência gerencial recente tem exigido dos gestores públicos esforços de monitoramento de resultados em diversas áreas para a controladoria e responsabilização (*accountability*) dos gastos e recursos públicos. 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 de 3,6% nestes 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, através de indicadores específicos sobre a extensão do progresso, do atingimento dos objetivos e dos progressos alcançados na utilização de recursos alocados (OECD, 2002; World Bank, 2009).

Chianca e Youker (2004) registram um crescente interesse no campo de avaliação de sistemas de saúde da América Latina e Caribe. Uma pesquisa integrativa aponta que quase os enfoques avaliativos usados inspiram-se 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). 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.

Diante desses contextos 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 na validação e consolidação de um *proxy* de painel de indicadores multidimensionais aplicados em saúde pública no Brasil?

O objetivo principal é 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 a gestão de performance de unidades do SUS. Dessa forma, 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 visando 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 e expandidos para outros serviços públicos de saúde.

2. Avaliação em saúde e o modelo teórico-lógico

Os primórdios da avaliação em saúde são citadas em (Dos Reis et al., 1990) com os trabalhos de Flexner (Flexner, 1910) e do 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.

Alkin (Alkin & Christie, 2004) *apud* (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.

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 (Brook & Lohr, 1985); (Dos Reis *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.

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 governação (Raposo, 2007).

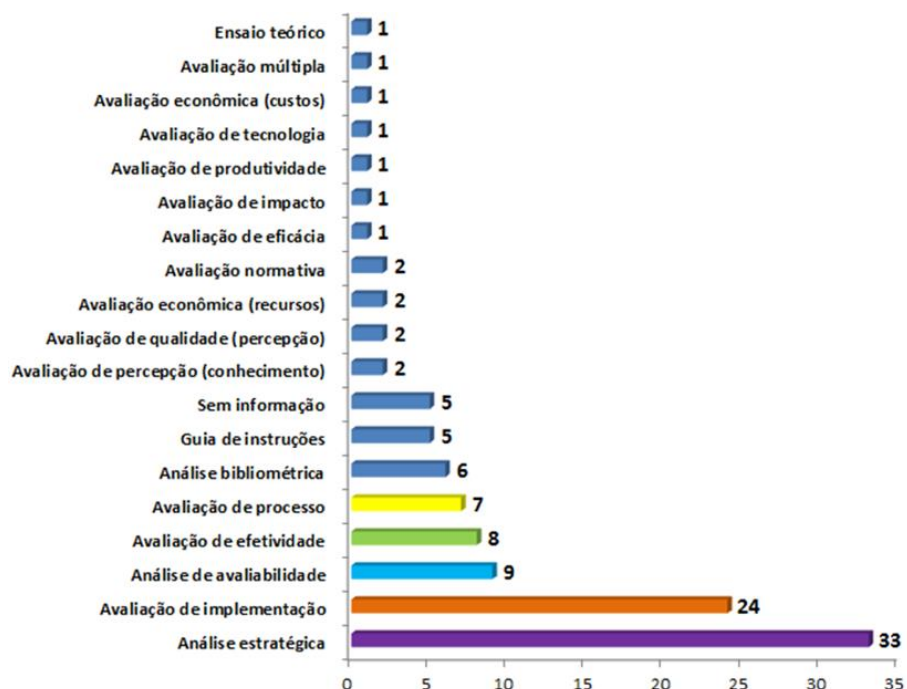
Este trabalho aborda aspectos de aplicabilidade da avaliação para a tomada de decisão em saúde no Brasil, como nos exemplos dos relatos de Tanaka e Tamaki (2012), de Contandriopoulos (2006), de Fernandes, *et al.* (2011), de Campos e Furtado (2011) e alguns outros estudos iniciais aplicados no sistema de saúde do Brasil. De acordo com Travassos *et al.* (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 governação 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 demandam 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, & Custers, 2003).

Outro destaque é o quadro teórico-conceitual proposto pelo 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 *et al.*, 2005). De forma mais instrumental, pode-se citar ainda alguns quadros teóricos-conceituais como o *Quality Indicator Project* (QIP) da Agência de Pesquisa e Qualidade nos Serviços de Saúde – *U.S. Agency for Healthcare Research and Quality* (AHRQ), com destaque no projeto da Fundação Europeia para Gerenciamento da Qualidade - *European Foundation for Quality Management* (EFQM), além de sistema de indicadores do *Project Assessment Tool of Quality Improvement in Hospitals* (PATH) da Organização Mundial de Saúde (OMS). No Reino Unido, duas abordagens, a primeira baseada na definição de indicadores do Serviço Nacional de Saúde/NHS, e a segunda interessada na concretização de uma metodologia para estabelecer “*hospital ratings*”; em Portugal, o *Sinas@Hospitais*¹ sob a coordenação da Entidade Reguladora de Saúde (ERS), que classifica os hospitais avaliados (*ranking*) que é utilizado pela Administração Central do Sistema de Saúde (ACSS), entre outros.

No contexto brasileiro, uma revisão sistemática (Machado, Martins, & Martins, 2013) voltada para avaliação de qualidade hospitalar mostra que os primeiros estudos surgiram entre 1991 a 2003, tendo encontrado apenas 48 publicações (quadro 26, anexo 4). Em outro estudo sobre a avaliação em saúde no Brasil (Fernandes *et al.*, 2011) acharam apenas 44 estudos. Neste propósito, procedeu-se 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 (gráfico 1).

Gráfico 1 - Bibliometria de estudos de avaliação em saúde no Brasil (1990 - 2018).



