

AVALIAÇÃO DA EFETIVIDADE DE SISTEMAS DE GD: Sistematização da Produção Científica

MARIANA LOPES DE ARAÚJO

UNIVERSIDADE DE BRASÍLIA (UNB)

PEDRO PAULO MURCE MENESES

UNIVERSIDADE DE BRASÍLIA (UNB)

Agradecimento à órgão de fomento:

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

AVALIAÇÃO DA EFETIVIDADE DE SISTEMAS DE GD: Sistematização da Produção Científica

Introdução

Desde o surgimento da teoria de Gestão Estratégica de Pessoas (GEP), na década de 1980, as pesquisas sobre o campo têm feito progresso em demonstrar a relação entre as políticas e práticas de Gestão de Pessoas (GP) e a melhora nos resultados organizacionais (Becker & Huselid, 1998). Como parte sucessora desta premissa, tem-se o sistema de Gestão de Desempenho (GD) que prevê atividades, políticas, procedimentos e intervenções destinados a auxiliar funcionários a melhorarem seu desempenho e, por conseguinte, o desempenho organizacional (DeNisi & Murphy, 2017; DeNisi & Pritchard, 2006). Para isso, considera-se parte primordial a mensuração dos resultados individuais de desempenho por meio da etapa de Avaliação de Desempenho (AD) (Den Hartog, Boselie & Paauwee, 2004).

Enquanto os estudos sobre AD datam da década de 1920 e mostram-se relevantes até os dias atuais, o interesse de estudiosos e praticantes sobre o campo de GD tem se desenvolvido e aumentado somente nas últimas décadas (DeNisi & Murphy, 2017; DeNisi & Pritchard, 2006). Apesar disso, são frequentes as queixas, principalmente dos praticantes, acerca da dificuldade e incerteza na condução de uma avaliação de desempenho composta por parâmetros adequados e que esteja inserida em um sistema de GD que seja realmente capaz de promover desempenho organizacional superior (Cawley et al., 1998; DeNisi & Pritchard, 2006; Levy & Williams, 2004).

Demonstrar a efetividade dos sistemas de GD, assim como de outras políticas de GP, é considerada uma tarefa complexa e de difícil execução. É por esta razão que uma elevada quantidade de estudos tem, nas últimas três décadas, se debruçado em demonstrar a relação entre o incremento das políticas de pessoal e o desempenho organizacional superior (Jiang & Messersmith, 2018). Apesar do progresso neste intento, ainda hoje, têm-se a ausência de consenso acerca do conceito de “efetividade” e as ferramentas para mensuração da efetividade em sistemas de GP, como o de GD, se mostram desconexas e não convincentes (Awan et al., 2020; Haines & St-Onge, 2012; Ikramullah et al., 2016; Modipane et al., 2019; Roberts, 1995).

Em geral, têm-se estudos aplicando diferentes modelos, com critérios, variáveis e medidas distintas, para avaliação da efetividade de GD que não estabelecem relação um com o outro (Ikramullah et al., 2016). Como não há uma sistematização desta produção, há um acúmulo de estudos que não se comunicam e não se integram, dificultando a identificação de teorias e o desenvolvimento de ferramentas claras e com critérios objetivos sobre a avaliação da efetividade de GD.

Além disso, a efetividade do sistema de GD tem como integrante o processo de implementação. Como a implementação de uma política organizacional envolve diversos aspectos contextuais e estes desempenham um papel significativo no andamento do sistema de GD (Levy & Williams, 2004), tem-se um ambiente incerto, uma vez que não se tem garantia de que os objetivos propostos inicialmente serão alcançados.

Logo, este artigo teve como objetivo sintetizar, a partir do empreendimento de revisão sistemática da literatura, dimensões e critérios utilizados nos modelos dos estudos dedicados à avaliação da efetividade de sistemas de GD. Para isso, serão identificados os padrões demográficos, teóricos, metodológicos e empíricos de cada dimensão identificada. Este estudo contribui com a sistematização dos critérios adotados até então pelos estudos revisados, permitindo contribuir que novas pesquisas possam aperfeiçoar a avaliação da efetividade do sistema de GD, sobretudo com foco no campo da Administração.

Efetividade de sistemas de Gestão de Desempenho (GD)

Acessar a efetividade de práticas organizacionais é apontado como um processo complexo, pois há pouca concordância acerca da definição do que é efetividade, assim como sobre como realizar sua mensuração, situação que se atribui, em geral, a natureza inerentemente subjetiva de qualquer que seja a definição do termo (Ahmed, 1999).

Basicamente, a efetividade de um sistema de GP é representada pelo quão bem o sistema opera como uma ferramenta para alcance do desempenho organizacional (e.g. Becker & Huselid, 1998; Huselid, 1995; Jackson et al., 2014). Neste intento, pode-se afirmar que constitui-se um sistema de avaliação da efetividade do sistema de GD que inclui tanto a investigação do lado preditor da equação, indicando a configuração e os fatores que influenciam um sistema de GD efetivo, quanto a perspectiva dos resultados desta equação (Keeping & Levy, 2000), normalmente representado pelo desempenho organizacional enquanto variável dependente.

Entretanto, argumenta-se que autores que esta visão, considerando apenas os resultados organizacionais, restringe o processo de mudança organizacional a apenas a perspectiva macro organizacional, e mostra-se insuficiente para avaliar a efetividade de sistemas de GP, tal como a GD (Baird et al., 2012). É por esta razão que, segundo Cardy e Dobbins (1986) a efetividade é um construto multidimensional ou o último critério a ser alcançado, o que o torna de difícil mensuração direta.

Por isso, os estudos têm buscado identificar a efetividade, sob a perspectiva dos resultados, a partir da adoção de uma seara de critérios distintos de forma a complementar os achados anteriores. Neste sentido, Levy e Williams (2004) também identificaram que novas formas de avaliar a efetividade de sistemas de GD emergiram com importância na literatura no início dos anos 2000.

