

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Departamento de Estatística - Instituto de Ciências Exatas (ICEx)

Especialização em Estatística: Ênfase em Estatística

Pedro Victor Moreira de Souza

**CLASSIFICAÇÃO DAS ESCOLAS ESTADUAIS DE ENSINO MÉDIO DE MINAS
GERAIS SEGUNDO RESULTADOS EDUCACIONAIS EM 2021: UMA ANÁLISE
POR COMPONENTES PRINCIPAIS**

Belo Horizonte

2022

Pedro Victor Moreira de Souza

**CLASSIFICAÇÃO DAS ESCOLAS ESTADUAIS DE ENSINO MÉDIO DE MINAS
GERAIS SEGUNDO RESULTADOS EDUCACIONAIS EM 2021: UMA ANÁLISE
POR COMPONENTES PRINCIPAIS**

Versão final

Monografia apresentada ao Programa de
Especialização em Estatística: Ênfase em
Estatística da Universidade Federal de
Minas Gerais.

Orientadora: Prof.^a Dr.^a Edna Afonso Reis

Belo Horizonte

2022

2022, Pedro Victor Moreira de Souza.
Todos os direitos reservados.

Souza, Pedro Victor Moreira de

S729c Classificação das escolas estaduais de ensino médio de Minas Gerais segundo resultados educacionais em 2021: [recurso eletrônico] uma análise por componentes principais / Pedro Victor Moreira de Souza — 2022.
1 recurso online (70 f. il, color.): pdf.

Orientadora: Edna Afonso Reis.

Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.

Referências: 65-70

1. Estatística. 2. Ensino de segundo grau -Escolas publicas 3. Análise de regressão . I. Reis, Edna Afonso. II. Universidade Federal de Minas Gerais I. Instituto de Ciências Exatas, Departamento de Estatística .III.Título.

CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa CRB 6/1510
Universidade Federal de Minas Gerais – ICEX



Universidade Federal de Minas Gerais

E-mail: pgest@ufmg.br

Instituto de Ciências Exatas

Tel: 3409-5923 – FAX: 3409-5924

Departamento de Estatística

Programa de Pós-Graduação / Especialização

Av. Pres. Antônio Carlos, 6627 - Pampulha

31270-901 – Belo Horizonte – MG

ATA DO 268º. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE PEDRO VICTOR MOREIRA DE SOUZA.

Aos vinte dias do mês de dezembro de 2022, às 10:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Pedro Victor Moreira de Souza**, intitulado: “Classificação das Escolas Estaduais de Ensino Médio de Minas Gerais segundo Resultados Educacionais em 2021: Uma Análise Por Componentes Principais”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, a Presidente da Comissão, Professora Edna Afonso Reis – Orientadora, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado por unanimidade condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje. O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 20 de dezembro de 2022.

A handwritten signature in black ink, appearing to read "Edna Afonso Reis".

Prof.^a Edna Afonso Reis (Orientadora)
Departamento de Estatística / ICEX / UFMG

A handwritten signature in black ink, appearing to read "Ilka Afonso Reis".

Prof.^a Ilka Afonso Reis
Departamento de Estatística / ICEX / UFMG



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.

Declaro para os devidos fins que Pedro Victor Moreira de Souza, número de registro 2019705170, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística e que entregou para sua orientadora, a professora Edna Afonso Reis, que aprovou a versão final. O trabalho foi apresentado no dia 21 de dezembro de 2022 com o título “*Classificação das Escolas Estaduais de Ensino Médio de Minas Gerais segundo Resultados Educacionais em 2021: Uma Análise Por Componentes Principais*”.

Belo Horizonte, 03 de fevereiro de 2023

A handwritten signature in blue ink, which appears to read "Roberto da Costa Quinino".