Fonte: elaborado pelos autores a partir de (Moita, 2019)

Na saúde brasileira, existem alguns sistemas informatizados, indicadores e instrumentos gerenciais, no campo de saúde pública, tais como: indicadores do DATASUS, Projeto Desenvolvimento de Metodologia de Avaliação do Desempenho do Sistema de Saúde Brasileiro (PROADESS), indicadores do Índice de Desempenho do SUS (IDSUS), Indicadores e Dados Básicos da Saúde (IDB) da Rede Interagencial de Informações para a Saúde (RIPSA) e indicadores do novo PNASS. No âmbito da Agência Nacional de Saúde Suplementar (ANS), Índices de Desempenho da Saúde Suplementar (IDSS) e indicadores QUALISS do Programa de Qualificação dos Prestadores de Serviços de Saúde Suplementar (QUALISS) – em fase de estruturação e; nas organizações privadas: indicadores PROHASA/CQH² e indicadores SIPAGEH³, dentre outros que surgirem na revisão da revisão integrativa da literatura.

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 pelos seus futuros utilizadores. Neste quesito, Uchimura e Bosi (Uchimura & Bosi, 2002) (2002) referem autores e listas das possíveis dimensões de avaliação: Gattinara *et al.* (Gattinara, Ibacache, Puente, Giaconi, & Caprara, 1995) (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) (1991; também citado por Acurcio *et al.* (Acurcio, Cherchiglia, & Santos, 1991), 1991; Akerman & Nadanovsky (Akerman & Nadanovsky, 1992), 1992; Santos, (Santos, 1995) 1995) aponta: [...] efetividade; [...] eficácia; [...] eficiência.

Em outro prisma, ao desenvolver uma abordagem da avaliação denominada “*avaliação focada para a utilização*”, Patton (Patton, 1997) inclui na sua definição a maneira ou o método de fazer a avaliação de resultados dos programas e fazer julgamentos sobre os programas e/ou para subsidiar o processo de tomada de decisões sobre futuras programações.

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 a seguir.

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*) (Alves, 2012). Todas estas teorias apresentam vantagens e desvantagens sintetizadas por Alves (Alves, 2012) que destaca a teoria dos *stakeholders* com diversas vantagens para a gestão dos sistemas produtivos de bens e serviços públicos.

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: a visão instrumental, descritiva e a de responsabilidade social (Carvalho, 2007).

Diante dos objetivos desta investigação, embasou-se 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.

Assim, metodologias e sistemas de indicadores de destaque no campo de saúde pública podem ser analisados e sistematizados de acordo com suas possíveis potencialidades, virtudes, limitações e desafios. Desta forma, é possível compor uma matriz de dimensões e de indicadores para avaliação e validação pelos agentes de decisão da amostra desta investigação.

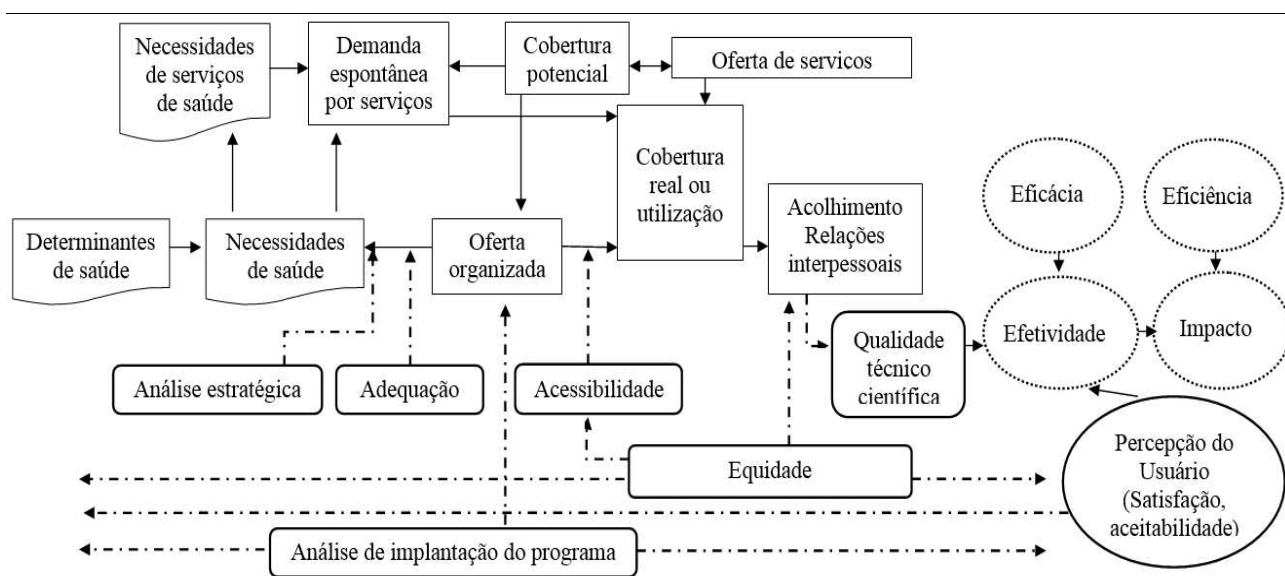
3. Método

A natureza da pesquisa caracteriza-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 chaves da amostra (especialistas, profissionais e gestores da saúde).

Na exploração inicial dos conhecimentos prévios dos agentes de decisão, seguiu-se os princípios de Sampieri *et al.* (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. Citam a limitação de generalização de resultados, mas a boa captação e compreensão dos significados em contextos específicos.

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 de marco teórico foi pautada em dois diferentes prismas. Um marco lógico da organização dos serviços e outro marco lógico centrado na equalização de oferta, na demanda e na mensuração de resultados (figura 1).