Colakoglu, Lepak & Hong (2006) demonstram que os estudos tradicionais de GP focam em critérios individuais, como motivação, satisfação, comprometimento e outros, diferentemente dos estudos estratégicos de pessoal que adotam critérios a nível da unidade e da organização, como produtividade e crescimento de vendas. Já Dyer e Reeves (1995) categorizaram os critérios de avaliação da efetividade em relação a proximidade com a prática de GP investigada. Para os autores, uma primeira categoria refere-se aos resultados do indivíduo que são influenciados de forma imediata com a intervenção realizada, adotando critérios de rotatividade, absenteísmo, comprometimento, motivação e satisfação; outra categoria é a de desempenho organizacional que inclui aspectos mais distantes da prática organizacional do que os de nível individual, incluindo resultados em um nível mais macro e cujos esforços individuais são agregados em indicadores de produtividade, qualidade dos produtos e serviço; a terceira categoria é referente aos aspectos financeiros e contábeis como retorno de investimentos (ROI) e de ativos (ROA); por último, têm-se os critérios mais distantes referentes aos resultados do mercado, como o crescimento de vendas e seus retornos.

Critérios como estes são, então, considerados em diferentes dimensões de avaliação de efetividade de GD. Cardy & Dobbins (1986), como precursores, postulam que as dimensões de erros do avaliador, precisão da avaliação e aspectos qualitativos da avaliação deveriam ser considerados para a avaliação da efetividade de GD. Estas três dimensões de avaliação passam a predominar durante certo período, porém começa-se a perceber que não bastava somente um sistema de GD psicometricamente perfeito, mas era necessário que ele fosse percebido pelos participantes do processo como justo, válido, útil, entre outros critérios (Levy & Williams, 2004). Neste sentido, a dimensão de reação dos avaliados para avaliação dos resultados do sistema de GD passa a ser considerada como uma dimensão de avaliação. Atualmente, a dimensão de reação conseguiu superar,

inclusive, o interesse anterior pelos aspectos psicométricos, e predomina os estudos sobre o tema (Ikramullah et al., 2016).

No estudo de Colakoglu, Lepak & Hong (2006), os autores concluem que a gama de dimensões de avaliação e de critérios de efetividade adotados pelos estudos tem variado em razão da (1) sua distância ou proximidade com os resultados do empregado, e (2) o foco do grupo de interesse. Isto é: os critérios variam conforme à distância ou proximidade que estes possuem com os resultados desejados, os quais se alteram de acordo com o ator-chave. Logo, suponha que o resultado desejado pelo gestor de GP é melhorar a percepção do empregado sobre o sistema de GD, é importante que ele compreenda que existem critérios específicos para avaliar a reação dos avaliados que podem alavancar mais ou menos este objetivo.

Apesar da linha aparentemente histórica e da narrativa estabelecida, é possível perceber que a adoção de dimensões e critérios de avaliação da efetividade do sistema de GD não consistem na superação de um por outro, mas sim em um processo de desenvolvimento em que os critérios são incluídos a partir de sua contribuição para os modelos de avaliação da efetividade. Além disto, há ainda maneiras de se avaliar a efetividade de uma prática organizacional que não foram ainda sistematizados.

Logo, mesmo diante da tendência de integração, é de esperar que determinados critérios se sobressaiam ou apareçam com maior frequência nos modelos de avaliação, sendo este o caso do critério de reação dos funcionários, o qual tem assumido um papel predominante nos estudos mais recentes. De toda forma, mesmo com a predominância de determinados critérios ao longo do tempo ou com o desenvolvimento de estudos mais recentes que busquem integrá-los, a produção científica sobre o tema ainda não apresenta um consenso acerca da condução de uma avaliação que seja capaz de identificar de forma clara e consistente a efetividade do sistema de GD.

Diante disto e das possibilidades distintas de avaliação da efetividade, percebe-se a necessidade de sistematização destes critérios de forma a permitir que estudos e organizações consigam identificar novas possibilidades de mensurar a avaliação da efetividade ou como o campo de Administração pode contribuir com novas abordagens de avaliação da efetividade de GD.

Métodos, técnicas e procedimentos de pesquisa

Com intuito de sistematizar dimensões e critérios utilizados nos modelos dos estudos dedicados à avaliação da efetividade de sistemas de GD, empreendeu-se uma revisão sistemática de literatura, a qual consiste em identificar, avaliar e interpretar o conteúdo disponível sobre uma questão de pesquisa, tópico ou fenômeno de interesse (Kitchenham, 2004). Em relação à operacionalização da revisão, foram estabelecidos critérios de seleção na identificação dos estudos. Não tendo sido identificada qualquer estudo de revisão sobre o tema que incluísse a produção brasileira, optou-se por não restringir as buscas por período com intuito de abarcar toda a produção. Buscando a completude dos dados e para abordar o campo de forma mais abrangente possível, a busca foi realizada nas bases de dados: *Scopus* e *Web of Science* – com resultados apenas para os descritores em inglês – e Scielo, Spell, Periódicos Capes e Google Acadêmico – para os resultados para os termos em português, buscando ampliar a busca em mais bases de dados do que os descritores em língua estrangeira.

Para fins da revisão, considerou-se os termos AD e GD de forma intercambiável, ou seja, apesar da diferença existente e indicada no início deste artigo, considerou-se por bem utilizar ambos os termos nas buscas como forma de abranger todos os estudos que abordem o tema da efetividade para evitar exclusão daqueles que porventura adotassem os termos como semelhantes. A busca foi feita no título, resumo e palavras-chaves com

os seguintes descritores, inicialmente, na língua inglesa: “*performance management*”, “*performance appraisal*” e “*performance evaluation*” combinados com o termo “*effectiveness*”; e, em seguida, em português: “gestão de desempenho”, “avaliação de desempenho” e “estágio probatório” em conjunto a palavra “efetividade”.

Para combinação dos termos, foram adotados operadores *booleanos*, porém, estes não retornaram bons resultados em alguns momentos, dando-se preferência pela utilização dos termos de forma isolada para ampliar os achados; além disso, algumas bases de dados não suportaram a aplicação dos termos de forma exata, tornando necessária a utilização combinada, o que aumentou a quantidade e a poluição dos resultados.