Prof. Roberto da Costa Quinino
Coordenador do curso de
Especialização em Estatística
Departamento de Estatística / UFMG

RESUMO

Este trabalho categorizou as escolas estaduais de Ensino Médio de Minas Gerais, identificando aquelas de melhores e piores resultados educacionais, como o desempenho e o fluxo, os contextualizando com indicadores de infraestrutura, gestão escolar, corpo docente e nível socioeconômico. Para a realização deste estudo, utilizou-se indicadores educacionais em âmbito estadual e federal por meio da técnica de Análise de Componentes Principais (PCA). Nas últimas décadas, o acesso à escola foi quase universalizado, porém a garantia da qualidade de ensino continua sendo um desafio. A expansão de avaliações de larga escala e de indicadores de resultado permitiram a análise comparativa entre escolas. Contudo, a interpretação desses resultados sem a devida contextualização, especialmente do contexto socioeconômico, vem se mostrando limitada, na medida em que parte de uma premissa segundo a qual a escola age de forma isolada para superar as desigualdades sociais e educacionais. Diante desse cenário, a síntese de vários aspectos da realidade escolar permite uma aproximação melhor do esforço despendido pela escola no processo de aprendizagem e possibilita a identificação de grupos de indivíduos semelhantes, para fins de análises comparativas. As novas variáveis obtidas foram capazes de categorizar as escolas estaduais segundo perfil social e administrativo, a partir da combinação de características como rica/pobre com grande/pequeno porte, assim como de acordo com a qualidade da infraestrutura e formação docente, por meio da combinação infraestrutura boa/ruim com corpo docente mais/menos qualificado. Por meio desses grupos de combinações, os resultados educacionais foram comparados, indicando uma relação mais forte com o contexto social e de gestão do que com os recursos humanos e de infraestrutura. A capacidade de gestão escolar, estimada através de um indicador da avaliação dos alunos acerca do retorno às aulas presenciais pós-pandemia, também mostrou uma relação positiva com os resultados educacionais, porém menos forte do que o verificado no nível socioeconômico.

Palavras-chave: Desempenho Escolar; Componentes Principais; Nível socioeconômico; Técnicas de Estatística Multivariada; Escolas Estaduais; Infraestrutura Escolar; Gestão Escolar.

ABSTRACT

This study categorized Minas Gerais' public high schools, identifying those with the best and worst educational outcomes, such as performance and pass rate, in the context of infrastructure, school management, faculty, and socioeconomic level. To accomplish this study, federal and state-level educational indicators were used through the Principal Component Analysis method. In the last decades, school access has been nearly widespread, but ensuring education quality still is a challenge. Expansion in large-scale assessments and performance indicators allowed comparative analysis between schools. However, comprehension of those results without due context, particularly socioeconomic context, has shown limits, insofar as it assumes that schools act in isolation to overcome social and educational inequalities. Within this setting, the synthesis of multiple aspects of schools' reality allows a better approximation of their effort in the learning process and identifies groups of similar individuals for comparative analysis. The new variables obtained were capable of categorizing state schools by social-administrative profile, by the combination of characteristics such as rich/poor with large/small sized, as well by quality of infrastructure and faculty, through the combination of good/bad infrastructure with more/less-qualified staff. Through those groups of combinations, educational outcomes were compared, showing a stronger relation with social and management context than with human resources and infrastructure. School management capability, estimated by an indicator of student evaluation about the return to post-pandemic in-class learning, also shows a positive relationship with educational outcomes, although weaker than verified for socioeconomic level.

Keywords: Educational Development; Principal Components; Socioeconomic Level; Multivariate Statistics; High Schools; School Infrastructure; School Management.