Figura 1. Diagrama do modelo lógico – atributos ou características para avaliação em saúde



Fonte: elaborado pelos autores a partir de Freire -Moita (2019)

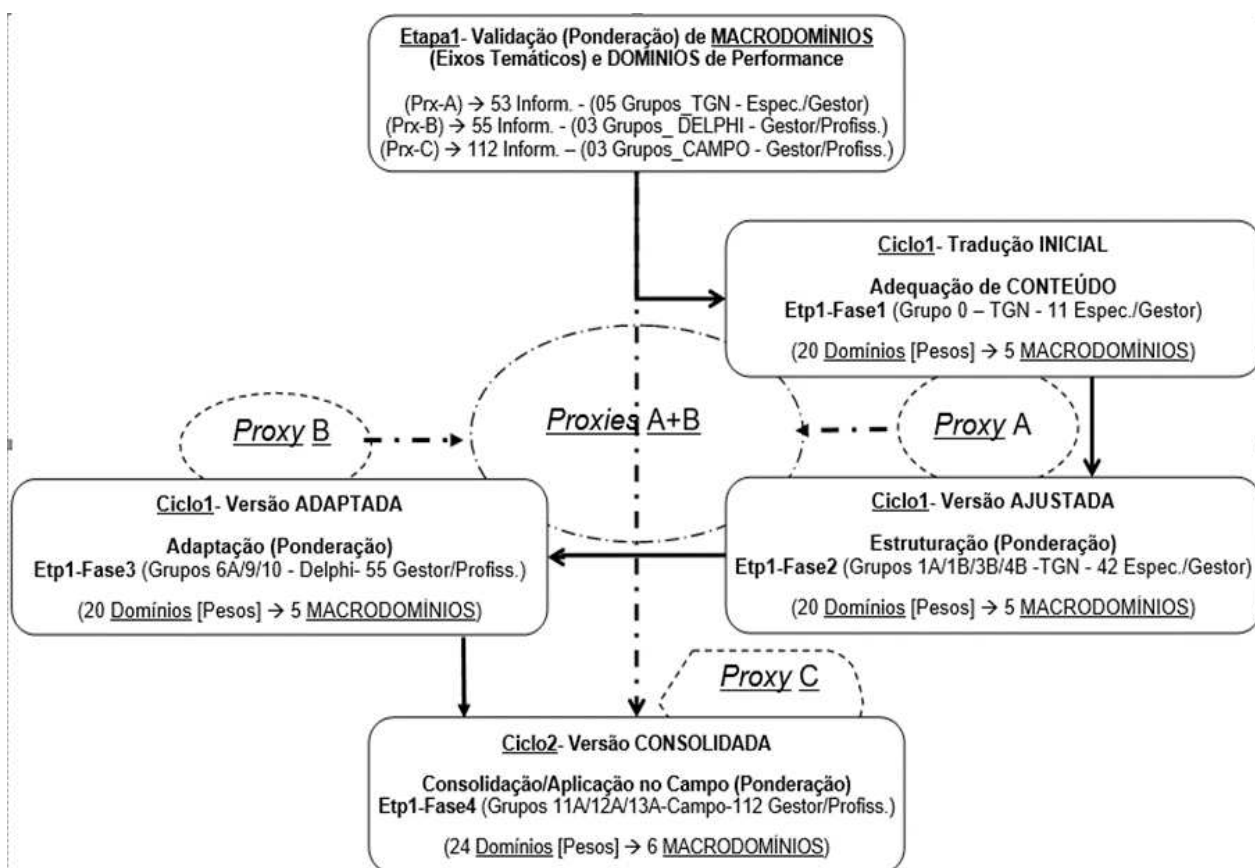
Foi inicialmente selecionado um grupo piloto de gestores e profissionais para formar um painel de 11 especialistas (fase 1), que fizeram a adequação de conteúdo, devidamente analisados por 42 especialistas e gestores de saúde em quatro grupos pilotos (fase 2) que, em seguida, foi validada por outros três grupos pilotos de 55 profissionais e gestores de saúde (fase 3) conformando um *proxy* inicial. Em seguida, aplicou-se uma pesquisa de campo de validação por 112 gestores de saúde que consolidaram um *proxy* final com 24 subcategorias de avaliação (domínios de avaliação) e seis categorias (macrodomínios).

Os informantes-chave selecionados foram perguntados sobre o grau de relevância e preferência dos macrodomínios (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, baseados em modelos que emergiram da revisão da literatura. Também foram instados a classificar o grau de usabilidade (importância/relevância), das macrodimensões e indicadores de performance pelos níveis governamentais/sistêmicos, organizacionais e gerenciais/chefia setoriais, com base na ponderação de itens por uma escala *Likert* de 5 níveis.

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.

Após as etapas de análise de literatura e de preparação de instrumentos de coleta de dados, a aplicação da pesquisa foi dividida em quatro fases sucessivas e incrementais de adaptação, sintetizadas na figura 2.

Figura 2. Fluxo do processo de validação de macrodomínios e domínios de avaliação (fases 1 a 4)



Fonte: elaborado pelos autores a partir de (Moita, 2019)

Estes 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 a análise estatística de dados das *proxies* da valoração dos domínios de avaliação, além de seus reflexos nos macrodomínios avaliação, descritos no quadro 1 a seguir.

Quadro 1 - 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 <u>domínios</u>)				
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 <u>macrodomínios</u>):				
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 <u>macrodomínios</u> 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 pilotos 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, tendo captado respostas de três grupos pilotos (55 gestores e profissionais de gestão de saúde pública) a fim de consolidar os níveis de conhecimentos específicos finais sobre os domínios e macrodomínios de avaliação em sistemas e serviços de saúde, poderia agregar novos itens de avaliação (domínios e/ ou macrodomínios) que julgassem relevantes.

Em seguida, os resultados obtidos nas respostas das fases 1 a 3 foram avaliados, categorizados e readequados conforme as macrodimensões de performance definidos pela literatura mais relevante (Acurcio et al., 1991; Akerman & Nadanovsky, 1992; Gattinara et al., 1995; Santos, 1995; Viacava et al., 2004; Vieira-Da-Silva, 2005; Vuori, 1991) subsidiando o ajuste e consolidação de um novo questionário semiestruturado (fase 4).

Por fim, na fase 4, organizou-se outros três grupos pilotos de informantes chaves, 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. 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), ambos 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, na fase 4.

A análise final e modelagem do instrumento definitivo deste estudo partiu das valorações atribuídas aos domínios e macrodomínios de avaliação por uma amostra de 112 gestores e profissionais, tendo sido 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.

