

O PAPEL DA ADMINISTRAÇÃO E ANÁLISE DE *BIG DATA* COMO HABILITADORAS DA GESTÃO DO DESEMPENHO CORPORATIVO

MAURICIUS M. MEDEIROS¹

 <https://orcid.org/0000-0001-5552-4897>

ANTONIO C. G. MAÇADA¹

 <https://orcid.org/0000-0002-8849-0117>

NORBERTO HOPPEN²

 <https://orcid.org/0000-0003-1827-8695>

Para citar este artigo: Medeiros, M. M., Maçada, A. C. G., & Hoppen, N. (2021). O papel da administração e análise de *big data* como habilitadoras da gestão do desempenho corporativo. *Revista de Administração Mackenzie*, 22(6), 1–32. doi:10.1590/1678-6971/eRAMD210063

Submissão: 15 mar. 2021. **Aceite:** 5 maio 2021.

¹ Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

² Universidade do Vale do Rio dos Sinos (Unisinos), São Leopoldo, RS, Brasil.



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

This paper may be copied, distributed, displayed, transmitted or adapted for any purpose, even commercially, if provided, in a clear and explicit way, the name of the journal, the edition, the year and the pages on which the paper was originally published, but not suggesting that RAM endorses paper reuse. This licensing term should be made explicit in cases of reuse or distribution to third parties.

Este artigo pode ser copiado, distribuído, exibido, transmitido ou adaptado para qualquer fim, mesmo que comercial, desde que citados, de forma clara e explícita, o nome da revista, a edição, o ano e as páginas nas quais o artigo foi publicado originalmente, mas sem sugerir que a RAM endosse a reutilização do artigo. Esse termo de licenciamento deve ser explicitado para os casos de reutilização ou distribuição para terceiros.

RESUMO

Objetivo: A transformação digital e o *big data* (BD) geraram uma verdadeira revolução no gerenciamento orientado a dados. Embora o BD melhore a gestão do desempenho corporativo (GDC), isso também implica aumentar a exposição a riscos em vários estágios do ciclo de vida do BD. À medida que aumentam os requisitos regulatórios e a necessidade de análise do banco de dados em diversas áreas de negócios, é necessário que a organização estabeleça definições, políticas e processos para garantir a qualidade dos dados, a fim de proteger e potencializar seus dados para obter vantagem competitiva. Portanto, compreender a administração de dados (AD) e a *business analytics* (BA) é essencial para o gerenciamento dos negócios. O objetivo deste estudo é analisar o papel do BD, da AD e da BA como habilitadores da GDC.

Originalidade/valor: Contribuímos para a teoria ao conceituarmos, validarmos e discutirmos o construto AD e ao destacarmos o papel dela com a BA na relação entre BD e GDC. As evidências deste estudo indicam que na prática a AD e a BA são um caminho crítico para as organizações obterem um melhor controle dos efeitos que o BD pode ter na GDC.

Design/metodologia/abordagem: Realizou-se uma *survey* com 312 gestores que utilizam *big data analytics* (BDA) em organizações brasileiras. Os dados foram analisados por meio de equações estruturais e testes de mediação.

Resultados: Os resultados sugerem que a administração e a analítica de dados de negócios, tanto isoladamente quanto em conjunto, podem transmitir o efeito BD para a GDC. No entanto, um melhor nível de ajuste do modelo é obtido quando há uma multimediação serializada nesse relacionamento, sendo a AD um antecedente para a BA.

PALAVRAS-CHAVE

Big data. Business analytics. Administração de dados. Governança de dados. Gestão do desempenho corporativo.

1. INTRODUÇÃO

O advento da transformação digital e do *big data* (BD) implica uma oportunidade para a adoção da *business analytics* (BA), a fim de suportar os recursos analíticos necessários para extrair *insights* que podem conduzir a melhores decisões na gestão de negócios (Chen, Chiang, & Storey, 2012; Kitchens, Dobolyi, Li, & Abbasi, 2018; Seddon, Constantinidis, Tamm, & Dod, 2017). Isso tem o potencial de mudar a forma como as empresas competem por meio de uma melhor compreensão do processamento e da exploração de grandes quantidades de dados provenientes de diferentes fontes externas e processos internos (Ferraris, Mazzoleni, Devalle, & Couturier, 2019). Além disso, a estratégia de BD é um recurso significativo para a competitividade, o desempenho, a agilidade e a inovação que está influenciando fortemente a formulação de estratégias de negócio, devido ao aumento de dados e capacidades analíticas (Côrte-Real, Ruivo, & Oliveira, 2019; Dubey, Gunasekaran, & Childe, 2019). Os principais *clusters* que descrevem a literatura sobre *big data analytics* (BDA) estão relacionados à tomada de decisão e ao gerenciamento de desempenho (Rialti, Marzi, Ciappei, & Busso, 2019).

No entanto, a infraestrutura de tecnologia da informação (TI) desatualizada, a complexidade e o caos inerentes ao BD, o gerenciamento, a qualidade e a segurança de dados, a falta de habilidades de ciência de dados nas organizações, as preocupações com a privacidade e as culturas organizacionais que não conduzem a operações orientadas a dados ou à tomada de decisão baseada em dados são as principais barreiras para a implementação eficaz de estratégias de BD. É necessário criar uma visão organizacional clara relacionada ao BD, mas, se a alta administração não valorizar a tomada de decisão orientada por dados, seu comportamento afetará os padrões de decisão em todos os níveis da organização (Alharthi, Krotov, & Bowman, 2017; Tabesh, Mousavidin, & Hasani, 2019). Isso denota a relevância de uma gestão estratégica baseada em dados. A gestão do desempenho corporativo (GDC) combina práticas de gestão e TI para possibilitar o planejamento, a medição e a avaliação da implementação e execução de estratégias organizacionais para alavancar o desempenho dos negócios (Richards, Yeoh, Chong, & Popovič, 2019). A GDC é orientada por dados (Acito & Khatri, 2014), o que requer a integração de uma ampla variedade de fontes de dados externas e internas.

No entanto, sem o controle de dados adequado, as organizações estão mais expostas a riscos em vários estágios do ciclo de vida do BD (Coyne,



Coyne, & Walker, 2018). Nesse contexto, a BA enfrenta alguns desafios relacionados aos dados, como em extrações, oscilações, duplicações e falhas de segurança, cujas irregularidades e inconsistências precisam ser identificadas e corrigidas antes que decisões sejam tomadas com base em dados incorretos (Fleckenstein & Fellows, 2018). Para que possam equilibrar a criação de valor e a exposição ao risco e alcançar a coordenação eficaz necessária para ter sucesso e manter a vantagem competitiva, as organizações precisam de um programa de governança de dados (Coyne et al., 2018; Riggins & Klamm, 2017). No entanto, o desenvolvimento de mecanismos e políticas de governança e administração de dados (AD) é um dos desafios organizacionais contemporâneos mais complexos (Khatri & Brown, 2010; Nielsen, 2017; Tallon, Ramirez, & Short, 2013).

Apesar da importância do tema e da existência de diversos estudos conceituais, as pesquisas empíricas sobre governança e gestão de dados são escassas e fragmentadas, conforme apontado na revisão de literatura realizada por Abraham, Schneider e vom Brocke (2019). E a pesquisa empírica existente sobre AD aborda a questão de uma forma exploratória por meio de pesquisa qualitativa – ver Nokkala, Salmela e Toivonen (2019) e Plomp, Dintzner, Teperek e Dunning (2019). Expressões como governança de dados, gerenciamento de dados e administração de dados às vezes são usadas de forma intercambiável, embora sejam elementos separados. Governança de dados refere-se a políticas e regras organizacionais para resolver problemas ou fornecer serviços às partes interessadas em dados (Harrison et al., 2019; Lillie & Eybers, 2018). Gerenciamento de dados lida com processos organizacionais relacionados ao ciclo de vida dos dados – aquisição, pré-processamento e tratamento, distribuição, exclusão e descarte (Côrte-Real et al., 2019; Fleckenstein & Fellows, 2018). Tais políticas, regras e processos de gestão são implementados pelos *data stewards*, responsáveis pelo exercício da AD, que é o elo entre a política de dados e a implementação das ações pelos usuários (Alhassan, Sammon, & Daly, 2018; Plomp et al., 2019).

Portanto, compreender e medir o impacto da administração e da analítica de dados é essencial para gerenciar o desempenho dos negócios. A AD, como guardião da governança de dados, desempenha um papel essencial na implementação de BA (Harrison et al., 2019). No entanto, também há uma escassez de pesquisas empíricas sobre as implicações do uso de BDA nas organizações (Mikalef, Boura, Lekakos, & Krogstie, 2019), especialmente na governança, no gerenciamento e na administração de BD (Abraham et al., 2019; Nokkala et al., 2019) e na relação entre BA e GDC (Richards et al., 2019). Embora as descobertas de Richards et al. (2019) indiquem que a BA afeta a GDC de



forma positiva e significativa, não há estudos que considerem sua inter-relação com a administração do BD. Portanto, este estudo busca responder às seguintes questões de pesquisa:

- Quanto o BD, a AD e a BA impactam a GDC?
- Como o BD afeta a GDC, direta e indiretamente, por meio da administração e da analítica de dados de negócios?