LISTA DE ILUSTRAÇÕES

Figura 1: Bloco sobre retorno às aulas presenciais do questionário contextual do Simave 2021	17
Figura 2: Fluxo de cálculo da Análise de Componentes Principais	25
Figura 3: Matriz de correlação das variáveis padronizadas	31
Figura 4: Scree plot - Percentual da variância explicada por Componente Principal	33
Figura 5 - Scree plot - Seleção das novas variáveis pelo critério de Kaiser.....	34
Figura 6: Pesos CP1 (20,5% da variância explicada)	35
Figura 7: Pesos CP2 (14,4% da variância explicada)	36
Figura 8: Pesos CP3 (11,7% da variância explicada)	37
Figura 9: Pesos CP4 (9,5% da variância explicada)	38
Figura 10: Pesos CP5 (9,2% da variância explicada)	39
Figura 11: Qualidade do ajustamento das variáveis (a) e matriz de correlação entre variáveis e componentes (b)	40
Figura 12: Matriz de contribuição percentual das variáveis para os CPs	41
Figura 13: Ordenamento das variáveis por percentual de contribuição no CP1.....	42
Figura 14: Ordenamento das variáveis por percentual de contribuição no CP2.....	42
Figura 15: Ordenamento das variáveis por percentual de contribuição no CP3.....	43
Figura 16: Ordenamento das variáveis por percentual de contribuição no CP4.....	43
Figura 17: Ordenamento das variáveis por percentual de contribuição no CP5.....	44
Figura 18: Distribuição dos indivíduos na análise CP1 x CP2.....	45
Figura 19: Agrupamentos na análise CP1 x CP2	46
Figura 20: Comparação das escolas por agrupamento CP1-CP2.....	48
Figura 21: Distribuição dos indivíduos na análise CP2 x CP3.....	49
Figura 22: Agrupamentos na análise CP2 x CP3	49
Figura 23: Comparação das escolas por agrupamento CP2-CP3.....	51
Figura 24: Distribuição dos indivíduos na análise CP2 x CP4.....	52
Figura 25: Agrupamentos na análise CP2 x CP4	52
Figura 26: Comparação das escolas por agrupamento CP2-CP4.....	54
Figura 27: Distribuição dos indivíduos na análise CP1 x CP5.....	55
Figura 28: Agrupamentos na análise CP1 x CP5	55
Figura 29: Comparação das escolas por agrupamento CP1-CP5.....	57
Figura 30: Distribuição dos indivíduos na análise CP2 x CP5.....	58
Figura 31: Agrupamentos na análise CP2 x CP5	59
Figura 32: Comparação das escolas por agrupamento CP2-CP5.....	61

LISTA DE QUADROS

Quadro 1: Lista de ambientes considerados no Diagnóstico de Infraestrutura	15
Quadro 2: Eixos da análise e tabela resumo das variáveis escolhidas, suas fontes e ano de coleta.....	21
Quadro 3: Escala e Interpretação das variáveis.....	22
Quadro 4: Resumo da interpretação dos Componentes Principais selecionados	44

LISTA DE TABELAS

Tabela 1: Análise descritiva das variáveis.....	30
Tabela 2: Resumo descritivo dos agrupamentos CP1-CP2	47
Tabela 3: Resumo descritivo dos agrupamentos CP2-CP3	50
Tabela 4: Resumo descritivo dos agrupamentos CP2-CP4	53
Tabela 5: Resumo descritivo dos agrupamentos CP1-CP5	56
Tabela 6: Resumo descritivo dos agrupamentos CP2-CP5	60

LISTA DE SIGLAS

COVID-19: Coronavirus Disease 2019 (Doença do Coronavírus de 2019)