4. Resultados

A presente investigação foi desenvolvida a partir dos modelos conceituais e conhecimentos gerais de um grupo de 220 especialistas, profissionais e gestores de saúde com o objetivo de medir os níveis de adequação e adaptabilidade de um instrumento de mensuração da performance em unidades de saúde aplicável ao SUS.

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 cinco macrodomínios de performance). Em fase posterior (fase 4), abordou-se a análise da *proxy* C - 112 informantes. (24 domínios e seis macrodomínios de performance).

Vale lembrar que cada um dos itens 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 de todas as respostas obtidas.

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% trabalhava no nível de complexidade Outros (estruturas não assistenciais, como apenas de gestão ou governança). Por sua vez, na fase 4, 61,64% trabalhava 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) possuía 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. A formação/profissão mais frequente entre os indivíduos foi a de Enfermeiro (23,36%; 40,18), seguida por Administrador Hospitalar (22,43%; 17,86%) e Assistente Social (10,28%; 9,82%).

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%

	Mestrado/ PhD	18	17,31%	10	9,01%
Tempo de experiência	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 - 60meses	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 - 60meses	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: elaborado pelos autores a partir de (Moita, 2019)

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)

Macrodomínios	Domínios	<i>Proxy</i> A+B (fases 1 a 3)				<i>Proxy</i> C (fase 4)			
		N	Média	D.P.	I.C-95%	N	Média	D.P.	I.C-95% ¹
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_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]
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]
Efeitos	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_SatisfUsuario	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_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: elaborado pelos autores 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 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]

Nota.¹ Intervalo Bootstrap

Fonte: elaborado pelos autores a partir de (Moita, 2019)

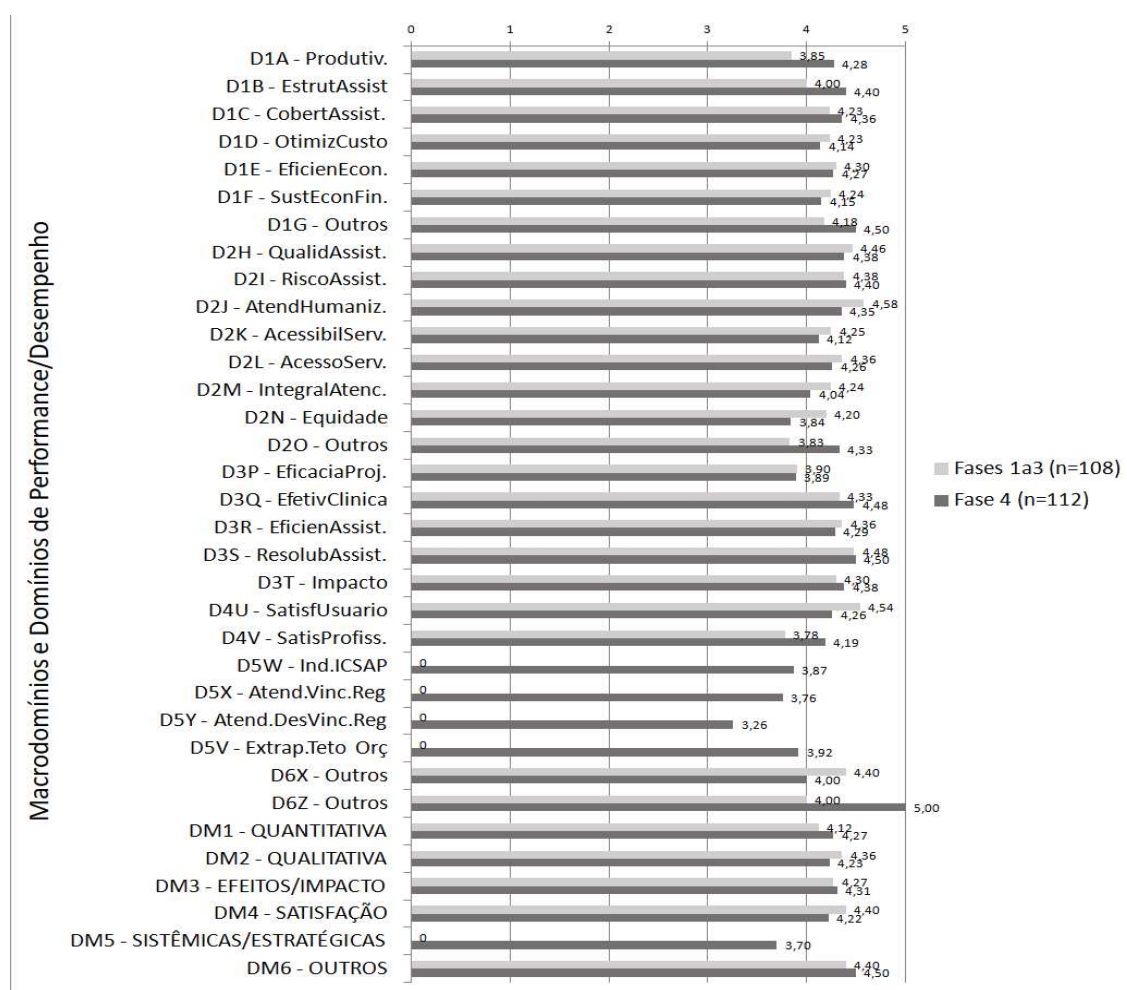
A análise global permite afirmar que todos os quatro macrodomínios de avaliação apresentaram valoração média elevada (superior ou cerca de 4, muito importante/relevante), nas *proxies* A e B (n=108) e na *proxy* C (n = 112), além do quinto macrodomínio (Sistêmico/Estratégico) adicional, composto por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

Quanto as análises de significância estatística quanto as 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 e que 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 importâncias/relevâncias 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 3. Além disso, em média, as valorações da ampla maioria destes itens de análise foram razoavelmente dentro das mesmas margens de valores, uma vez que a maioria dos intervalos de confiança se sobrepuseram.