Assim, o objetivo deste estudo é analisar o papel do BD, da administração e da analítica de dados de negócios como habilitadores da GDC.

Os achados sugerem que a AD e a BA, isoladamente ou em conjunto, podem transmitir o efeito BD à GDC, porém um melhor nível de ajuste do modelo é obtido quando há uma multimediação serializada nessa relação, sendo a AD um antecedente da BA. Tais achados são especialmente relevantes, pois o uso do BDA exige que os dados sejam coletados e analisados de forma centralizada, garantindo a aplicação de padrões, protocolos, métodos e ferramentas (Grover, Chiang, Liang, & Zhang, 2018). Isso reforça a importância da AD, cujo papel é garantir a aplicação efetiva de definições, políticas e procedimentos para o uso de dados em toda a organização (Koltay, 2016). De um lado, cresce a necessidade de democratização da análise de BD, pois o treinamento e uso de ferramentas de análise de dados (por exemplo, PowerBI, Tableau, QlikView, SAS e outras) criaram atividades de autosserviço BA (*self-service* BA) devido às novas habilidades de especialistas de negócios com competência em manuseio de dados, o que pode possibilitar uma maior flexibilidade para os gestores na gestão corporativa (Riggins & Klamm, 2017). No entanto, isso requer um entendimento comum desses dados em toda a organização (Abraham et al., 2019). E de outro lado, quando se lida com dados pessoais em negócios, é necessário garantir o cumprimento de requisitos regulatórios mais rigorosos (Alharthi et al., 2017; Khatri & Brown, 2010; Plotkin, 2020; Thompson, Ravindran, & Nicosia, 2015), tais como o *General Data Protection Regulation* (GDPR) na União Europeia, o *California Consumer Privacy Act* (CCPA) no estado norte-americano da Califórnia ou a Lei Geral de Proteção de Dados (LGPD) no Brasil.

A seguir, este estudo parte da fundamentação teórica para definição de construtos e desenvolvimento do modelo de pesquisa. Em seguida, apresentam-se o método, a análise e a discussão dos resultados, e, por fim, as considerações finais.

2. FUNDAMENTOS TEÓRICOS

2.1 Gerenciamento de desempenho corporativo habilitado por *big data analytics*

As estratégias de negócios são cada vez mais dependentes de dados, e o BDA está redefinindo inovação, competição e produtividade (Côrte-Real et al., 2019). O BDA suporta o conjunto de tecnologias, métodos e aplicativos avançados para armazenamento, gerenciamento e análise de dados, para melhorar a tomada de decisão (Chen et al., 2012). No entanto, BD e BA são elementos distintos. A Figura 2.1.1 apresenta essas definições.

(Figura 2.1.1)

BIG DATA E BUSINESS ANALYTICS

<p><i>Big data</i></p>	<ul style="list-style-type: none"> • O BD pode ser definido como a utilização de uma grande quantidade de dados, de diferentes fontes e tipos, criados rapidamente, o que implica maiores desafios para coletá-los, gerenciá-los e processá-los por meio de sistemas e recursos tradicionais de apoio à tomada de decisão (Ghasemaghaei & Calic, 2020). • As características essenciais do BD são volume, variedade e velocidade, bem como variabilidade (Chen, Mao, & Liu, 2014). • A expressão <i>big data</i> é usada para descrever o grande volume de dados digitais produzidos pela atividade humana que é muito difícil de gerenciar usando ferramentas convencionais de análise de dados, sendo caracterizado pelos três Vs: volume, variedade e velocidade (Alharthi et al., 2017).
<p><i>Business analytics</i></p>	<ul style="list-style-type: none"> • A BA inclui o uso de modelos, fórmulas e algoritmos para configurar o conjunto de regras ou instruções elaboradas para resolver problemas de negócios e muitas vezes é subdividida em quatro dimensões: descritiva, diagnóstica, preditiva e prescritiva (Appelbaum, Kogan, Vasarhelyi, & Yan, 2017; Duan, Cao, & Edwards, 2020; Fleckenstein & Fellows, 2018). • Contribui para a análise do BD, melhorando a compreensão dos padrões de desempenho e a preparação de análises de tendências, de forma a possibilitar a projeção de previsões e potenciais riscos e resultados futuros, além da identificação e adoção das melhores estratégias, de forma a otimizar os objetivos, maximizando oportunidades e potencialidades ou minimizando riscos e fraquezas (Acito & Khatri, 2014; Appelbaum et al., 2017; Duan et al., 2020).

Fonte: Elaborada pelos autores.

Assim, o BD está mais relacionado à tecnologia, enquanto a BA suporta a tomada de decisões em vários níveis dentro da organização. Portanto, as organizações precisam primeiro coletar e armazenar dados e, em seguida,

aplicar os processos de BA. Ao possibilitar que as decisões sejam tomadas com mais rapidez ou precisão, isso permite que eles criem uma vantagem competitiva (Dubey et al., 2019; Fernando & Engel, 2018). Com o gerenciamento estratégico de dados, a organização pode proteger e aproveitar seus dados (DalleMule & Davenport, 2017), tornando-o mais competitivo para atingir metas e obter valor (Grover et al., 2018). As descobertas de Ferraris et al. (2019) demonstram que as empresas que desenvolveram mais capacidades de BDA do que outras, tanto tecnológicas quanto gerenciais, aumentaram seu desempenho.

A ascensão da BA está, de certa forma, relacionada à noção de usar medidas de desempenho como uma forma de testar as hipóteses de negócios (Acito & Khatri, 2014). Suportada por *scorecards* e painéis, a GDC ajuda a analisar e visualizar uma variedade de métricas de desempenho (Chen et al., 2012). A GDC contempla as atividades, os processos, as metodologias, as métricas e as tecnologias utilizadas pelas organizações para estabelecer, implementar, medir, monitorar e gerenciar o desempenho das estratégias de negócios (Weeserik & Spruit, 2018). A GDC é sinônimo de gestão de desempenho empresarial ou de negócios (Weeserik & Spruit, 2018) e busca entender como o desempenho dos negócios pode ser medido e quais são os determinantes do desempenho dos negócios, combinando práticas de gestão e TI, para habilitar o desempenho organizacional (Richards et al., 2019). No entanto, a GDC depende da integração de dados externos e internos de uma ampla variedade de fontes (BD) para, por exemplo, traçar um curso para a organização, definir indicadores de desempenho, coletar e analisar dados de desempenho e tomar ações corretivas (Richards et al., 2019).

A GDC pode ser observada por meio do desenvolvimento e da aprendizagem de competências organizacionais, da otimização da eficiência e eficácia dos processos internos, e da melhoria na gestão do relacionamento com clientes, fornecedores e partes interessadas, resultando no aumento do valor acrescentado econômico e financeiro para o negócio, que é essencial para agilidade, inovação e desempenho competitivo nos ambientes de negócios contemporâneos (Côrte-Real et al., 2019; Dubey et al., 2019; Mikalef & Pateli, 2017; Wamba & Akter, 2019; Mikalef, Krogstie, Pappas, & Pavlou, 2020). Portanto, a GDC habilitada pelo BDA facilita a formulação da estratégia e o controle e a avaliação do desempenho do negócio.

2.2 Administração de dados: o elo entre a governança e o gerenciamento de dados

As expressões governança de dados, gerenciamento de dados e administração de dados são geralmente usadas como sinônimos, mas, embora sejam

interdependentes, são conceitos distintos (Koltay, 2016). Inicialmente, é importante esclarecer a diferença entre eles. A Figura 2.2.1 apresenta os conceitos.

(Figura 2.2.1)

GOVERNANÇA, ADMINISTRAÇÃO E GERENCIAMENTO DE DADOS

Governança de dados	<ul style="list-style-type: none">• Refere-se à estruturação de políticas de gerenciamento de dados para garantir a qualidade, padronização e segurança dos dados, incluindo a análise de como as informações são distribuídas e administradas, e também a definição das estruturas organizacionais e dos agentes responsáveis pela gestão de todo o ciclo de vida dos dados (Khatri & Brown, 2010; Nielsen, 2017; Harrison et al., 2019; Abraham et al., 2019).• Especifica uma estrutura de gerenciamento de dados multifuncional como um ativo estratégico da organização e direitos e responsabilidades de tomada de decisão, formalizando políticas, padrões e procedimentos, e monitorando a conformidade de dados (Abraham et al., 2019).
Administração de dados	<ul style="list-style-type: none">• Denota a responsabilidade dos administradores para com os dados, no que se refere à forma como eles cuidam de seus dados, a fim de garantir que o trabalho seja realizado de acordo com as políticas e práticas estabelecidas pela governança e que os dados sejam precisos, acessíveis, utilizáveis e atuais (Koltay, 2016; Brous, Janssen, & Vilminko-Heikkinen, 2016).• É um conjunto de métodos de gestão de dados que abrange aquisição, armazenamento, integração e procedimentos de manutenção, distribuição e utilização, cujo objetivo é garantir exatidão, validade, qualidade, segurança, gestão e retenção de dados (Nokkala et al., 2019; Rosenbaum, 2010).
Gerenciamento de dados	<ul style="list-style-type: none">• O objetivo é garantir que as informações mais importantes para a organização sejam bem definidas. Diz respeito a como os dados são coletados, armazenados, transformados, distribuídos e consumidos em toda a organização ao longo de seu ciclo de vida, e isso inclui as regras que padronizam formatos estruturados, como bancos de dados e sistemas de arquivamento, sistemas de integração de dados e processos de gerenciamento que os consomem (Fleckenstein & Fellows, 2018).• Consiste nos processos necessários para gerenciar todo o seu ciclo de vida – aquisição, pré-processamento e tratamento, distribuição, exclusão e descarte (Côrte-Real et al., 2019).• Conjunto de processos que inclui governança de dados, armazenamento de dados e documentos e gerenciamento de arquitetura, qualidade e segurança de dados e desenvolvimento de dados por meio de <i>business intelligence/ analytics</i> (Surbakti, Wang, Indulska, & Sadiq, 2020).