Ao todo, foram obtidos mais de 47 mil resultados de busca, sem a aplicação de filtros e com estudos duplicados entre uma base e outra. Destes resultados, quase sua totalidade se distanciaram dos termos desejados de pesquisa. Em razão disso, optou-se por classificar os resultados de busca por relevância, o que permitiu identificar mais precisamente os artigos que possuíam mais relação com os termos pesquisados, tendo a coleta dos dados sido finalizada em cada base de dados quando os achados já não apresentavam mais relação com o termo pesquisado.

Para seleção dos artigos, foi aplicado o primeiro filtro, ainda durante a busca, por meio da leitura do título, resumo e palavras-chaves. Com isto, este primeiro filtro resultou em 153 artigos internacionais e nacionais. Retirando-se duplicidades, restaram 120 artigos.

Neste ponto de uma revisão sistemática de literatura, é comum que, para otimizar a seleção dos artigos relevantes, seja aplicado, um protocolo específico e reconhecido (e.g. *Methodi in Ordination* – Pagani, Kovalski e Resende, 2015; *Preferred Reporting Items for Systematic Protocol Review and Meta-Analysis* – PRISMA, 2009; Diretrizes do *Cochrane Handbook de Higgins and Green*, 2011); todavia, com receio de excluir estudos importantes da revisão se aplicado tais protocolos, principalmente aqueles baseados no fator de impacto do periódico, e tendo em vista a quantidade já reduzida de artigos selecionados, optou-se por avaliar a pertinência dos estudos de forma qualitativa com relação à adequação ao objeto de interesse.

Foi feita uma leitura inicial de forma dinâmica dos artigos selecionados até então. Após isso, foram excluídos aqueles que não tratavam da avaliação de efetividade dos sistemas de GD e os artigos de eventos, assim como trabalhos de finalização de curso por serem tidos como trabalhos ainda em desenvolvimento. Diante da amostra final de artigos, foi realizado o fichamento e a compilação dos dados que permitiu a identificação das dimensões e das categorias. Durante esta etapa, a identificação foi feita, inicialmente, a partir da declaração dos autores dos artigos revisados, porém, quando isso não foi possível por falta de clareza, os autores desta pesquisa analisaram o instrumento de pesquisa aplicado.

As dimensões e categorias compiladas foram, então, sistematizados por meio de análise de conteúdo pré e pós-categórica (Bardin, 2011) que permitiu a identificação das dimensões de avaliação da efetividade de sistemas de GD com base na similaridade. Destaca-se que, diante da ausência de sistematização das dimensões e dos critérios de avaliação específicos de GD, esta revisão buscou, a princípio, inspiração na literatura de treinamento, a qual possui modelos de avaliação já consolidados (e.g. Abbad, 1999; Borges-Andrade, 1982; Hamblin, 1978; Kirkpatrick, 1987). Com isso, foram propostas as seguintes categorias: reação, aprendizagem, transferência (ou comportamento no cargo) e impacto. Porém, ao longo dos resultados da revisão, houve a identificação do estudo de Schleicher et al. (2019) em que os autores propõem um modelo de efetividade de GD, fundamentado em uma extensa revisão da literatura, que considera, entre as

dimensões de seu o modelo, as mesmas categorias de reação, aprendizagem, transferência, resultados operacionais e resultados financeiros. Tal achado auxiliou a validar as dimensões até então escolhidas para sistematizar a produção sobre avaliação de efetividade de sistemas de GD. Além destas cinco dimensões, a revisão específica dos critérios de avaliação da efetividade de GD indicou uma sexta dimensão, de resultados sociais.

Com isso, a produção foi sistematizada em seis categorias de dimensões de avaliação da efetividade de GD: Reação, Aprendizagem, Transferência, Resultados Operacionais, Resultados Financeiros e Resultados Sociais. Para cada dimensão também foram sistematizados os seus respectivos critérios de avaliação, aqui representados pelas medidas ou variáveis. Tendo em vista estas categorias, buscou-se demonstrar as características metodológicas aplicadas em cada categoria identificada.

Resultados e discussão

A revisão identificou 49 artigos, conforme os critérios de busca demonstrados anteriormente, referentes à avaliação da efetividade de sistemas de GD. As dimensões identificadas estão sistematizadas na Tabela 1 com suas respectivas definições, assim como a indicação dos autores e a quantidade deles que adotaram estas dimensões.

Todos os estudos revisados adotaram, ao menos, uma das categorias de dimensão propostas de avaliação da efetividade de sistemas de GD. De forma paralela, alguns estudos aplicaram mais de uma dimensão de avaliação, figurando em duas ou mais dimensões.

A dimensão de Reação foi a categoria adotada pela maior quantidade de estudos (24 artigos). Com a diferença de apenas um estudo em relação à dimensão de Reação, têm-se a de Aprendizagem, a qual foi por uma quantidade também elevada de artigos, 23. Já a dimensão de Resultados Operacionais, foi a terceira dimensão mais adotada pelos estudos (20), seguida da dimensão de Transferência, com 18. As quatro primeiras dimensões mais adotadas pelos estudos revisados não possuem quantidades tão díspares entre si. Este grupo de dimensões inclui, principalmente, critérios de nível individual, no caso da Reação e Aprendizagem; de Transferência, a nível de unidade; e no caso dos Resultados Operacionais, a nível organizacional. Já as dimensões de Resultados Financeiros e de Resultados Sociais figuram como um grupo com menor frequência entre os estudos revisados, com, respectivamente, sete e quatro artigos, e abordam, em especial, critérios de nível organizacional.

As dimensões de avaliação predominantes, principalmente, a de Reação, são achados convergentes com a tendência apontada por estudos mais recentes (e.g. Denisi & Smith, 2014; Ikramullah et al., 2016; Pichler, 2012), assim como o foco principal em aspectos individuais e da unidade e não mais exclusivamente focalizado em aspectos organizacionais, como normalmente feito anteriormente (Colakoglu et al., 2006).