Figura 3 - 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: elaborado pelos autores 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. Estes resultados subsidiaram a construção de um novo instrumento, devidamente alinhado pela literatura relevante de indicadores e dimensões de performance (Arah 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

Em busca de validar os cinco macrodomínios de avaliação, além de estruturar e validar a variável latente Performance Global Percebida, procedentes da *proxy C* (112 informantes), aplicou-se métodos de análise inferencial a fim de consolidar os modelos desenhados nesta investigação, em três categorias iniciais de análises inferenciais: estudo das cargas fatoriais entre variáveis, validação por critérios de qualidade e validade além de análise da qualidade (ajustamento) do modelo final resultante, que se apresentam a seguir, sendo complementadas pela análise das correlações entre as variáveis e suas influências na Performance Global Percebida.

Para validar os macrodomínios de avaliação estudados (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico) foi feita uma Análise Fatorial Confirmatória (AFC) (Hair, William, Babin, & Anderson, 2009) para os macrodomínios de avaliação, na *proxy C* (n = 112). Logo, pode-se destacar 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 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*, com as cargas fatoriais conforme tabela 4.

Tabela 4 - Análise Fatorial Confirmatória dos domínios de avaliação – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Macrodomínios	Domínios	Modelo Inicial		Modelo Final	
		Peso ¹	C.F. ²	Peso ¹	C.F. ²
Quantitativos	D1A_Produtiv	15%	0,48		
	D1B_EstrutAssist	11%	0,39		
	D1C_CobertAssit	10%	0,37		
	D1D_OtimizCusto	21%	0,71	29%	0,63
	D1E_EficienEcon	20%	0,66	37%	0,82
	D1F_SustEconFin	21%	0,65	34%	0,71
Qualitativos	D2H_QualidAssist	12%	0,47		
	D2I_RiscoAssist	13%	0,60	23%	0,66
	D2J_AtendHumaniz	15%	0,61	29%	0,75
	D2K_AcessibilServ	14%	0,54	22%	0,54
	D2L_AcessoServ	16%	0,62	26%	0,60
	D2M_IntegralAtenc	19%	0,65		
Efeitos	D2N_Equidade	12%	0,53		
	D3P_EficaciaProj	20%	0,51		
	D3Q_EfetivClinica	17%	0,58	21%	0,58
	D3R_EficienAssist	26%	0,80	32%	0,79
	D3S_ResolubAssist	19%	0,63	23%	0,61
Satisfação	D3T_Impacto	18%	0,61	24%	0,64
	D4U_SatisfUsuario	45%	0,71	44%	0,70
Sistêmico/Estratégia	D4V_SatisfProfiss	55%	0,81	56%	0,83
	D5W_IndICSAP	22%	0,50	22%	0,50
	D5X_AtendVincReg	24%	0,72	24%	0,72
	D5Y_AtendDesVincReg	31%	0,81	31%	0,81
	D5V_ExtrapTetoOrç	23%	0,59	23%	0,59

Nota.¹ Peso relativo; ² Carga Fatorial.

Fonte: elaborado pelos autores a partir de (Moita, 2019)

Na Tabela 5, pode-se observar que a qualidade e a validade dos cinco macrodomínios de avaliação (Quantitativos, Qualitativos, Efeitos, Satisfação e Sistêmico/Estratégico) foi assegurada, uma vez que todos 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).

Tabela 5 - Validação por critérios de qualidade e validade dos macrodomínios de avaliação - Proxy C (ciclo 2, estudo 1) (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

Fonte: elaborado pelos autores a partir de (Moita, 2019)

Para validar a variável latente Performance Global Percebida foi feita uma Análise Fatorial Confirmatória (AFC) (Hair et al., 2009) da valoração média obtida para os macrodomínios de avaliação, 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, com as cargas fatoriais vistas na tabela 6.

Tabela 6 - Análise Fatorial Confirmatória da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Macrodomínios	Modelo Inicial		Modelo Final	
	Peso ¹	C.F. ²	Peso ¹	C.F. ²
Quantitativa	15%	0,63	15%	0,48
Qualitativa	19%	0,90	21%	0,82
Efeitos	24%	0,82	19%	0,76
Satisfacao	25%	0,74	27%	0,78
Sistemico	17%	0,53	19%	0,56

Nota.¹ Peso relativo; ² Carga Fatorial.

Fonte: elaborado pelos autores a partir de (Moita, 2019)

Na Tabela 7, pode-se observar que a qualidade e a validade da variável latente Performance Global Percebida foram asseguradas, uma vez que apresentou 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).

Tabela 7 - Validação da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Macrodomínios	A.C. ¹	C.C. ²	Dim. ³	AVE ⁴	VCM ⁵
5	0,73	0,82	1	0,48	-

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

Fonte: elaborado pelos autores a partir de (Moita, 2019)

Há uma diversidade de parâmetros e índices de ajustamento e qualidade dos modelos de equações estruturais (Bollen & Long, 1993). Decidiu-se por usar um grupo de índices amplamente aplicados na literatura relevante ($X^2/G.L.$ – teste de qui-quadrado por grau de liberdade, CFI – índice de ajustamento comparativo, TLI – índice de ajustamento não normalizado ou índice Tucker-Lewis e o índice RMSEA – erro quadrático médio de aproximação) a fim de medir a qualidade (ajustamento) do modelo construído para cada um dos itens em análise.

Também se aplicou testes para medir a qualidade (ajustamento) do modelo construído para os macrodomínios e domínios de avaliação da *proxy* C (112 informantes), através de alguns índices selecionados na literatura relevante ($X^2/G.L.$, CFI, TLI e RMSEA). Para um bom ajuste espera-se que $X^2/G.L.$ seja menor que 3 (Arbuckle, 2008; Wheaton, 1987), CFI seja maior que 0,80 (Bentler, 1990), TLI seja maior que 0,80 (Bentler & Bonnet, 1980) e que o RMSEA (Steiger, Shapiro, & Browne, 1985) seja menor que 0,10, sendo o ideal abaixo de 0,05.