Fonte: Elaborada pelos autores.

Com base na literatura revisada, governança de dados refere-se às decisões que são tomadas e a quem as toma, para garantir a gestão e o uso eficazes dos recursos; e o gerenciamento de dados envolve a implementação de decisões como parte da implementação diária das políticas de governança de dados (Khatri & Brown, 2010; Alhassan et al., 2018). Embora sejam conceitos distintos, ambos são interdependentes. E a AD é definida como a função ou o conjunto de atividades em que o trabalho relacionado aos dados é realizado de acordo com as políticas e práticas estabelecidas pela governança, com o objetivo de garantir que as informações da organização sejam confiáveis, saudáveis, de boa qualidade e preservadas (Alhassan et al., 2018; Brous et al., 2016; Khatri & Brown, 2010; Nokkala et al., 2019; Plomp et al., 2019; Rosenbaum, 2010). Considera-se que a AD é a responsável por operacionalizar a política de dados e a implementação das ações dos utilizadores, sendo, portanto, o elo entre a governança e a gestão dos dados.

Conforme apresentado na Figura 2.2.1, para definir o construto AD, foram revisados outros conceitos já consolidados na literatura: “governança de dados” e “gerenciamento de dados”. Isso ajudou a diferenciar os elementos e a delimitar as características específicas desse novo construto. Assim, “governança de dados” e “gerenciamento de dados” não são construtos abordados neste estudo, mas servem como pano de fundo conceitual para: 1. auxiliar na definição do contexto e do problema de estudo; 2. possibilitar a diferenciação conceitual para definir o papel do construto AD; e 3. ajudar, como contexto, na discussão das implicações do estudo. Na verdade, a AD está interligada com a governança e o gerenciamento de dados (Nokkala et al., 2019), porque contempla precisamente a finalidade prática de ambos, uma vez que se refere à efetiva operacionalização e aplicação das ações de gerenciamento de dados de acordo com os princípios estabelecidos pela governança de dados.

Para Plotkin (2020), cabe à AD melhorar a qualidade dos dados a partir das seguintes iniciativas: padronizar os dados em nível corporativo; mediar a resolução de problemas relacionados a dados (por exemplo, divergência nas regras e requisitos de qualidade de dados); garantir o alinhamento e a comunicação dos objetivos de governança de dados em toda a organização, comunicando as regras aos usuários dos dados; colaborar com as partes interessadas no gerenciamento de definições, políticas, procedimentos e questões relacionadas a dados; e promover seu uso como ativo para obtenção de vantagem competitiva. Nesse contexto, os *data stewards* são administradores e operadores de negócios responsáveis por traduzir a política de governança de dados na implementação de práticas de acordo com as recomendações e



os requisitos por ela estabelecidos (Plomp et al., 2019) e geralmente são grupos organizados (comitês, conselhos, grupos de usuários) que representam as partes interessadas do negócio e são responsáveis por tomar decisões sobre o tratamento dos ativos que contribuem para a comunicação entre os usuários e o armazenamento, gerenciamento e compartilhamento de dados. Eles atuam como o canal entre a TI e os negócios, cujo desafio é garantir que um dos ativos mais críticos da corporação – seus dados – seja usado em sua capacidade máxima.

2.3 Desenvolvimento do modelo de pesquisa

A literatura indica que um volume e uma variedade de dados gerados a partir de uma alta velocidade (BD) é um recurso relevante para uma tomada de decisão ágil e eficaz (Conboy, Mikalef, Dennehy, & Krogstie, 2020; Ghasemaghahi & Calic, 2020). As soluções BD suportam a integração de dados, possibilitando um aumento considerável na quantidade e qualidade das informações, o que torna maior o impacto no desempenho (Wamba et al., 2017). Para Urbinati, Bogers, Chiesa e Frattini (2019), o papel do BD é auxiliar na tomada de decisão, ou seja, apoiar a GDC. Assim, a primeira hipótese sugere que lidar com BD influencia a GDC.

- H1: O BD impacta a GDC.

A integração de várias fontes externas e internas de BD pode se tornar incontrollável (Appelbaum et al., 2017; Coyne et al., 2018), implicando maiores desafios para controlar a veracidade, vulnerabilidade, volatilidade e validade dos dados (Fleckenstein & Fellows, 2018). Para gerenciar o BD, além dos repositórios de dados tradicionais, são necessários *data lakes* – repositórios de dados estruturados e não estruturados, distribuídos dentro da organização e *on-line* na *web*. Assim, lidar com o BD implica mais desafios para a AD, conforme proposto na segunda hipótese.

- H2: O BD impacta a AD.

O valor BD é desbloqueado apenas quando usado para conduzir a tomada de decisão por meio de processos analíticos (Mikalef et al., 2019). As organizações precisam primeiro coletar e armazenar dados e, em seguida, aplicar processos BA (Fernando & Engel, 2018). Portanto, a terceira hipótese sustenta que o BD afeta a BA.

- H3: O BD impacta a BA.



Recursos de governança e gerenciamento de dados são essenciais para a análise de dados (Harrison et al., 2019; Lillie & Eybers, 2018). Considerando que o papel da AD é a própria aplicação de políticas, práticas de governança e gerenciamento de dados (Alhassan et al., 2018; Plomp et al., 2019; Rosenbaum, 2010; Koltay, 2016; Thompson et al., 2015), argumenta-se que a AD é essencial para BA e GDC.

- H4: A AD impacta a BA.
- H5: A AD impacta a GCD.

A BA contribui para a compreensão dos padrões de desempenho, para a avaliação do meio ambiente, para a análise de tendências e projeção de resultados futuros, e para a formulação das melhores estratégias (Duan et al., 2020). Isso permite que a organização obtenha um melhor entendimento de seus clientes e direcione a otimização de processos para melhorar sua eficiência e eficácia (Grover et al., 2018; Mikalef et al., 2019; Wamba et al., 2017), aumentando a produtividade e o desempenho organizacional, o que está apoiado na sexta hipótese.

- H6: A BA impacta a GCD.

A integração e gestão de dados facilitam seu compartilhamento e o estabelecimento de uma visão de dados corporativos (Fleckenstein & Fellows, 2018; Grover et al., 2018; Seddon et al., 2017). Isso habilita a desenvolver análises de forma compartilhada por toda a organização, estimulando novos e interessantes usos, além da resolução de problemas. Integrar BD em toda a organização é importante para fornecer informações de boa qualidade para fins analíticos (Harrison et al., 2019; Lillie & Eybers, 2018). Vale ressaltar que, ao se tratar de BD, a qualidade é considerada fundamental para a tomada de decisão (Wamba & Akter, 2019). No entanto, isso requer estruturas de gestão e AD para garantir precisão, validade, qualidade, segurança, gestão e conservação de dados (Rosenbaum, 2010). Portanto, a AD pode influenciar a qualidade e a eficiência dos resultados analíticos e das informações gerenciais.

- H7: A AD medeia a relação de BD com BA (H7a) e GDC (H7b).