Com relação aos critérios de avaliação de efetividade de GD identificados, nenhum foi majoritariamente adotado em uma quantidade representativa de estudos da amostra que indicasse um padrão, porém, o critério de Justiça (Reação) e de Habilidade (Aprendizagem) foram os que figuraram com maior recorrência entre os estudos revisados (14).

Tabela 1. Dimensões de avaliação de efetividade de sistema de GD

Dimensão	Definição	Autores	Critério	Definição	Autores	Quantidade
Reação	Sensação ou sentimento do avaliado acerca do sistema de GD	Abdullah & Abraham (2012); Ahmed (1999); Awan, Habib, Ahktar & Naveed (2020); Baird, Schoch & Chen (2012); Costigan, Insinga, Berman, Ilter, Kranas & Kureshov (2009); Da Luz & Figueiredo Filho (2018); Dhas, Joseph & Krishnan (2015); Iqbal, Akbar, Budwar & Shah (2019); Ishaq, Mansor, Khan & Ahmed (2013); Kakkar, Dash, Vohra & Saha (2020); Kim (2016); Kivipõld, Türk & Kivipõld (2020); Longenecker & Nykodym (1996); Longenecker, Liverpool & Wilson (1988); Modipane, Botha & Blom (2019); Phuong (2018); Pinto & Behr (2015); Ponmuthumari (2020); Roberts (1996); Schleicher et al. (2019); Selden & Sowa (2011); Sharma, Sharma & Agarwal (2016); Tuytens & Devos (2012); Waeyenberg & Decramer (2018)	Justiça	Quão justo é o sistema de GD	Ahmed (1999); Awan, Habib, Ahktar & Naveed (2020); Ishaq, Mansor, Khan & Ahmed (2013); Kakkar, Dash, Vohra & Saha (2020); Kim (2016); Kivipõld, Türk & Kivipõld (2020); Longenecker & Nykodym (1996); Modipane, Botha & Blom (2019); Phuong (2018); Roberts (1996); Schleicher et al. (2019); Selden & Sowa (2011); Sharma, Sharma & Agarwal (2016); Tuytens & Devos (2012)	14
			Utilidade	Utilidade ou valor do sistema de GD	Abdullah & Abraham (2012); Ahmed (1999); Baird, Schoch & Chen (2012); Boroujerdi, Baradaran & Habibi (2016); Cislaghi & Zugno (2019); Dhas, Joseph & Krishnan (2015); Maley & Moeller (2014); Pinto & Behr (2015); Schleicher et al. (2019); Teeroovengadum, Nunkoo & Dulloo (2019); Tuytens & Devos (2012); Waeyenberg & Decramer (2018); Yu, Baird & Tung (2018)	13
			Satisfação	Satisfação geral com o sistema de GD	Abdullah & Abraham (2012); Ahmed (1999); Dhas, Joseph & Krishnan (2015); Kim (2016); Iqbal, Akbar, Budwar & Shah (2019); Pinto & Behr (2015); Ponmuthumari (2020); Waeyenberg & Decramer (2018); Schleicher et al. (2019)	9

			Precisão	Quão preciso é o sistema de GD	Awan, Habib, Ahkter & Naveed (2020); Costigan, Insinga, Berman, Ilter, Kranas & Kureshov (2009); Kakkar, Dash, Vohra & Saha (2020); Modipane, Botha & Blom (2019); Roberts (1996); Schleicher et al. (2019); Sharma, Sharma & Agarwal (2016)	7	
			Afetividade	Sentimentos relacionados ao sistema de GD	Baird, Schoch & Chen (2012); Costigan, Insinga, Berman, Ilter, Kranas & Kureshov (2009); Da Luz & Figueiredo Filho (2018); Longenecker, Liverpool & Wilson (1988); Phuong (2018); Pinto & Behr (2015); Schleicher et al. (2019)	5	
Aprendizagem	Conhecimentos, habilidades e atitudes adquiridos após a intervenção do sistema de GD	Ahmed (1999); Akhtar & Mittal (2010); Akhtar & Sishil (2018); Baird, Schoch & Chen (2012); Boroujerdi, Baradaran & Habibi (2016); Cislighi & Zugno (2019); Dewettinck & van Dijk (2013); Dewettinck (2008); Guimarães, Nader & Ramagem (1998); Ikramullah, Van Prooijen, Iqbal & Ul-Hassan (2016); Iqbal, Akbar & Budwar (2015); Iqbal, Akbar, Budwar & Shah (2019); Ishaq, Iqbal & Zaheer (2009); Ishaq, Mansor, Khan & Ahmed (2013); Kand & Choi (2019); Longenecker & Nykodym (1996); Longenecker, Liverpool & Wilson (1988); Schleicher et al. (2019); Soltani, Van der Meer, Gennard & Williams (2003); Selden & Sowa (2011);	Habilidade	Desenvolvimento individual de habilidades	Ahmed (1999); Akhtar & Mittal (2010); Akhtar & Sishil (2018); Baird, Schoch & Chen (2012); Cislighi & Zugno (2019); Dewettinck & van Dijk (2013); Dewettinck (2008); Iqbal, Akbar, Budwar & Shah (2019); Ishaq, Mansor, Khan & Ahmed (2013); Schleicher et al. (2019); Soltani, Van der Meer, Gennard & Williams (2003); Selden & Sowa (2011); Teeroovengadum, Nunkoo & Dulloo (2019); Yang & Hsieh (2007); Yu, Baird & Tung (2018);	14	23
			Conhecimento	Desenvolvimento individual de conhecimentos (declarativo, procedimental e tácito)	Baird, Schoch & Chen (2012); Ikramullah, Van Prooijen, Iqbal & Ul-Hassan (2016); Kand & Choi (2019); Dewettinck & van Dijk (2013); Ishaq, Iqbal & Zaheer (2009);	9	