Na Tabela 8, 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 do foi menor que o limite máximo de 0,10.

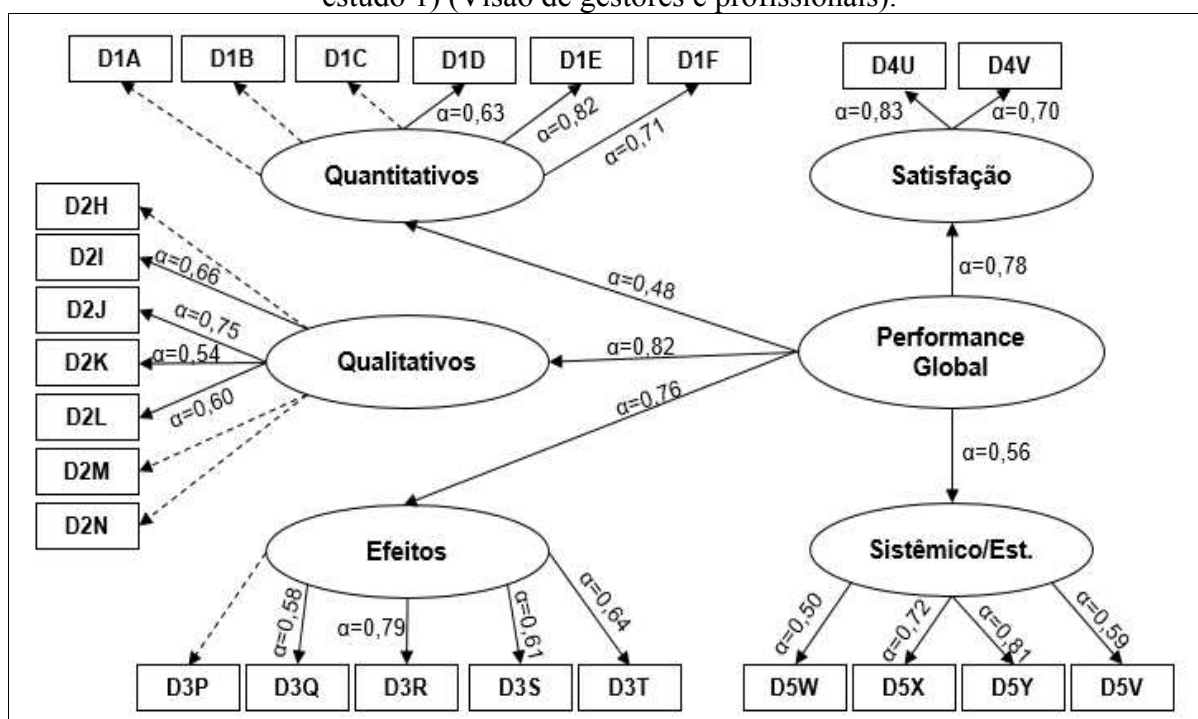
Tabela 8 - Parâmetros de qualidade do modelo final da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).

Medidas de Qualidade do Ajuste	Estatística de Teste	Estatística de Teste Reescalada (Satorra e Bentler 1994)
χ^2	166,61	150,32
G.L.	114	114
$\chi^2/G.L.$	1,46	1,32
TLI	0,888	0,888
CFI	0,906	0,906
RMSEA	0,064	0,053
valor-p (RMSEA \leq 0,05)	[0,042; 0,084]	[0,028; 0,074]
	0,137	0,389

Fonte: elaborado pelos autores a partir de (Moita, 2019)

Diante destes resultados, a figura 4 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 optou-se por omitir suas nomenclaturas, que ficaram identificados como D1A a D5Y.

Figura 4 - Análise Fatorial Confirmatória da Performance Global Percebida – Proxy C (ciclo 2, estudo 1) (Visão de gestores e profissionais).

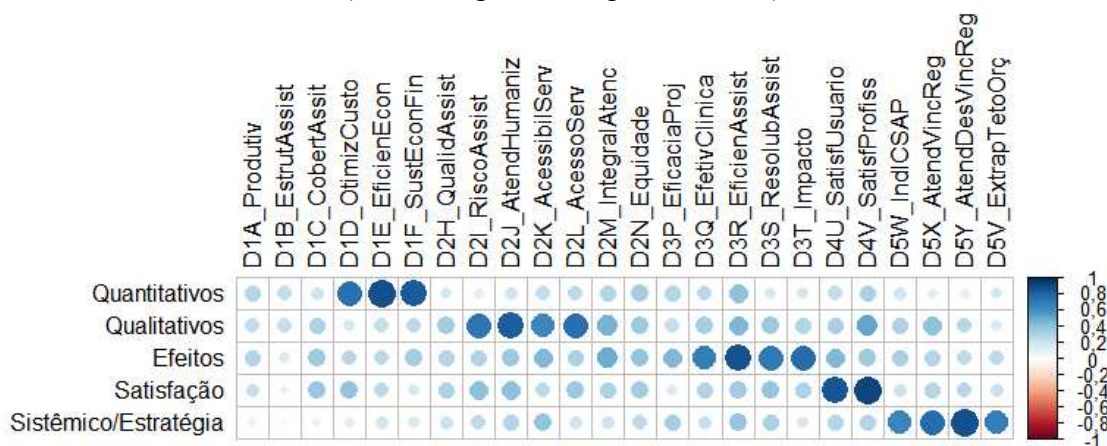


Fonte: elaborado pelos autores a partir de (Moita, 2019)

Pode-se destacar que, no modelo final, cinco dos 24 domínios de avaliação (indicadores) propostos pelo painel de especialistas apresentaram cargas fatoriais validadas elevadas (0,75 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. 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 (0,75 a 1,00), 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.