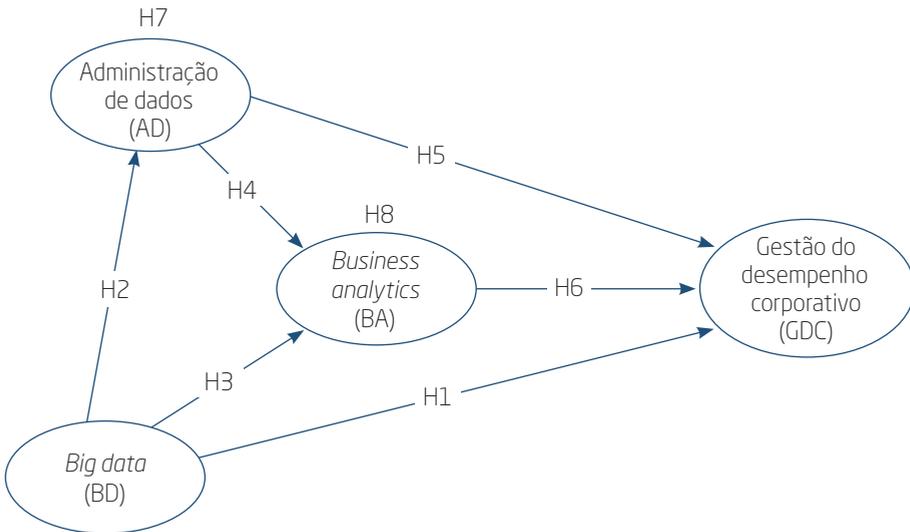
Para superar a complexidade e a sobrecarga causadas pelo BD, é fundamental gerar *insights* (Seddon et al., 2017). Para possibilitar que os gerentes encontrem apenas os dados capazes de fornecer percepções valiosas, as organizações devem obter uma rica combinação de informações por meio

deles (Kitchens et al., 2018). Portanto, há a hipótese de que a BA pode mediar a relação entre BD e AD com GDC.

- H8: A BA medeia a relação entre BD e GDC (H8a), e entre AD e GDC (H8b).

Acito e Khatri (2014) indicam que extrair valor dos dados requer alinhamento entre recursos e capacidades analíticas com a GDC, estruturando uma ordenação da relação entre esses elementos. O quadro estrutural para BA proposto por Acito e Khatri (2014) foi considerado pano de fundo para a organização das relações discutidas neste estudo. Entende-se que a AD está no nível informacional operacional, construindo a ponte entre os dados (BD) e a BA. Com base nas hipóteses desenvolvidas, foi elaborado o modelo de pesquisa (Figura 2.3.1).

(Figura 2.3.1)
MODELO DE PESQUISA



Fonte: Elaborada pelos autores.

Esse modelo preenche uma lacuna de pesquisa, pois observa o construto AD, que, apesar de desempenhar um papel importante na operacionalização da governança e gestão de dados, não é encontrado na literatura de sistemas de informação, e analisa sua relação com o construto que trata da análise de dados de negócios (BA) e os processos para explicar como o BD afeta a GDC.

3. MÉTODO

Realizou-se um estudo quantitativo para avaliação empírica do modelo de pesquisa. Para a coleta de dados, foi desenvolvida uma *survey* com base em um questionário estruturado, uma vez que o foco da pesquisa está em mensurar a percepção dos participantes sobre o uso do BD e o papel da AD e da BA na GDC. A seguir, descrevem-se o desenvolvimento do instrumento, os procedimentos de coleta de dados, os dados demográficos da amostra e as análises estatísticas realizadas.

3.1 Desenvolvimento do instrumento de pesquisa

Para o desenvolvimento do instrumento de pesquisa, após conceituarmos os construtos, utilizamos a literatura existente para desenvolver itens que representem a definição dos construtos. Além disso, a validação de face e conteúdo do instrumento foi realizada com o apoio de cinco especialistas da área (dois profissionais, dois mestres em Administração e um doutor em Sistemas de Informação). Durante o processo de validação de face do instrumento, foram feitos ajustes na estrutura do questionário, como a retirada de itens que continham algum nível de ambiguidade em sua definição, a integração de itens com definições semelhantes e a adequação na descrição de alguns itens.

O instrumento de pesquisa foi desenvolvido pelos autores, por meio da operacionalização dos construtos em variáveis de acordo com as referências presentes na literatura, conforme indicado no Apêndice A. Para mensurar as variáveis, adotou-se uma escala do tipo Likert de cinco pontos de concordância, de 1 (não concordo) a 5 (concordo plenamente). Além desses itens, solicitaram-se dados demográficos e perguntou-se a frequência de uso do BDA – em uma escala de cinco pontos, de “raramente” a “muito frequente”. Para a qualificação da amostra, estabeleceram-se duas questões de triagem dos participantes, com respostas (“sim” ou “não”). A primeira questionou se o entrevistado era um usuário BDA na organização em que trabalhava. Para isso, apresentaram-se uma definição e exemplos de ferramentas de BDA. A segunda questão tinha como objetivo verificar se o respondente possuía experiência profissional.

3.2 Procedimento de coleta e tratamento da amostra

O tamanho mínimo da amostra foi estimado com o apoio do *software* G*Power V3.1.9.4 (Faul, Erdfelder, Buchner, & Lang, 2009). De acordo com

as recomendações de Cohen (1988) e Hair, Hult, Ringle e Sarstedt (2017), utilizando os valores do poder do teste como 0,80, mediana do tamanho do efeito (f^2) (igual a 0,15), e, considerando que a variável latente GDC possui três preditores, calculou-se que a amostra mínima é de 77 respondentes. A amostra mínima foi superada, pois a amostra é de 312 casos.

A população desejada eram gestores de organizações brasileiras, de diversos setores da economia, que trabalham com BD. Os dados foram coletados por meio de instrumento de pesquisa, por meio de formulário eletrônico, respondido por 366 pessoas. Para a qualificação da amostra, estabeleceram-se duas questões de triagem dos participantes. Assim, coletaram-se 312 observações qualificadas de gestores e profissionais experientes, que confirmaram a utilização do BDA na organização em que trabalhavam. Mais de 70% da amostra possui pós-graduação. Aproximadamente 75% dos participantes têm formação nas áreas de gestão empresarial, informática e estatística, o que é condizente com os utilizadores da ciência de dados. A maioria dos participantes (61%) trabalha em organizações do setor de serviços, destacando-se as áreas de TI (18%), gestão empresarial, *marketing* e inteligência de mercado (18%) e serviços financeiros/bancários (11%). Cada informante que respondeu à pesquisa corresponde a uma organização, portanto a amostra representa 173 organizações brasileiras. Há uma concentração de organizações das regiões Sul e Sudeste (86% do total), que são as regiões mais economicamente ativas do país. A Figura 3.2.1 mostra o perfil dos respondentes e das organizações.

(Figura 3.2.1)

PERFIL DOS RESPONDENTES

Total informantes/organizações (n = 312)					
Idade (anos)	(%)	Experiência profissional (anos)	(%)	Frequência de uso de BDA	(%)
≤ 25	8	≤ 2	3	1 - Raramente	1
26 ≤ x ≤ 35	44	2 ≤ x ≤ 5	11	2	6
36 ≤ x ≤ 45	32	6 ≤ x ≤ 10	31	3	19
46 ≤ x ≤ 55	12	11 ≤ x ≤ 15	25	4	26
x > 55	4	x > 15	30	5 - Muito frequentemente	48
Tamanho da firma	(%)	Setor econômico	(%)	Região geográfica	(%)
Pequeno	24	Serviços	61	Norte	1
Médio	31	Indústria	14	Noroeste	6

(continua)

(Figura 3.2.1 (conclusão))
PERFIL DOS RESPONDENTES

Total informantes/organizações (n = 312)					
Tamanho da firma	(%)	Setor econômico	(%)	Região geográfica	(%)
Grande	45	Comércio	6	Centro-Oeste	8
		Governo	12	Sudeste	39
		Outros	7	Sul	46

Fonte: Elaborada pelos autores.

Como se trata de dados primários, foi necessário garantir que nenhum viés sistemático estivesse influenciando as informações coletadas. Assim, realizou-se o teste de fator único de Harman (1976), cujo resultado foi de 39,99%, inferior ao mínimo de 50%. O teste AFC no *software* SPSS, com rotação varimax e autovalor igual a 1,0, indicou a existência dos cinco componentes previstos no modelo de pesquisa para uma variação total, explicada por 66,47%, confirmando todas as dimensões previstas no modelo. Além disso, realizou-se a análise do viés de “não respondentes”, segundo Armstrong e Overton (1977). Quando se realizaram esses testes, verificou-se que tanto o viés do método comum quanto o viés “não respondentes” não são problemas significativos.

4. ANÁLISE DE DADOS

Para a análise dos dados, adotaram-se as ferramentas IBM SPSS Statistics V20 e SmartPLS V3. Avaliou-se o modelo de medição, e as equações estruturais foram modeladas por mínimos quadrados parciais (*partial least squares structural equation modeling* – PLS-SEM). O método PLS-SEM foi adotado por permitir trabalhar com modelos complexos, sendo preferido para desenvolvimento teórico e explicação da variância do construto (Hair et al., 2017; Hair, Risher, Sarstedt, & Ringle, 2019), recomendado para pesquisa de gestão e amplamente utilizado em estudos de sistemas de informação (Mikalef & Pateli, 2017).

4.1 Modelo de mensuração

O modelo lida com construtos reflexivos, com suporte do *software* SmartPLS, consistência interna, confiabilidade composta, validade conver-

gente e validade discriminante (Hair et al., 2019). Os resultados são apresentados na Figura 4.1.1.