		Teeroovengadum, Nunkoo & Dulloo (2019); Yang & Hsieh (2007); Yu, Baird & Tung (2018)			Iqbal, Akbar, Budwar & Shah (2019); Longenecker, Liverpool & Wilson (1988); Schleicher et al. (2019); Yu, Baird & Tung (2018)		
			Atitude	Mudanças atitudinais voltadas para o trabalho	Boroujerdi, Baradaran & Habibi (2016); Dewettinck & van Dijk (2013); Guimarães, Nader & Ramagem (1998); Ikramullah, Van Prooijen, Iqbal & Ul-Hassan (2016); Iqbal, Akbar & Budwar (2015); Iqbal, Akbar, Budwar & Shah (2019); Ishaq, Iqbal & Zaheer, (2009); Longenecker & Nykodym (1996); Longenecker, Liverpool & Wilson (1988); Schleicher et al. (2019)	7	
Resultados Operacionais	Melhoria em aspectos de operação organizacional	Ahmed (1999); Akhtar & Mittal (2010); Akhtar & Sishil (2018); Baird, Schoch & Chen (2012); Biulchi & Pauli (2012); Dhas, Joseph & Krishnan (2015); Haines & St-Onge (2012); Ikramullah, Van Prooijen, Iqbal & Ul-Hassan (2016); Ishaq, Iqbal & Zaheer (2009); Kand & Choi (2019); Longenecker & Nykodym (1996); Longenecker, Liverpool & Wilson (1988); Nankervis, Staton & Foley (2012); Schleicher et al. (2019); Soltani, Van der Meer, Gennard & Williams (2003); Stanton & Nankervis (2011); Teeroovengadum, Nunkoo & Dulloo (2019); Thompson &	Produtividade	Melhoria na capacidade produtiva	Ahmed (1999); Biulchi & Pauli (2012); Dhas, Joseph & Krishnan (2015); Kand & Choi (2019); Longenecker, Liverpool & Wilson (1988); Nankervis, Staton & Foley (2012); Soltani, Van der Meer, Gennard & Williams (2003); Schleicher et al. (2019); Stanton & Nankervis (2011); Thompson & Davies (2020); Yang & Hsieh (2007)	11	20
			Outros resultados previstos estrategicamente	Alcance de outros resultados previstos estrategicamente	Ahmed (1999); Akhtar & Mittal (2010); Akhtar & Sishil (2018); Baird, Schoch & Chen (2012); Ikramullah, Van Prooijen, Iqbal & Ul-Hassan (2016); Teeroovengadum, Nunkoo & Dulloo (2019); Yu, Baird & Tung (2018)	7	

		Davies (2020); Yang & Hsieh (2007); Yu, Baird & Tung (2018)	Inovação	Desenvolvimento de novas ferramentas	Akhtar & Mittal (2010); Haines & St-Onge (2012); Nankervis, Staton & Foley (2012); Schleicher et al. (2019); Stanton & Nankervis (2011)	5	
			Rotatividade	Melhoria na taxa de funcionários que entram e saem da organização	Longenecker & Nykodym (1996); Ishaq, Iqbal & Zaheer (2009); Schleicher et al. (2019)	3	
			Força de trabalho	Melhoria na qualidade e quantidade da força de trabalho	Nankervis, Staton & Foley (2012); Schleicher et al. (2019); Stanton & Nankervis (2011)	3	
			Absenteísmo	Melhoria na taxa de funcionários que deixam de ir trabalhar	Schleicher et al. (2019)	1	
Transferência	Mudanças de atitude e comportamento referentes ao papel desempenhado no cargo	Ahmed (1999); Baird, Schoch & Chen (2012); Dewettinck & van Dijk (2013); Dewettinck (2008); Dhas, Joseph & Krishnan (2015); Haines & St-Onge (2012); Ishaq, Iqbal & Zaheer (2009); Longenecker & Nykodym (1996); Longenecker, Liverpool & Wilson (1988); Maley & Moeller (2014); Monsur & Akkas (2015); Phuong (2018); Ponmuthumari (2020); Roberts (1992); Schleicher et al. (2019); Soltani, Van der Meer, Gennard & Williams (2003); Teeroovengadam, Nunkoo & Dulloo (2019); Yu, Baird & Tung (2018)	Desempenho	Melhoria no desempenho de tarefa	DeNisi & Murphy (2017); DeNisi & Pritchard (2006)	12	18
			Motivação	Melhoria da motivação	Ahmed (1999); Baird, Schoch & Chen (2012); Dewettinck & van Dijk (2013); Dewettinck (2008); Ishaq, Iqbal & Zaheer, 2009; Longenecker & Nykodym (1996); Longenecker, Liverpool & Wilson (1988); Maley & Moeller (2014); Monsur & Akkas (2015); Schleicher et al. (2019); Soltani, Van der Meer, Gennard & Williams (2003); Teeroovengadam, Nunkoo & Dulloo (2019); Yu, Baird & Tung (2018)	12	
			Relação interpessoal	Melhoria nas relações interpessoais	Dewettinck (2008); Longenecker, Liverpool & Wilson (1988);	5	

					Ponmuthumari (2020); Roberts (1992); Schleicher et al. (2019)		
			Bem-estar	Melhoria no clima organizacional	Dhas, Joseph & Krishnan (2015); Dewettinck & van Dijk (2013); Schleicher et al. (2019)	3	
Resultados Sociais	Melhoria em indicadores referentes ao público-alvo do negócio	Akhtar & Mittal (2010); Biulchi & Pauli (2012); Ikramullah, Van Prooijen, Iqbal & Ul-Hassan (2016); Maley & Moeller (2014); Pinto & Behr (2015); Yang & Hsieh (2007); Zhang, Tantardini, Kim & Julnes (2020)	Desempenho empresarial/governamental	Melhoria do desempenho da organização/governo perante seu público-alvo	Akhtar & Mittal (2010); Biulchi & Pauli (2012); Ikramullah, Van Prooijen, Iqbal & Ul-Hassan (2016); Maley & Moeller (2014); Pinto & Behr (2015); Yang & Hsieh (2007); Zhang, Tantardini, Kim & Julnes (2020)	7	7
Resultados Financeiros	Melhoria em aspectos de finanças	Akhtar & Mittal (2010); Nankervis, Stanton & Foley (2012); Schleicher et al. (2019); Stanton & Nankervis (2011); Zhang, Tantardini, Kim & Julnes (2020)	Outros resultados financeiros	Outros resultados como crescimento de vendas, aumento do lucro, etc	Akhtar & Mittal (2010); Nankervis, Stanton & Foley (2012); Schleicher et al. (2019)	5	5
			Retorno sobre o investimento	Aumento da taxa de Retorno sobre investimento	Nankervis, Stanton & Foley (2012); Schleicher et al. (2019); Stanton & Nankervis (2011);	3	
			Competividade	Aumento da participação no mercado	Nankervis, Stanton & Foley (2012); Schleicher et al. (2019); Stanton & Nankervis (2011)	3	

Nota. Fonte: elaborada pelos autores.