EAD: Ensino à Distância

Ideb: Índice de Desenvolvimento da Educação Básica

Inep: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

MEC: Ministério da Educação

PBF: Programa Bolsa Família

PCA: Análise de Componentes Principais

PROEB: Programa de Avaliação da Rede Pública de Educação Básica

TRI: Teoria da Resposta ao Item

Saeb: Sistema de Avaliação da Educação Básica

SEE-MG: Secretaria de Estado de Educação de Minas Gerais

SIMADE: Sistema Mineiro de Administração Escolar

Simave: Sistema Mineiro de Avaliação e Equidade da Educação Pública

SUMÁRIO

1.	INTRODUÇÃO	12
2.	INDICADORES EDUCACIONAIS	13
2.1.	Resultado Educacional	13
2.2.	Infraestrutura	15
2.3.	Gestão escolar.....	16
2.4.	Recursos humanos.....	18
2.5.	Contexto socioeconômico.....	19
3.	METODOLOGIA	21
3.1.	Banco de Dados	21
3.2.	Análise de Componentes Principais (PCA)	23
4.	RESULTADOS	30
4.1.	Estatísticas Descritivas	30
4.2.	Autovalores e autovetores	32
4.3.	Escolha do número de componentes principais	33
4.4.	Interpretação dos componentes principais	35
4.5.	Análise do ajuste das variáveis aos componentes principais	40
4.6.	Análise dos grupos de escolas caracterizados pelos componentes principais.....	45
	Análise 1: Porte e nível socioeconômico da escola (CP1) por resultado educacional (CP2).....	45
	Análise 2: Ajuste ou desajuste entre o desempenho (CP2) e o fluxo escolar (CP3).....	48
	Análise 3: Resultados educacionais (CP2) e condições de infraestrutura e corpo docente (CP4).....	51
	Análise 4: Relação entre o porte e nível socioeconômico da escola (CP1) e a adequação da comunidade escolar aos protocolos sanitários (COVID) (CP5) ..	54
	Análise 5: Relação entre os resultados educacionais (CP2) e a adequação da comunidade escolar aos protocolos sanitárias (COVID) (CP5).....	58
5.	CONCLUSÃO	62
6.	REFERÊNCIAS.....	65

1. INTRODUÇÃO

O direito à educação foi fundamentado na Constituição Federal de 1988 nos artigos 205 e 206, nos quais está definido que a Educação visa o pleno desenvolvimento da pessoa, o seu preparo para a cidadania e sua qualificação para o trabalho, tendo que ser baseada nos princípios da igualdade de condições de acesso e permanência, assim como da garantia do padrão de qualidade (BRASIL, 1988).

Contudo, a realização efetiva desse direito vem se mostrando um desafio para a sociedade brasileira, que observa a superação de determinados problemas históricos como a universalização do acesso, porém ainda esbarra na promoção da qualidade e equidade (SILVA e HASENBALG, 2000). Solucionar essa equação é entendido muitas vezes como o passo faltante para o desenvolvimento do país, assim como para a superação de suas desigualdades estruturais, colocando a avaliação educacional no centro desse debate.

A urgência crescente dessa discussão fez surgir, desde a redemocratização, vários esforços de mensurar aspectos da vida escolar com o objetivo de dar insumos para políticas públicas que visam a melhoria da qualidade educacional. Essa disponibilidade de vários indicadores facilita o processo de avaliação educacional, contudo pode apresentar também dois desafios: a dificuldade de analisar conjuntamente todos esses indicadores e a necessidade de contextualizar certos resultados educacionais mediante a realidade específica de cada escola.

A análise conjunta dos indicadores educacionais citada pode ser feita por meio de métodos de Análise Multivariada, que permitem sintetizar a informação contida em diversas variáveis. Já o segundo problema pode ser mitigado incluindo, no modelo de análise, indicadores de contexto acerca da realidade escolar nos seus aspectos socioeconômicos, de infraestrutura ou de corpo docente.

O presente trabalho se propôs a categorizar as escolas estaduais de Ensino Médio de Minas Gerais a partir de um levantamento dos principais indicadores educacionais disponíveis. Por meio da Análise de Componentes Principais (PCA), as escolas de resultados com maior destaque, positivo ou negativo, foram identificadas em grupos que podem ser utilizados para priorização em políticas públicas ou estudos futuros focalizados.