(Figura 4.1.1)
AVALIAÇÃO DO MODELO DE MENSURAÇÃO

Variável ¹	Indicadores ²			Critério HTMT			
	α	CC	AVE	BA	BD	GDC	AD
BA	,875	,914	,728				
BD	,811	,876	,638	,518			
GDC	,910	,930	,690	,578	,332		
AD	,863	,897	,593	,639	,540	,561	

¹ Big data (BD); *business analytics* (BA); administração de dados (AD); gestão do desempenho corporativo (GDC).

² Alfa de Cronbach (α); confiabilidade composta (CC); *average variance extracted* (AVE).

Fonte: Elaborada pelos autores.

Para todos os construtos, os indicadores alfa de Cronbach e confiabilidade composta (CC) são maiores que 0,7 (Hair et al., 2017), sugerindo, respectivamente, a consistência interna e a confiabilidade dos construtos. A validade convergente, calculada a partir da variância média extraída (*average variance extracted* – AVE) de cada fator, indica o quanto uma determinada composição de variáveis observáveis representa uma única variável latente. O indicador AVE de cada um deles foi superior ao recomendado – de 0,5 (Hair et al., 2017). Além disso, as cargas externas de cada item em seu respectivo construto (ver Apêndice A) são maiores que 0,7 (Hair et al., 2017). Portanto, conclui-se que os construtos possuem validade convergente.

A validade discriminante indica o quanto um construto difere dos outros. O critério mais moderno é a *Henseler's Heterotrait-Monotrait Ratio* (HTMT), de Henseler, Ringle e Sarstedt (2015). De acordo com Mikalef et al. (2020), a HTMT é melhor que o critério de Fornell-Larcker. Por meio da HTMT, a correlação média de indicadores entre construtos que medem diferentes aspectos do modelo é comparada em relação à média de correlações de indicadores dentro de um mesmo construto, cujos resultados devem atingir o limite de 0,85 (Henseler et al., 2015; Hair et al., 2019). Portanto, observa-se que há validade convergente.

4.2 Modelo estrutural

A avaliação do modelo estrutural é realizada por meio dos coeficientes de caminho, do nível de significância das relações, do tamanho do efeito (f^2), dos coeficientes de correlação de Pearson (R^2) e validade preditiva (Q^2), e da *standardized root mean square residual* (SRMR), conforme recomendado por Hair et al. (2019). Preliminarmente, a colinearidade entre os construtos foi analisada por meio do *variance inflation factor* (VIF), observando que ela não é um problema, pois todos os valores são inferiores a 3 (Hair et al., 2019). Em seguida, o procedimento de *bootstrapping* (com cinco mil amostras) foi utilizado para testar se as hipóteses são significantes, obtendo-se $p < 0,05$ (Hair et al., 2017). Exceto para H1, todas as outras hipóteses foram sustentadas com significância menor que 0,1%.

Em seguida, passamos a avaliar a porção da variância das variáveis endógenas, que é explicada pelo modelo estrutural pelo coeficiente de determinação de correlação de Pearson (R^2). Os construtos BA e GDC têm efeitos grandes, pois possuem $R^2 > 26\%$, e o construto AD tem efeito médio, pois tem $R^2 > 13\%$ (Cohen, 1988). As figuras 4.2.1 e 4.2.2 contêm os resultados do modelo estrutural.

(Figura 4.2.1)

SIGNIFICÂNCIA DOS COEFICIENTES DO CAMINHO DO MODELO ESTRUTURAL

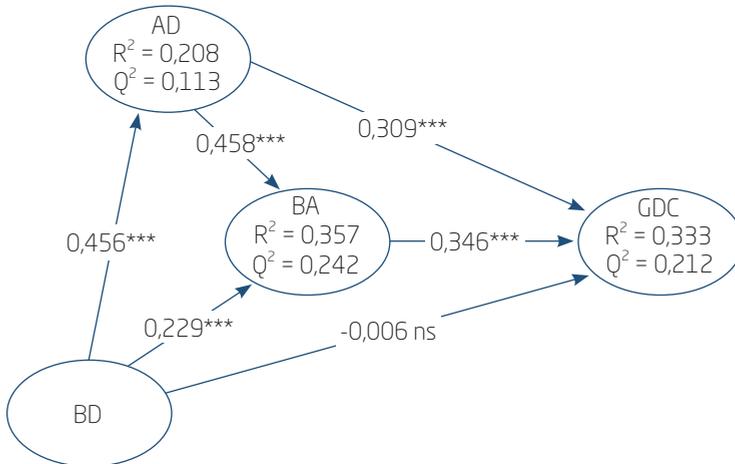
Hipótese	Caminho	Coefficiente	Estatística T	p-valores	Tamanho do efeito (f^2)	Análise do f^2 de Cohen	Evidência empírica
H1	BD -> GDC	-,006	0,009	,927 ns	,000	Não relevante	Não suportada
H2	BD -> AD	,456	9,707	,000***	,263	Médio	Suportada
H3	BD -> BA	,229	4,097	,000***	,064	Pequeno	Suportada
H4	AD -> BA	,458	9,062	,000***	,258	Médio	Suportada
H5	AD -> GDC	,309	5,284	,000***	,090	Pequeno	Suportada
H6	BA -> GDC	,346	5,131	,000***	,115	Pequeno	Suportada

*** $p < 0,001$, ns - não significante.

Fonte: Elaborada pelos autores.

(Figura 4.2.2)

MODELO DE PESQUISA COM RESULTADOS DE TESTE DE HIPÓTESES



*** $p < 0,001$, ns - não significante.

Fonte: Elaborada pelos autores.

Para verificar a porção que cada variável exógena representa na explicação das variáveis endógenas do modelo, avaliou-se o tamanho do efeito. Observa-se que, exceto para a relação entre BD -> GDC, todos os outros efeitos (f^2) do coeficiente de Cohen foram acima de 0,02, o que evidencia resultados adequados para fatores latentes (Henseler, Ringle, & Sinkovics, 2009). De acordo com Cohen (1988) e Hair et al. (2017), $f^2 > 0,02$ representa um efeito de pequeno porte, já $f^2 > 0,15$, um efeito de médio porte, e $f^2 > 0,35$, um efeito de grande porte. Conforme a Figura 4.2.1, os efeitos de tamanho médio são evidenciados na relação entre BD ->AD e AD -> BA; e os efeitos de tamanho pequeno, na relação entre BD -> BA, AD -> GDC e BA -> GDC. A relevância preditiva, por sua vez, é medida pelo indicador Stone-Geisser (Q^2). Observa-se que todas as variáveis endógenas apresentam $Q^2 > 0$, o que sugere relevância preditiva significativa (Hair et al., 2019).

O único critério de ajuste de modelo recomendado para modelagem de equações estruturais por mínimos quadrados parciais (PLS) é a SRMR (Hu & Bentler, 1999). Ressalta-se que o índice SRMR (0,059) atende aos parâmetros recomendados na literatura para sua validação, valor inferior a 0,08 (Hair et al., 2019; Hu & Bentler, 1999).

4.3 Análise de mediações

Para a mediação, é necessário que a variável independente afete com significância a variável dependente ao remover a variável mediadora (Zhao, Lynch, & Chen, 2010). Essa suposição é atendida porque todos os relacionamentos têm efeitos diretos significantes quando analisados sem a intervenção das variáveis mediadoras. A Figura 4.3.1 apresenta os resultados da análise de mediação.

(Figura 4.3.1)
TESTE DE MEDIAÇÃO

Modelo	Hipótese	SRMR	Efeito direto	Efeito indireto	Efeito total	Tipo de mediação
Mediações simples	H7a: BD -> AD -> GDC	,066	,073 ns	,215***	,288***	Completa
	H7b: BD -> AD -> BA	,072	,229***	,209***	,438***	Parcial
	H8a: BD -> BA -> GDC	,054	,074 ns	,213***	,287***	Completa
	H8b: AD -> BA -> GDC	,062	,307***	,194***	,501***	Parcial
Múltipla tipo A	H7a, H8a: BD -> AD+BA -> GDC	,106	-,007 ns	,294***	,287***	Completa
Múltipla tipo B	H7b, H8b: BD -> AD -> BA -> GDC	,059	-,006 ns	,293***	,287***	Completa

*** $p < 0,001$, ns - não significante.

Fonte: Elaborada pelos autores.