Neste sentido, a dimensão de Reação, relativa à sensação ou sentimento do participante do processo de GD acerca do sistema, teve como a categoria mais recorrente nos estudos a Justiça (14 artigos), a qual se refere ao quanto os participantes do processo de GD entendem que o sistema é justo. Este critério foi consolidado, principalmente nos últimos anos, por meio da construção de escalas específicas, como a de Sharma et al. (2016). Na dimensão de Aprendizagem, referente aos conhecimentos, habilidades e atitudes adquiridos após a intervenção do sistema de GD, o desenvolvimento de Habilidades figura como o critério mais adotado pelos estudos revisados, seguido dos Conhecimentos (9 artigos) e da Atitude (7 artigos).

Acerca da dimensão de Resultados Operacionais, que aborda as melhorias em aspectos da operação e execução à nível da organização, a Produtividade predomina como critério entre os estudos revisados (11). Na dimensão de Transferência, a qual apontam as mudanças de atitude e comportamento no desempenho do cargo, o critério mais adotado pelos estudos revisados (12) é o de Desempenho, resultando em melhorias na execução das tarefas previstas.

A dimensão de Resultados Sociais, identificada de forma pós-categórica, está representada por apenas uma categoria de avaliação que se refere à melhoria de desempenho organizacional/ governamental perante o seu público-alvo. Por último, os Resultados Financeiros abordam principalmente indicadores de finanças da organização, como crescimento de vendas, aumento de lucro, solvência financeira, entre outros.

Os critérios marginalizados, ou seja, aqueles menos adotados pelos estudos representam possibilidades de pesquisa, principalmente se adotados de forma a complementar com outros critérios mais consolidados, como o de Justiça, na dimensão de Reação.

Posto isto, entende-se que as características metodológicas são definidas em razão, essencialmente, da dimensão adotada de avaliação da efetividade de GD. Logo, a predominância de determinadas características metodológicas se justifica pela adoção de uma ou outra dimensão, assim como de critérios indicados anteriormente, utilizados pelos estudos para mensurar a efetividade de GD. Neste sentido, tendo em vista que a efetividade de GD pode ser avaliada por diferentes dimensões, critérios e procedimentos de pesquisa e que o seu estudo tem sido prejudicado pela falta de articulação entre eles (Schleicher et al., 2019), buscou-se a identificação das características metodológicas relacionadas às dimensões adotadas, conforme disposto na Tabela 2.

Diante dos resultados metodológicos identificados, percebe-se a predominância dos estudos empíricos em todas as dimensões, resultado explicado, em partes, pelo objeto de pesquisa desta revisão, o qual discorre, principalmente, acerca da aplicação ou adoção de dimensões e critérios de avaliação da efetividade. Destaca-se que as dimensões de Reação, Transferência e Resultados Financeiros não foram apontadas em nenhum estudo teórico.

Tabela 2. Características metodológicas das dimensões de avaliação da efetividade de sistemas de GD

		DIMENSÃO DE AVALIAÇÃO DA EFETIVIDADE DE GD											
		Reação		Aprendizagem		Transferência		Resultados Operacionais		Resultados Financeiros		Resultados Sociais	
		Quantidade											
Tipificação	Categoria	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
Natureza	Teórica-empírica	24	100	21	91	18	100	18	90	5	100	6	86
	Teórica	0	0	2	9	0	0	2	10	0	0	1	14
Origem dos dados	Primária	20	83	16	76	17	94	16	89	3	60	3	43
	Secundária	2	8	3	14	1	6	2	11	2	40	3	43
	Primária e secundária	2	8	2	10	0	0	0	0	0	0	1	14
Abordagem	Qualitativa	1	4	2	10	2	11	2	11	1	20	3	50
	Quantitativa	16	67	15	71	15	83	14	78	2	40	2	33
	Mista	5	21	4	19	1	6	2	11	2	40	1	17
Recorte temporal	Transversal	23	96	19	90	17	94	17	94	3	60	5	83
	Longitudinal	1	4	2	10	1	6	1	6	2	40	1	17
Amostra	Probabilística	16	67	15	71	13	72	12	67	3	60	3	50
	Não probabilística	5	21	4	19	4	22	5	28	1	20	2	33
	Teórica	2	8	2	10	1	6	0	0	1	20	1	17
	Censitária	1	4	0	0	0	0	0	0	0	0	0	0
Método	Survey	20	83	17	81	16	89	15	83	2	40	2	33
	Estudo de Caso	3	13	2	10	1	6	1	6	2	40	3	50
	Pesquisa documental	1	4	2	10	1	6	2	11	1	20	1	17
Instrumento	Questionários e escalas	22	69	18	72	16	84	16	84	3	75	3	38
	Roteiro de entrevista	5	16	3	12	2	11	1	5	0	0	2	25
	Roteiros de observação	1	3	0	0	0	0	0	0	0	0	0	0
	Auto registros	4	13	4	16	1	5	2	11	1	25	3	38

Nota. Fonte: elaborado pelos autores.

Com relação à origem dos dados, esta não varia entre as dimensões, sendo predominante a fonte primária em todas as dimensões identificadas. Destaca-se a participação reduzida dos dados de origem secundária ou mista (primária e secundária), o que sugere uma possibilidade das pesquisas de complementação de seus resultados com dados secundários, principalmente os produzidos internamente às organizações estudadas.