Noutra vertente a análise de correlação mostrou que os domínios de avaliação apresentaram maiores correlações com seus respectivos macrodomínios, na *proxy C* (fase 4) (figura 4).

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



Fonte: elaborado pelos autores a partir de (Moita, 2019)

Por fim, a comparação da Performance Global Percebida (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 demonstrou 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) para a PGP média; o nível Primário de atenção apresentou a maior variabilidade (0,34) na Performance Global Percebida 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 Performance Global Percebida, 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 nem entre os níveis do cargo/função (0,332). Pode-se observar uma variabilidade na valoração média da variável latente PGP, conforme os níveis de atenção à saúde ou de nível de cargo/função. Porém, 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*.

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 importâncias/relevâncias 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 2) para a aplicação da análise inferencial. Além disso, em média, as valorações da ampla maioria destes 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 as 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 et al., 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 a ampla maioria dos indicadores e dimensões. Também se aplicou testes para medir a qualidade (ajustamento) do modelo construído para as macrodimensões de qualidade percebida, através de alguns índices selecionados na literatura relevante ($X^2/G.L.$, 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 3).

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.

Destes 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 revelam a necessidade de novos estudos, em amostras maiores, tendo sido um processo inovador que estabeleceu subcategorias (indicadores) e categorias (dimensões) que pode subsidiar a construção de um novo instrumento aplicável aos gestores do SUS, já em planejamento para avanços da pesquisa, inclusive através do desenvolvimento de protótipo de uma plataformas de informática, usado como suporte tecnológico para aplicação do instrumento validado.

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. Referencias Bibliográficas

- Acurcio, F. A., Cherchiglia, M. L., & Santos, M. A. (1991). Avaliação de qualidade de serviços de saúde. *Saúde em debate*, 33, 4.
- Akerman, M., & Nadanovsky, P. (1992). Evaluation of Health Services - What to Evaluate? *Caderno de Saúde pública*, 8(4), 5.
- Alkin, M. C., & Christie, C. A. (2004). *Evaluation Roots: Tracing Theorist's Views and Influences*. Sage publications.
- Alves, A. S. D. (2012). *Responsabilidade e Governança na Moderna Gestão Hospitalar*. (Doctoral). Universidade de Aveiro, Portugal: Aveiro.
- Arah, O. A., Klazinga, N. S., Delnoij, D. M. J., Asbroek, A. H. A. T., & Custers, T. (2003). Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement *International Journal for Quality in Health Care*, 15(5), 22. doi:10.1093/intqhc/mzg049
- Arbuckle, J. L. (2008). *Amos 17 users' guide* (SPSS Ed.). Chicago, IL.
- Barbosa, L., & Neto, A. P. (2017). Ludwik Fleck (1896-1961) e translação do conhecimento: considerações sobre a genealogia de um conceito. *Saúde Debate*, 41(Especial), 13.

- Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, 107(2), 238-246.
- Bentler, P. M., & Bonett, P. G. (1980). Significance Tests and Goodness-of-Fit in the Analysis of Covariance Structures. *Psychological Bulletin*, 88(3), 588-606.
- Bollen, K. A., & Long, J. S. (1993). *Testing Structural Equation Models*. In S. F. Ediths (Ed.), (Vol. 154, pp. 320).
- Bosi, M. L., & Mercado, F. J. (2006). *Pesquisa qualitativa de serviços de saúde* (Vozes Ed.).
- Brook, R. H., & Lohr, K. (1985). Efficacy, effectiveness, variations and quality: boundary-crossing research. *Med Care*, 23(5), 23.
- Bryson, J. (2004). What to do when stakeholders matter: Stakeholder identification and analysis techniques. *Public Management Review*, 6(1), 33.
- Campos, R. T. O., & Furtado, J. P. (2011). *Desafios da Avaliação de Programas e Serviços em saúde: Novas Tendências e Questões Emergentes* (Editora Unicamp ed.). Campinas, São Paulo: Editora Unicamp.
- Carvalho, C. (2007). *Organizações, actores envolvidos e partes interessadas: Determinantes da saliência dos stakeholders e sua relação com o desempenho organizacional*. (Doctoral). Universidade de Coimbra, Coimbra, Portugal.
- Chianca, T., & Youker, B. (2004). *La evaluación em América Latina y El Caribe. Visión general de los desarrollos recientes*.
- CIHR, C. I. o. H. R. (2004). *Knowledge Translation Strategy 2004-2009: Innovation in Action*. Retrieved from Ottawa: <http://www.cihr-irsc.gc.ca/e/26574.html>
- Clavier, C., Sénéchal, Y., & Potvin, L. (2011). A theory - based model of translation practices in public health participatory research. *Sociology of Health and Illness*, 34(5), 16.
- Contandriopoulos, A. P. (2006). Avaliando a institucionalização da avaliação. *Ciência Saúde Coletiva*, 10(3), 7.
- Donabedian, A. (1980). Basic approaches to assessment: structure, process and outcome. In H. A. Press (Ed.), *Explorations in Quality Assessment and Monitoring* (Vol. 1, pp. 77-125). Michigan: Health Administration Press.
- Dos Reis, E. J. F. B., Dos Santos, F. P., De Campos, F. E., Acúrcio, F. A., Leite, M. T. T., Leite, M. L. C., . . . Dos Santos, M. A. (1990). Avaliação da Qualidade dos Serviços de Saúde: Notas Bibliográficas. *Cadernos de Saúde Pública*, 6(1), 12.
- Déniz-Déniz, M., & Zárraga-Oberty, C. (2004). 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*, 13(4), 17.
- Fals Borda, O., & Mora-Osejo, L. E. (2004). La superación del Eurocentrismo. Enriquecimiento del saber sistémico y endógeno sobre nuestro contexto tropical. *Polis, Revista de la Universidad Bolivariana*, 2(7).
- Fernandes, F. M. B., Ribeiro, J. M., & Moreira, M. R. (2011). Reflexões sobre Avaliação de Políticas de Saúde no Brasil. *Cad. Saúde Pública*, 27(9), 11.
- Flexner, A. (1910). *Medical Education in United States and Canada: Report to Carnegie Foundation for Advancement of Teaching*. New York: Merrymount Press
- Freeman, R. (1984). *Strategic management: A stakeholder approach* (Pitman Ed.). Boston: Pitman.
- Gattinara, B. C., Ibacache, J., Puente, C., Giacconi, J., & Caprara, A. (1995). Percepción de la comunidad acerca de la calidad de los servicios de salud públicos en los distritos Norte e Ichilo, Bolívia. *Caderno de Saúde Pública*, 11, 14.
- Hair, J. F., William, B., Babin, B., & Anderson, R. E. (2009). *Análise multivariada de dados* (Bookman Ed. 6a ed.). Porto Alegre.
- Hart, S., & Sharma, S. (2004). Engaging fringe stakeholders for competitive imagination. *Academy of Management Executive*, 18(1), 12.
- Hartz, S., & John, J. (2008). 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*, 24(4), 9. doi:10.1017/S0266462308080616
- Machado, J. P., Martins, A. C. M., & Martins, M. S. (2013). Avaliação da qualidade do cuidado hospitalar no Brasil: uma revisão sistemática. *Cad. Saúde Pública*, 29(6), 20.
- Mellahi, K., & Wood, J. (2003). The role and potential of stakeholders in "holow participation": Conventional stakeholder theory and institutionalism alternatives. *Business and Society Review*, 108(2), 20.