2. INDICADORES EDUCACIONAIS

O foco das políticas educacionais no Brasil tem se voltado gradualmente para a melhoria da qualidade do ensino, uma vez que o acesso à educação praticamente se universalizou nas últimas décadas com a ampliação das vagas disponíveis para a população, especialmente na rede pública (SOARES, RIGOTTI e ANDRADE, 2008). Nesse contexto, a criação de indicadores educacionais se mostrou um esforço institucional verificado em várias esferas da federação. Em âmbito nacional, o Sistema de Avaliação da Educação Básica (Saeb) tem coordenado um conjunto de avaliações externas em larga escala, enquanto o Sistema Mineiro de Avaliação e Equidade da Educação Pública (Simave) busca objetivos similares e complementares em âmbito estadual.

2.1. Resultado Educacional

Desde a criação do Simave em 2000, as escolas públicas de Minas Gerais, estaduais e municipais, participam de uma avaliação em larga escala por meio do Programa de Avaliação da Rede Pública de Educação Básica (PROEB). O 5º e 9º anos do Ensino Fundamental e o 3º ano do Ensino Médio são contemplados, buscando averiguar o conhecimento estimado dos estudantes em Língua Portuguesa e Matemática, com avaliações baseadas na Teoria da Resposta ao Item (TRI). A partir das proficiências dos alunos, é feita uma média aritmética para o cálculo da proficiência média das escolas (MINAS GERAIS, 2021).

A proficiência média pode ser então padronizada, conforme também é feito no cálculo do Índice de Desenvolvimento da Educação Básica (Ideb), para variar entre 0 e 10, considerando os limites inferiores e superiores do Saeb de 1997 para todo o Brasil (INEP, 2007). Além de facilitar a interpretação e comparação do **desempenho** das escolas, a padronização a partir dos resultados de 1997 permite uma análise temporal, novamente com o foco na melhoria da qualidade ao longo do tempo.

O Ideb também multiplica o desempenho padronizado dos alunos por um **indicador de fluxo**, que consiste na média (harmônica) das taxas de aprovação das etapas de um determinado nível de ensino (INEP, 2020). Por exemplo, no Ensino Médio, o indicador de fluxo é a média das taxas de aprovação do 1º, 2º e 3º ano. Logo, a qualidade aferida por um indicador como o Ideb é bidimensional: avalia-se o

aprendizado dos estudantes e a capacidade da escola de formar aquele aluno dentro da idade escolar ideal, evitando a reprovação ou o abandono.

Algo idêntico pode ser calculado com os resultados do Simave, que tem como âmbito somente o estado de Minas Gerais. Nesse caso, o desempenho no Saeb é substituído pelo PROEB e o indicador de fluxo se mantém o mesmo, aferido a partir de informações prestadas pelas escolas no Censo Escolar da Educação Básica, que ocorre anualmente. Neste trabalho, o estudo de caso focou nas escolas estaduais de Ensino Médio de Minas Gerais e foram utilizadas informações das avaliações de rede estaduais para o cálculo do desempenho.

Apesar da construção de um índice como o Ideb contemplar tanto o desempenho quanto o fluxo de aprovação escolar, existe um debate se essas dimensões seriam suficientes para aferir de fato a qualidade educacional ou se, na verdade, ele se limitaria a ser um indicador de resultado (CHIRINÉA e BRANDÃO, 2015).

Por exemplo, uma escola poderia alcançar notas altas a partir de uma política de não retenção dos alunos, de forma que somente a análise separada dos dois componentes do Ideb (desempenho e fluxo) evidenciariam o desajuste entre a aprovação e a aprendizagem dos estudantes.

Ampliando essa discussão, ao considerar somente esses dois componentes, como o Ideb faz, o que se avalia ao fim e ao cabo é a produtividade, desconsiderando outros aspectos da política educacional e da ação pedagógica (CHIRINÉA, 2010). O ambiente físico, a capacidade de gestão escolar, a formação e condições de trabalho dos profissionais da escola também são aspectos da qualidade, que inclusive interagem com a já referida produtividade dessa escola sendo avaliada.