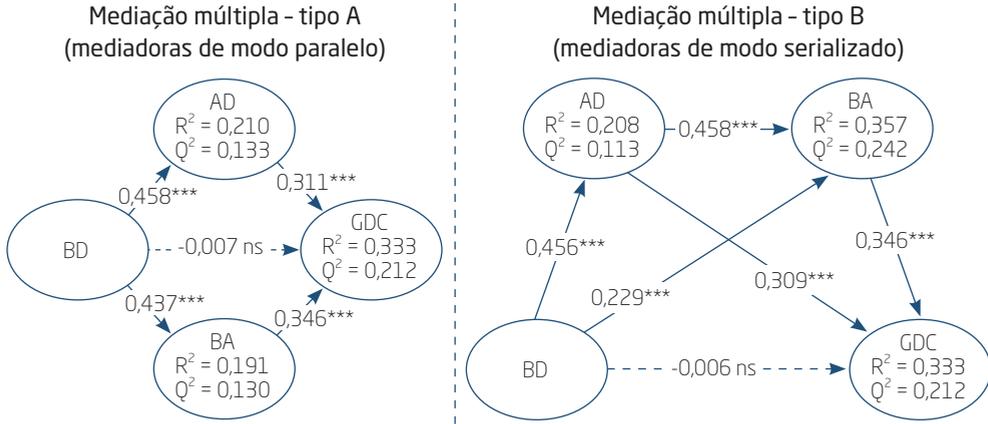
Quanto à análise de mediações simples, os resultados mostram que tanto AD quanto BA medeiam completamente a relação entre BD e GDC (o que suporta H7 e H8). Mediações completas denotam que os construtos AD e BA podem transmitir totalmente os efeitos de BD para GDC. Além disso, a AD mede parcialmente a relação entre BD e BA, e a BA mede parcialmente a relação entre AD e GDC. Isso significa que a AD pode transmitir uma parte do efeito BD para a BA e que a BA transmite parte do efeito da AD para a GDC.

Além de mediações simples, segundo Zhang, Zyphur e Preacher (2009), dois tipos diferentes de mediação múltipla das variáveis mediadoras foram analisados: “tipo A”, uma mediação múltipla que considera mediadores com causalidade em paralelo, e “tipo B”, uma mediação múltipla que considera a relação de causalidade serializada entre as variáveis mediadoras, partindo da

premissa de que a AD é antecedente da BA, conforme proposto como hipótese no modelo de pesquisa. O objetivo desse procedimento é avaliar qual modelo de mediação múltipla tem maior poder explicativo. A Figura 4.3.2 mostra os resultados dos vários modelos de mediação analisados.

(Figura 4.3.2)

MODELOS DE MEDIAÇÃO MÚLTIPLA



*** p < 0,001, ns - não significante.

Fonte: Elaborada pelos autores.

No que se refere à análise das mediações múltiplas, observa-se que o efeito do BD na GDC é totalmente transmitido por meio de variáveis explicativas, quando os mediadores ocorrem com causalidade em paralelo – modelo mediação múltipla do “tipo A” – e quando ocorrem em forma seriada – modelo de mediação múltipla do “tipo B”. Porém, quando ocorrem de forma serializada, em que a mediadora AD precede a mediadora BA, embora não sejam observadas diferenças no efeito total, diferenças importantes são observadas no ajuste do modelo. Em primeiro lugar, o modelo de mediação múltipla do “tipo B” apresenta coeficientes de determinação (R²) e relevância preditiva (Q²) da variável endógena BA maiores. Em segundo lugar, foi observada uma diferença considerável no indicador de ajuste do modelo, uma vez que a SRMR do modelo estimado “tipo B” é menor (0,059) do que o “tipo A” (0,106). Portanto, verifica-se que, de fato, o efeito de multimediação dessas variáveis se comporta de modo serializado (tipo B), pois, para um bom nível de ajuste, o indicador SRMR deve ser menor que 0,08 (Hair et al., 2017; Hu & Bentler, 1999).

5. DISCUSSÃO

Como o mundo e os dados dentro dele estão em um ciclo contínuo de mudança, as organizações que podem reconhecer a mudança e reagir de forma rápida e inteligente terão uma vantagem competitiva maior; a incorporação da estratégia BDA tem implicações decisivas para impulsionar e direcionar estratégias de negócios (Fleckenstein & Fellows, 2018). A capacidade de gerenciar, analisar e agir constitui um sistema de decisão “orientado por dados”, sendo caracterizado como um recurso significativo para a competitividade, o desempenho e a inovação (Tabesh et al., 2019). O uso estratégico do BDA implica a transformação dos processos de gestão e da cultura de tomada de decisão das organizações (Frisk & Bannister, 2017).

Estudos precedentes indicaram que quanto maiores forem o volume, a velocidade, a variedade e a variabilidade dos dados – BD –, maiores serão os riscos e as oportunidades para a GDC (Appelbaum et al., 2017; Conboy et al., 2020; Coyne et al., 2018; Ghasemaghaei & Calic, 2020; Riggins & Klamm, 2017; Urbinati et al., 2019). O BDA garante que os dados possam ser transformados em informações e conhecimentos de negócio úteis para processos de tomada de decisão eficientes, melhorando assim o desempenho (Ferraris et al., 2019). Este estudo confirma essa ideia ao identificar que a relação entre BD e GDC tem um efeito significativo quando a relação é analisada de modo isolado, isto é, sem a intervenção de outras variáveis latentes. Além disso, os achados deste estudo ampliam o conhecimento sobre esse fenômeno ao descrever como outros construtos podem explicar a transmissão do efeito de BD para a GDC.

Na presença das variáveis mediadoras AD e BA, a hipótese de que o BD influencia a GDC (H1) não foi sustentada, resultando na descoberta da mediação completa dessas variáveis intervenientes. Além disso, verificou-se que o BD influencia tanto a AD quanto a BA (H2 e H3, suportadas), e isso indica que, quando se lida com BD, o papel da AD e da BA é ainda mais relevante, pois um maior volume e uma grande variedade e variabilidade de dados, gerados com maior velocidade, requerem tratamentos de governança, gestão e analíticos mais sofisticados (Abraham et al., 2019; Fleckenstein & Fellows, 2018). Isso significa que, quando se lida com dados de diferentes domínios de negócios que são gerados, manuseados, analisados e interpretados por diferentes agentes organizacionais, torna-se imprescindível que os dados organizacionais sejam administrados e analisados de modo eficaz (Alhassan et al., 2018; Coyne et al., 2018; Kitchens et al., 2018).



Descobriu-se que a AD exerce um papel central no modelo, uma vez que influencia com significância a BA (com suporte H4), conforme documentado por Harrison et al. (2019), e a GDC (H5 suportada), confirmando as relações hipotéticas deste estudo. Além disso, a BA influencia diretamente a GDC (com suporte H6), o que é consistente com os achados de Richards et al. (2019). Assim, descobriu-se um efeito de mediação completa de AD e BA na relação entre BD e GDC (H7 e H8 suportadas), o que sugere que ambas as mediadoras, tanto isoladamente quanto de modo combinado, são capazes de transmitir o efeito do BD para a GDC. Constatou-se ainda uma mediação parcial de BA na relação entre AD e GDC (H9 e H10). Contudo, na presença de outras variáveis intervenientes, apesar de o caminho AD -> GDC ser significativo ($p < 0,000$), apresenta um efeito de tamanho pequeno na análise do f^2 . Diante desses achados, investigou-se outro tipo de mediação múltipla de AD e BA no modelo, considerando-se a AD como antecedente de BA. Essa nova configuração do modelo de multimediação provou-se mais eficiente na transmissão do efeito de BD para GDC. Isso confirma que a analítica de negócios exerce um papel crucial na GDC (Richards et al., 2019; Kitchens et al., 2018), pois a análise de dados pode contribuir para a geração de *insights* para a identificação de padrões, projeção de tendências, estudo de cenários e otimização de processos (Duan et al., 2020; Seddon et al., 2017).

Os achados evidenciam que a AD é uma abordagem eficaz para a execução e supervisão de planos, políticas, programas e práticas que controlam, protegem, entregam e aprimoram o valor dos ativos de dados (Koltay, 2016; Thompson et al., 2015). A AD é capaz de garantir a operacionalização e aplicação das políticas e mecanismos de governança de dados importantes tanto para protegê-los de irregularidades, inconsistências, fraudes ou falhas de segurança, o que garante sua veracidade e previne vulnerabilidades, quanto para alavancá-los ao propiciar a integração e o controle de todo o fluxo na organização, servindo para lidar com a variabilidade e a volatilidade dos dados, a fim de garantir o fornecimento de informação de boa qualidade para uso de BA, o que garante um dos ativos mais críticos da corporação – seus dados –, os quais podem ser usados em sua capacidade máxima (DalleMule & Davenport, 2017; Fleckenstein & Fellows, 2018; Lillie & Eybers, 2018; Rosenbaum, 2010).

6. CONSIDERAÇÕES FINAIS

Este estudo analisou a multimediação da AD e da BA na relação entre BD e GDC. Esse modelo preenche uma lacuna de pesquisa, pois observa o

construto AD que, apesar de exercer um papel importante na operacionalização da governança e gestão de dados, não é encontrado na literatura de sistemas de informação e analisa sua relação com o construto que trata da análise de dados de negócios (BA) para explicar como o BD afeta a GDC. As contribuições, implicações, limitações e propostas para estudos futuros são descritas a seguir.

6.1 Contribuições teóricas e implicações gerenciais

Diferentemente de outros estudos sobre o assunto, o desenho da pesquisa é baseado em um método quantitativo. Foi realizada pesquisa com 312 gestores usuários do BDA em organizações brasileiras de diferentes segmentos econômicos e regiões geográficas. Os dados foram analisados por meio de modelagem de equações estruturais e testes de mediação. Isso tornou possível fornecer evidências empíricas valiosas para a teoria e a prática da administração.