- Moita, G. F. (2019). *Avaliação Integrativa de Performance Multidimensional e Decisão Multicritério: Um Proxy de Painel de Indicadores de Eficiência, Efetividade e Qualidade para Governança de Organizações Hospitalares e Serviços de Saúde no Brasil [Integrating Evaluation of Multidimensional Performance and Multicriterary Decision: A Panel Proxy of Efficiency, Effectiveness and Quality Indicators for Governance of Hospital Organizations and Health Services in Brazil]*. (Doutoramento Doctoral). Universidade de Coimbra, Coimbra.
- Morra Imas, L. G., & Rist, R. C. (2009). *The Road to Results Designing and Conducting Effective Development Evaluations*. Retrieved from Washington: <https://openknowledge.worldbank.org/handle/10986/2699>
- OECD. (2017). *Health at a Glance 2017: OECD Indicators*. Retrieved from Paris: http://dx.doi.org/10.1787/health_glance-2017-en
- OECD (Ed.) (2002). *Glossary of Key Terms in Evaluation and Results-Based Management*. Paris.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation: The New Century Text* (S. Publications Ed.). London: Sage Publications.
- Porterfield, J. D. (1976). Evaluation of Patients: Codman Revisited. *Acad. Med*, 52(1), 9.
- Raposo, V. (2007). *Governança hospitalar - uma proposta conceptual e metodológica para o caso português*. Universidade de Coimbra, Coimbra.
- Samico, I., Felisberto, E., Figueiró, A. C., & Frias, P. G. (2010). *Avaliação em saúde: bases conceituais e operacionais* (Medbook Ed.). Rio de Janeiro: Medbook.
- Sampieri, R. H., Collado, C. F., & Lúcio, P. B. (2006). *Metodologia de pesquisa* (M. Hill Ed.). São Paulo: McGraw Hill.
- Santos, M. P. (1995). Avaliação da qualidade dos serviços públicos de atenção à saúde da criança sob a ótica do usuário. *Revista Brasileira de Enfermagem*, 48, 11.
- Souza, L. E. P. F., Vieira-Da-Silva, L. M., & Hartz, Z. M. A. (2005). *Avaliação em Saúde: dos modelos teóricos à prática na avaliação de programas e sistemas de saúde*. Paper presented at the Conferência de consenso sobre a imagem-objetivo da descentralização da atenção à saúde no Brasil, Rio de Janeiro.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253–264.
- Tanaka, O. Y., & Tamaki, E. M. (2012). O papel da avaliação para a tomada de decisão na gestão de serviços de saúde. *Ciência & Saúde Coletiva*, 17(4), 8.
- Travassos, C., Noronha, J. C., & Martins, M. (1999). Mortalidade Hospitalar como Indicador de Qualidade: Uma Revisão. *Ciência & Saúde Coletiva*, 4(2).
- Uchimura, K. Y., & Bosi, M. L. M. (2002). Qualidade e subjetividade na avaliação de programas e serviços em saúde *Cad. Saúde Pública*, 18(6), 9.
- UNDP. (2009). *Handbook on Planning, Monitoring and Evaluation for Development Results*.
- Veillard, J., Champagne, F., Klazinga, N., Kazandjian, V., Arah, O. A., & Guisste, A. L. (2005). A performance assessment framework for hospitals: the WHO regional office for Europe PATH Project. *International Journal for Quality in Health Care*, 17(6), 9.
- Viacava, F., Almeida, C., Caetano, R., Fausto, M., Macinko, J., Martins, M., . . . Szwarcwald, C. L. (2004). Uma metodologia de avaliação do desempenho do sistema de saúde brasileiro. *Ciência & Saúde Coletiva*, 9(3), 14.
- Vieira-Da-Silva, L. M. (2005). Conceitos, abordagens e estratégias para a avaliação em saúde. In *Avaliação em saúde* (pp. 15-39). Salvador: EDUFBA.
- Vuori, H. (1991). A qualidade da saúde. *Divulgação em saúde para debate*, 1, 9.
- Wheaton, B. (1987). Assessment of fit in overidentified models with latent variables. *Sociological Methods & Research*, 16(1), 118-154. doi:10.1177/0049124187016001005
- WHO, W. H. O. (2006). *Bridging the "Know-Do" Gap: Meeting on Knowledge Translation in Global Health*. Retrieved from Geneva:

1 Sinas@Hospitais – Sistema de Indicadores de Performance de Hospitais do Serviço Nacional de Saúde de Portugal

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

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