A abordagem de pesquisa predominante entre todas as dimensões é a quantitativa, quase não havendo a presença de pesquisas qualitativas ou de abordagem mista. O recorte temporal é predominantemente transversal, mas possui alguns estudos longitudinais em todas as dimensões, o que é interessante para o campo em questão, pois séries históricas podem auxiliar na comparação e aferição de resultados. Acerca da amostra dos estudos revisados, a probabilística predomina entre todas as dimensões. Destaca-se um estudo censitário na dimensão de Reação em que os autores Modipane, Botha & Blom (2019) identificaram a percepção sobre a efetividade de sistemas de GD com base nos critérios de justiça e de precisão com todos os servidores públicos de uma província.

Identificou-se a triangulação de dados em nove estudos da amostra dos 49 artigos revisados, ou seja, fazendo uso de mais de uma técnica de pesquisa, o que é favorável para o campo, pois visa suprir imperfeições de um método com a combinação de outro (Denzin & Lincoln, 2011). Entre os estudos que aplicaram mais de uma técnica, oito mesclaram técnicas de abordagem qualitativa e quantitativa e dois fizeram uso somente da abordagem qualitativa.

Com relação ao método de pesquisa adotado, o método survey foi predominante em todas as dimensões identificadas, seguido de estudo de caso e pesquisa documental. Nenhum outro método de pesquisa além destes três foi aplicado nos estudos revisados, indicando que a área de estudo é dominada por levantamentos de opinião dos participantes do processo de GD. Isto é confirmado quando comparado com o resultado referente à aplicação de instrumentos de pesquisa em que a maioria dos estudos, em todas as dimensões, adotam instrumentos como questionários, porém, com escalas perceptuais, mesmo em dimensões ou critérios de avaliação (e.g. produtividade, rotatividade, ROI, ROA etc.) que caberiam a aplicação de escalas objetivas.

Por fim, as análises estatísticas básicas e inferenciais também se mostram dominantes, até mesmo pelos padrões metodológicos demonstrados anteriormente, sendo a análise de conteúdo a outra única análise aplicada e em minoria entre todas as dimensões.

Portanto, percebe-se uma tendência dos estudos em avaliar a efetividade de GD com dados primários a partir de levantamentos com os participantes do processo de GD por meio de, principalmente, questionário e de entrevistas. Este cenário é convergente com o que já fora apontado por Cardy e Dobbins (1986), Levy e Williams (2004) e Schleicher et al. (2019), os quais afirmaram que a produção sobre efetividade de GD é marcadamente avaliada sob a perspectiva da percepção dos participantes da GD.

Chama a atenção a preferência dos autores de quase toda a amostra dos estudos revisados em avaliar a efetividade de GD a partir de critérios mensurados de forma perceptual, mesmo quando os critérios se mostram passíveis de avaliação com métricas e medidas objetivas. Pode-se atribuir a decisão por uma ou outra forma de mensuração dos critérios adotados em razão da interface estabelecida a partir da década de 1960 entre a área de Administração e a Psicologia, o que fez com que a área de GP se concentrasse, principalmente, no comportamento humano na organizações e em uma atuação voltada

para a vertente humanista-desenvolvimentista (Schmidt et al., 2013). Com isso, impera sobre a atuação da área de GP um forte arsenal da Psicologia e da psicometria que se reflete na condução de instrumentos, métodos e procedimentos da GP (Fischer et al., 2002).

Logo, percebe-se a influência predominante das vertentes psicológicas nos estudos sobre GP em Administração, demonstrando o reflexo na forma praticamente absoluta de mensurar a efetividade de GD a partir de medidas perceptuais e sob a ótica humanista-desenvolvimentista.

Portanto, destaca-se a necessidade de um modelo de avaliação da efetividade de sistemas de GD integrador que consiga relacionar (1) a prática já adotada pelo campo de pesquisa por estudos perceptuais com outras possibilidades metodológicas de pesquisa e (2) as dimensões e os critérios menos adotados, os quais podem clarear, tanto para acadêmicos, como praticantes, a relação entre os critérios de avaliação, favorecendo o alcance de objetivos mais distantes (e.g. resultados financeiros) por meio da consecução de objetivos mais próximos.

Conclusão

Este estudo objetivou sistematizar dimensões e critérios utilizados nos modelos dos estudos dedicados à avaliação da efetividade de sistemas de GD em razão da fragmentação do campo. O objetivo foi cumprido por meio de extensa e meticulosa revisão sistemática de literatura que resultou em uma amostra de 49 artigos nacionais e internacionais identificados.

A produção foi sistematizada em seis dimensões de avaliação da efetividade de sistemas de GD: Reação, Aprendizagem, Transferência, Resultados Operacionais, Resultados Financeiros e Resultados Sociais. Além das dimensões, também foram identificados os critérios de Justiça e de Habilidades com os critérios mais adotados pelos estudos revisados para avaliação da efetividade de GD. Entretanto, percebeu-se que diversos outros critérios são ignorados pelos estudos com recorrência pouco expressiva entre eles.

Outro achado identificado nesta revisão refere-se à operacionalização dominante das avaliações de efetividade por meio de levantamento de opinião com escalas perceptuais, mesmo quando as dimensões e critérios de avaliação permitiram outra forma de operacionalização metodológica. Assim, percebe-se um viés dos estudos da Psicologia no padrão metodológico de realização dos estudos desta temática de GP na Administração.

Portanto, sugere-se que a partir desta conclusão e das dimensões e categorias sistematizadas, estudos futuros possam buscar integrar diferentes ferramentas e tecnologias da Administração para promover uma avaliação de efetividade de GD que consiga estabelecer uma relação entre os resultados mais acessíveis à área de GP que gere efeito nos resultados da organização. Destaca-se como limitação os descritores restritos à GD e AD, mas que podem ser ampliados para mensuração dos resultados de GEP; e a amostra que não incluiu os artigos provenientes de eventos, livros, dissertações, teses e anais por serem tidos como artigos em aperfeiçoamento.