Primeiramente, contribui para a teoria ao conceituar, validar e discutir o papel do construto AD na gestão corporativa. Outra contribuição relevante está relacionada à mensuração de quanto BD, AD e BA impactam diretamente a GDC. Descobriu-se ainda que a AD exerce um papel fundamental, pois afeta a GDC, direta e indiretamente, por meio da BA, e medeia a relação entre BD e GDC. Com base nesses achados, avançou-se na análise da mediação múltipla para investigar se ela poderia ser mediada em paralelo (tipo A), tanto a partir do mesmo antecedente (BD) quanto de modo serializado (tipo B), tendo a AD como um precedente para BA. Assim, constatou-se que a multimediação do “tipo B” apresenta um melhor nível de ajuste do modelo, evidenciando como o efeito BD é transmitido à GDC por meio da multimediação da AD e BA.

As evidências deste estudo indicam que investir no desenvolvimento de capacidades de AD e BA é uma forma assertiva de controlar melhor os efeitos positivos e negativos que o BD pode gerar na GDC (Dubey et al., 2019; Harrison et al., 2019; Richards et al., 2019; Nokkala et al., 2019; Plomp et al., 2019; Rosenbaum, 2010). Ademais, é importante ressaltar que a AD é benéfica mesmo para organizações que não lidam com o BD, pois ela é capaz de influenciar com significância a GDC, direta e indiretamente, por meio da BA. Logo, as implicações deste estudo são consideravelmente abrangentes, pois, embora algumas organizações não lidem com BD e outras ainda estejam construindo sua capacidade de BA, de fato toda e qualquer organização possui AD de alguma forma, porque, mesmo na ausência de um programa de governança ou de processos de gestão padronizados, sempre haverá



alguém na organização a quem as pessoas recorrem com perguntas sobre o significado dos dados.

Usar BD e BA pode ajudar as organizações a mudar sua cultura de tomada de decisão, resultando em decisões melhores e mais eficazes (Frisk & Bannister, 2017). Muitas organizações têm adotado o *self-service* BA para facilitar a expansão do uso de dados na organização (Riggins & Klamm, 2017). Contudo, o excesso de flexibilidade e acessibilidade aos dados pode implicar uma desorganização no uso de dados e ferramentas analíticas, o que pode comprometer a qualidade dos dados e gerar interpretações e decisões equivocadas (Appelbaum et al., 2017).

A AD, como um instrumento de governança de dados (Koltay, 2016; Rosenbaum, 2010), precisa avaliar e garantir o equilíbrio entre o nível de acessibilidade aos dados de modo distribuído nas áreas de negócio e a disponibilidade de competências em ciência de dados, bem como zelar pelo cumprimento das políticas, dos processos e dos padrões de uso de dados, a fim de evitar o caos no uso deles, garantir sua proteção e alavancar a obtenção de *insights* e de valor. A integridade dos dados, com a proteção e aproveitamento deles, é a base da excelência futura. A AD impõe a preparação básica para enfrentar uma ameaça contra a integridade dos dados e garantir o máximo de aproveitamento dos investimentos neles (Keywell, 2020).

Naturalmente, as preocupações com a privacidade muitas vezes impedem a adoção e o uso de BDA nas organizações. Os novos requisitos regulatórios do GPDR, do CCPA ou da LGPD, apesar de serem barreiras, podem ser utilizados como uma grande oportunidade para as organizações. Assim, as organizações precisam acomodar essa legislação, bem como incorporar as melhores práticas gerais para lidar com dados confidenciais de clientes em suas políticas e operações (Alharthi et al., 2017). Por exemplo, as organizações podem se beneficiar de boas práticas de AD e BA para desenvolver uma estrutura ética e legal para o compartilhamento de dados (Zarkadakis, 2020), a fim de lidar com o tratamento do grande volume e variedade de dados gerados em alta velocidade (BD) na *web* e em operações de negócios, disponível em *data lakes* organizacionais, em dados abertos ou outras fontes, de modo a estabelecer uma gestão estratégica para proteger e alavancar dados, tanto aqueles da organização quanto os de clientes e partes interessadas (Dalle-Mule & Davenport, 2017; Thompson et al., 2015).

6.2 Limitações e estudos futuros

No estudo, especialmente duas limitações devem ser reconhecidas. Em primeiro lugar, a confirmação da mediação parcial de BA na relação entre AD



e GDC sugere que pode haver outras variáveis intervenientes não consideradas nesse modelo. Em segundo lugar, vale ressaltar a limitação geográfica, visto que o estudo foi aplicado apenas no contexto brasileiro. Embora as limitações mencionadas não comprometam os resultados obtidos, a generalização dos resultados deve ser avaliada de acordo com as características específicas do contexto e da amostra analisada.

Apesar da relevância da AD para o sucesso da governança de dados, há uma escassez de estudos empíricos sobre esta (Abraham et al., 2019; Nokkala et al., 2019). Portanto, estudos futuros podem aprofundar a investigação do fenômeno por meio de abordagens qualitativas e estudos de caso comparativos, para estabelecer uma compreensão das implicações dos modelos de gestão e das práticas de AD. Sugere-se também que estudos futuros busquem avaliar a influência de outros fatores organizacionais intervenientes que são promissores no contexto do BD e da ciência de dados, como estratégia de dados, cultura orientada a dados, literacia em dados, qualidade de dados, entre outras capacidades analíticas.

THE ROLE OF BIG DATA STEWARDSHIP AND ANALYTICS AS ENABLERS OF CORPORATE PERFORMANCE MANAGEMENT

ABSTRACT

Purpose: Digital transformation and big data (BD) have generated a real revolution in data-driven management. While BD improves corporate performance management (CPM), this also implies increasing exposure to risks at various BD's life cycle stages. As regulatory requirements and the need for database analysis in various business areas increase, the organization must establish definitions, policies, and processes to ensure data quality in order to protect and leverage its data to obtain a competitive advantage. Therefore, understanding data stewardship (DS) and business analytics (BA) is essential for business management. The purpose of this study is to analyze the role of BD, DS, and BA as enablers of CPM.

Originality/value: We contribute to the theory by conceptualizing, validating, and discussing the DS construct and by highlighting its role together with BA in the relationship between BD and CPM. The evidence in this study indicates that, in practice, DS and BA are critical

paths for organizations to obtain better control over the effects that BD can have on business performance management.

Design/methodology/approach: A survey was conducted with 312 managers who use big data analytics (BDA) in Brazilian organizations. The data were analyzed through structural equation and mediation tests.

Findings: The findings suggest that DS and BA, both alone and jointly, can transmit the BD effect to CPM. However, a better level of model adjustment is obtained when there is a serialized multi-mediation in this relationship, being DS an antecedent to BA.

KEYWORDS

Big data. Business analytics. Data governance. Data stewardship. Corporate performance management.

REFERÊNCIAS

- Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424–438.
- Acito, F., & Khatri, V. (2014). Business analytics: Why now and what next? *Business Horizons*, 57(5), 565–570.
- Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*, 60(3), 285–292.
- Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities: A comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management*, 31(2), 300–316.
- Appelbaum, D., Kogan, A., Vasarhelyi, M., & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting. *International Journal of Accounting Information Systems*, 25, 29–44.
- Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 14(3), 396–402.
- Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating decision-making in data management activities: A systematic review of data governance principles. In H. Scholl et al. (eds.), *Electronic Government. EGOV 2016. Lecture Notes in Computer Science* (pp. 115–125). Cham: Springer.



- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: L. Erlbaum Associates.
- Conboy, K., Mikalef, P., Dennehy, D., & Krogstie, J. (2020). Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. *European Journal of Operational Research*, 281(3), 656–672.
- Coyne, E. M., Coyne, J. G., & Walker, K. B. (2018). Big data information governance by accountants. *International Journal of Accounting & Information Management*, 26(1), 153–170.
- Côrte-Real, N., Ruivo, P., & Oliveira, T. (2019). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value? *Information & Management*, 57(1), 103–141.
- DalleMule, L., & Davenport, T. H. (2017). What's your data strategy. *Harvard Business Review*, 95(3), 112–121.
- Delen, D., & Zolbanin, H. M. (2018). The analytics paradigm in business research. *Journal of Business Research*, 90, 186–195.
- Duan, Y., Cao, G., & Edwards, J. S. (2020). Understanding the impact of business analytics on innovation. *European Journal of Operational Research*, 281(3), 673–686.
- Dubey, R., Gunasekaran, A., & Childe, S. J. (2019). Big data analytics capability in supply chain agility: The moderating effect of organizational flexibility. *Management Decision*, 57(8), 2092–2112.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Method*, 41(4), 1149–1160.
- Fernando, F., & Engel, T. (2018). Big data and business analytic concepts: A literature review. *Americas Conference on Information Systems*, New Orleans, USA, 24.
- Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). Big data analytics capabilities and knowledge management: Impact on firm performance. *Management Decision*, 57(8), 1923–1936.