Referências

- Abbad, G. (1999). Um modelo integrado de avaliação de impacto de treinamento no trabalho. Tese (Doutorado) – Instituto de Psicologia, Universidade de Brasília, Brasil.
- Ahmed, S. (1999). The emerging measure of effectiveness for human resource

management: An exploratory study with performance appraisal. *Journal of Management Development*, 18(6), 543–556. <https://doi.org/10.1108/02621719910279644>

Awan, S. H., Habib, N., Shoaib Akhtar, C., & Naveed, S. (2020). Effectiveness of Performance Management System for Employee Performance Through Engagement. *SAGE Open*, 10(4), 215824402096938. <https://doi.org/10.1177/2158244020969383>

Baird, K., Schoch, H., & Chen, Q. (James). (2012). Performance management system effectiveness in Australian local government. *Pacific Accounting Review*, 24(2), 161–185. <https://doi.org/10.1108/01140581211258461>

Becker, B. E., & Huselid, M. a. (1998). High Performance Work Systems and Firm Performance : a Synthesis of Research and. *Research in Personnel and Human Resources Management*, 16, 53–101.

Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, 71(4), 672–678. <https://doi.org/10.1037//0021-9010.71.4.672>

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, 83(4), 615–633. <https://doi.org/10.1037/0021-9010.83.4.615>

Colakoglu, S., Lepak, D. P., & Hong, Y. (2006). Measuring HRM effectiveness: Considering multiple stakeholders in a global context. *Human Resource Management Review*, 16(2), 209–218. <https://doi.org/10.1016/j.hrmmr.2006.03.003>

Den Hartog, D. N., Boselie, P., & Paauwe, J. (2004). Performance management: A model and research agenda. *Applied Psychology*, 53(4), 556–569. <https://doi.org/10.1111/j.1464-0597.2004.00188.x>

DeNisi, A. S., & Murphy, K. R. (2017). Performance Appraisal and Performance Management: 100 Years of Progress? *Journal of Applied Psychology*, 102(3), 421–433. <https://doi.org/10.1037/apl0000085.supp>

DeNisi, A. S., & Pritchard, R. D. (2006). Performance Appraisal, Performance Management and Improving Individual Performance: A Motivational Framework. *Management and Organization Review*, 2(2), 253–277. <https://doi.org/10.1111/j.1740-8784.2006.00042.x>

Denisi, A. S., & Smith, C. E. (2014). Performance Appraisal, Performance Management, and Firm-Level Performance: A Review, a Proposed Model, and New Directions for Future Research. *Academy of Management Annals*, 8(1), 127–179. <https://doi.org/10.5465/19416520.2014.873178>

Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage handbook of qualitative research*.

Dyer, L., & Reeves, T. (1995). Human resource strategies and firm performance: What do we know and where do we need to go? *The International Journal of Human Resource Management*, 6(3), 656–670. <https://doi.org/10.1080/09585199500000041>

Fischer, T., Heber, F., Fadul, É., & Fachin, R. (2002). Capacitação avançada em regulação: desafios institucionais às interorganizações do setor de energia elétrica no Brasil e alternativas críticas à retórica da competência *. *Revista de Administração Pública*, 36(3), 485–506.

Haines, V. Y., & St-Onge, S. (2012). Performance management effectiveness: practices

or context? *International Journal of Human Resource Management*, 23(6), 1158–1175. <https://doi.org/10.1080/09585192.2011.561230>

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, 38(3), 635–672. <https://doi.org/10.2307/256741>

Ikramullah, M., Van Prooijen, J. W., Iqbal, M. Z., & Ul-Hassan, F. S. (2016). Effectiveness of performance appraisal: Developing a conceptual framework using competing values approach. *Personnel Review*, 45(2), 334–352. <https://doi.org/10.1108/PR-07-2014-0164>

Jackson, S. E., Schuler, R. S., & Jiang, K. (2014). An Aspirational Framework for Strategic Human Resource Management. *The Academy of Management Annals*, 8(1), 1–56. <https://doi.org/10.1080/19416520.2014.872335>

Jiang, K., & Messersmith, J. (2018). On the shoulders of giants: a meta-review of strategic human resource management. *International Journal of Human Resource Management*, 29(1), 6–33. <https://doi.org/10.1080/09585192.2017.1384930>

Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology*, 85(5), 708–723. <https://doi.org/10.1037/0021-9010.85.5.708>

Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*.

Levy, P. E., & Williams, J. R. (2004). The Social Context of Performance Appraisal: A Review and Framework for the Future. *Journal of Management*, 30(6), 881–905. <https://doi.org/10.1016/j.jm.2004.06.005>

Modipane, P. I., Botha, P. A., & Blom, T. (2019). Employees' perceived effectiveness of the performance management system at a north-west provincial government department. *South African Journal of Childhood Education*, 17. <https://doi.org/10.4102/sajhrm.v17i0.1081>

Pichler, S. (2012). The social context of performance appraisal and appraisal reactions: A meta-analysis. *Human Resource Management*, 51(5), 709–732. <https://doi.org/10.1002/hrm.21499>

Roberts, G. E. (1995). Municipal Government Performance Appraisal System Practices: Is the Whole Less Than the Sum of its Parts? *Public Personnel Management*, 24(2), 197–221. <https://doi.org/10.1177/009102609502400209>

Schleicher, D. J., Baumann, H. M., Sullivan, D. W., & Junhyok, Y. (2019). Evaluating the effectiveness of performance management: A 30-year integrative conceptual review. *Journal of Applied Psychology*, 104(7), 851–887. <https://doi.org/10.1037/apl0000368>

Schmidt, B., Krawulski, E., & Marcondes, R. C. (2013). Psicologia e Gestão de Pessoas em Organizações de Trabalho: investigando a perspectiva estratégica de atuação. *Revista de Ciências Humanas*, 47(2), 344–361.

Sharma, N. P., Sharma, T., & Agarwal, M. N. (2016). Measuring employee perception of performance management system effectiveness: Conceptualization and scale development. *Employee Relations*, 38(2), 224–247. <https://doi.org/10.1108/ER-01-2015-0006>