- Fleckenstein, M., & Fellows, L. (2018). Implementing a data strategy. In M. Fleckenstein & L. Fellows, *Modern data strategy* (pp. 35–54). Cham: Springer.
- Frisk, J. E., & Bannister, F. (2017). Improving the use of analytics and big data by changing the decision-making culture: A design approach. *Management Decision*, 55(10), 2074–2088.
- Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of Business Research*, 108, 147–162.
- Grover, V., Chiang, R. H., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), 388–423.
- Hair, J. F., Hult, G., Ringle, C., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Los Angeles: Sage.
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24.
- Harman, H. H. (1976). *Modern factor analysis*. Chicago: University of Chicago Press.
- Harrison, T. M., Luna-Reyes, L. F., Pardo, T., De Paula, N., Najafabadi, M., & Palmer, J. (2019, June). The data firehose and AI in government: Why data management is a key to value and ethics. *Annual International Conference on Digital Government Research, Dubai, United Arab Emirates*, (pp. 171–176). ACM, New York, NY, USA, 20.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In R. R. Sinkovics & P. N. Ghauri (Eds.), *New challenges to international marketing* (pp. 277–319). Bingley, UK: Emerald Group Publishing.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152.
- Keywell, B. (2020, October 21). Your board needs a data-integrity committee. *Harvard Business Review*. Recuperado de <https://hbr.org/2020/10/your-board-needs-a-data-integrity-committee>



- Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced customer analytics: Strategic value through integration of relationship-oriented big data. *Journal of Management Information Systems*, 35(2), 540–574.
- Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal*, 42(4), 303–312.
- Lillie, T., & Eybers, S. (2018, August). Identifying the constructs and agile capabilities of data governance and data management: A review of the literature In K. Krauss, Turpin, M. Naude F. (eds.), *Locally Relevant ICT Research. IDIA 2018. Communications in Computer and Information Science* (pp. 313–326). Cham: Springer.
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, 98, 261–276.
- Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, 57(2), 103–169.
- Mikalef, P., & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. *Journal of Business Research*, 70, 1–16.
- Nielsen, O. B. (2017). A comprehensive review of data governance literature. *IRIS: Selected Papers of the Information Systems Research Seminar in Scandinavia*, (8), 120–133.
- Nokkala, T., Salmela, H., & Toivonen, J. (2019). Data governance in digital platforms. *Americas Conference on Information Systems*, Cancun, México, 25.
- Plomp, E., Dintzner, N., Teperek, M., & Dunning, A. (2019). Cultural obstacles to research data management and sharing at TU Delft. *Insights*, 32(1), 1–11.
- Plotkin, D. (2020). *Data stewardship: An actionable guide to effective data management and data governance*. London, UK: Elsevier.
- Rialti, R., Marzi, G., Ciappei, C., & Busso, D. (2019). Big data and dynamic capabilities: A bibliometric analysis and systematic literature review. *Management Decision*, 57(8), 2052–2068.
- Richards, G., Yeoh, W., Chong, A. Y. L., & Popovič, A. (2019). Business intelligence effectiveness and corporate performance management: An empirical analysis. *Journal of Computer Information Systems*, 59(2), 188–196.



- Riggins, F. J., & Klamm, B. K. (2017). Data governance case at KrauseMcMahon LLP in an era of self-service BI and big data. *Journal of Accounting Education*, 38, 23–36.
- Rosenbaum, S. (2010). Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research*, 45(5), 1442–1455.
- Seddon, P. B., Constantinidis, D., Tamm, T., & Dod, H. (2017). How does business analytics contribute to business value? *Information Systems Journal*, 27(3), 237–269.
- Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*, 57(1), 103–146.
- Tabesh, P., Mousavidin, E. & Hasani, S. (2019). Implementing big data strategies: A managerial perspective. *Business Horizons*, 62(3), 347–358.
- Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance: Toward a theory of information governance. *Journal of Management Information Systems*, 30(3), 141–178.
- Thompson, N., Ravindran, R., & Nicosia, S. (2015). Government data does not mean data governance: Lessons learned from a public sector application audit. *Government Information Quarterly*, 32(3), 316–322.
- Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from big data: A multiple-case study analysis of provider companies. *Technovation*, 84, 21–36.
- Wamba, S. F., & Akter, S. (2019). Understanding supply chain analytics capabilities and agility for data-rich environments. *International Journal of Operations & Production Management*, 39(6–7–8), 887–912.
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365.
- Weeserik, B. P., & Spruit, M. (2018). Improving operational risk management using business performance management technologies. *Sustainability*, 10(3), 640–659.
- Zarkadakis, G. (2020, November 10). “Data trusts” could be the key to better AI. *Harvard Business Review*. Recuperado de <https://hbr.org/2020/11/data-trusts-could-be-the-key-to-better-ai>
- Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models. Problems and solutions. *Organizational Research Methods*, 12(4), 695–719.



Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37, 197–206.

(Apêndice A)

OPERACIONALIZAÇÃO DOS CONSTRUTOS

Construto (referências)	Item	Carga	Descrição do item
<i>Big data</i> Alharthi et al. (2017) e Chen et al. (2014)	BD-01	,756	No ambiente de negócios da organização lida com um grande volume de dados.
	BD-02	,820	... os dados são gerados (coletados, criados) em alta velocidade.
	BD-03	,835	... lida com uma grande variedade (tipos e formatos) de dados.
	BD-04	,782	... lida com grande variabilidade de dados.
<i>Business analytics</i> Delen e Zolbanin (2018) e Duan et al. (2020)	BA-01	,799	Os dados de negócio são analisados por meio de técnicas descritivas (categorização, consolidação e classificação) para convertê-los em informações úteis para os gestores.
	BA-02	,847	... explorados para descobrir as causas de problemas específicos.
	BA-03	,863	... para identificar comportamentos e prever tendências.
	BA-04	,900	... identificar as melhores alternativas e otimizar objetivos.
<i>Administração de dados</i> Fleckenstein e Fellows (2018), Lillie e Eybers (2018), Khatri e Brown (2010), Koltay (2016), Plotkin (2014), Rosenbaum (2010), Surbakti et al. (2020) e Tallon et al. (2013)	AD-01	,725	Na organização adotam-se políticas de governança e gerenciamento de dados formalmente estabelecidas.
	AD-02	,790	... uma arquitetura de dados que permite identificar quais dados são coletados e em quais processos são utilizados.
	AD-03	,775	... uma estrutura organizacional, ou comitê de gestão, responsável pela qualidade dos dados e orientação aos usuários do BDA.
	AD-04	,721	... BDA em ambiente corporativo acessível, de forma compartilhada entre as áreas de negócios.
	AD-05	,799	... práticas para garantir a proteção dos dados utilizados na gestão ou operação do negócio.
	AD-06	,807	... compartilhamento de dados para reduzir “ilhas” ou “silos” de dados nas áreas de negócios.

(continua)

Construto (referências)	Item	Carga	Descrição do item
			O BDA...
Gestão do desempenho corporativo Acito e Khatri (2014), Chen et al. (2012), Mikalef et al. (2019), Richards et al. (2019) e Weeserik e Spruit (2018)	GDC-01	,843	... auxilia na formulação e no planejamento da estratégia de negócios.
	GDC-02	,792	... auxilia no controle e na avaliação do desempenho empresarial.
	GDC-03	,844	... contribui para agregar valor econômico-financeiro ao negócio.
	GDC-04	,842	... ajuda a otimizar a eficiência e eficácia dos processos internos de negócios.
	GDC-05	,818	... favorece a gestão do relacionamento com clientes, fornecedores e <i>stakeholders</i> .
	GDC-06	,845	... favorece o desenvolvimento e aprendizagem de outras capacidades organizacionais.

NOTAS DOS AUTORES

Mauricius M. Medeiros, mestre pelo Programa de Pós-Graduação em Administração (PPGAdm), Universidade do Vale do Rio dos Sinos (Unisinos); **Antonio C. G. Maçada**, doutor pelo Programa de Pós-Graduação em Administração (PPGA), Universidade Federal do Rio Grande do Sul (UFRGS); **Norberto Hoppen**, doutor pelo Departamento de Administração, Université Pierre Mendès France (UPMF).

Mauricius M. Medeiros é agora doutorando do PPGA-UFRGS; Antonio C. G. Maçada é agora professor do PPGA-UFRGS; Norberto Hoppen é agora professor do PPGAdm-Unisinos.

Correspondências sobre este artigo devem ser enviadas para Mauricius M. Medeiros, Rua Washington Luiz, 855, Centro Histórico, Porto Alegre, Rio Grande do Sul, Brasil, CEP 90010-460. *E-mail*: mauricius.medeiros@gmail.com

CORPO EDITORIAL

Editor-chefe
Gilberto Perez

Editor associado
Luís Hernan Contreras Pinochet

Suporte técnico
Vitória Batista Santos Silva

PRODUÇÃO EDITORIAL

Coordenação editorial
Jéssica Dametta

Preparação de originais
Carlos Villarruel

Revisão
Paula Di Sessa Vavlis

Diagramação
Emap

Projeto gráfico
Libro