Relacionada ao próprio indicador de fluxo já citado, outra medida que possibilita ampliar o conceito de qualidade é a **taxa de distorção idade-série**. Ela é o percentual de alunos enturmados em uma etapa de ensino cuja idade real difere da idade adequada para aquela série.

Os motivos para a ocorrência dessa defasagem são a reprovação, o abandono escolar ou a matrícula tardia do estudante na escola (PORTELLA, BUSSMANN e OLIVEIRA, 2017). Haveria uma relação entre esse problema de distorção e a evasão escolar.

2.2. Infraestrutura

O próprio ambiente físico da escola também pode ser objeto de análise para fins de mensuração da qualidade educacional. No âmbito das escolas estaduais, a Secretaria de Estado de Educação de Minas Gerais (SEE-MG) dispõe de dois indicadores que permitem estimar as condições de infraestrutura à disposição de alunos e funcionários escolares.

O primeiro é a **taxa média de ocupação das salas de aula**, que consiste no número de alunos por m², segundo informações disponíveis no Sistema Mineiro de Administração Escolar (SIMADE) e obtidas diretamente com a SEE-MG. O intuito desse indicador é medir a eventual superlotação de salas de aula, o que prejudicaria o aprendizado dos estudantes enturmados e a capacidade de gestão da turma por parte do(a) professor(a).

O segundo dado é o **Diagnóstico de Infraestrutura**, um índice construído pela SEE-MG para subsidiar políticas de investimento em rede física e manutenção predial, o qual também foi disponibilizado para este estudo. Ele é calculado a partir de questionários respondidos pelos inspetores escolares de Minas Gerais em visita a todos os endereços ativos da rede pública estadual.

Nesses questionários, os inspetores assinalam os problemas de infraestrutura existentes para os ambientes do prédio escolar listados no Quadro 1 (MINAS GERAIS, 2022). A partir das respostas, uma nota é calculada para cada um dos ambientes ou serviços e, por fim, um índice geral para a escola é obtido a partir da média ponderada da nota de todos os ambientes, pesando mais aqueles indispensáveis para o funcionamento escolar.

Quadro 1: Lista de ambientes considerados no Diagnóstico de Infraestrutura

Ambiente/Serviço			
Abastecimento de água	Biblioteca	Fachada	Pátio
Reservatório de água	Corredores	Sala de Aula	Quadra
Acessibilidade	Coleta de Esgoto	Laboratório de Informática	Fornecimento de Rede Elétrica
Auditório	Cozinha	Casa de Gás	Itens da Rede Elétrica
Banheiro	Despensa	Muro	Refeitório
Sala da Coordenação	Secretaria	Sala dos Professores	Tratamento de Esgoto
Sala da Direção	Telhado	Outros Ambientes	

Fonte: Secretaria de Estado de Educação de Minas Gerais (SEE-MG).

O índice geral da escola no Diagnóstico de Infraestrutura varia em uma escala de 0 a 1, na qual 0 representaria aquela escola para a qual o inspetor identificou todos os problemas do questionário, enquanto 1 representaria o exato oposto, em que nenhum problema foi identificado. A última campanha de visita dos inspetores escolares para preenchimento do Diagnóstico de Infraestrutura ocorreu no primeiro semestre de 2022 e serve de base para os dados que serão utilizados neste trabalho para o referido indicador.

2.3. Gestão escolar

Além das condições físicas, a capacidade de gestão escolar também é objeto de discussão no debate sobre qualidade educacional, colocando essa capacidade como sendo as condições objetivas e subjetivas de dispor insumos, concretizar objetivos pedagógicos e, em sentido ainda mais amplo, realizar a função social da escola como formadora de sujeitos engajados (PASSADOR e SALVETTI, 2013). Essa modalidade tem sido chamada de gestão escolar democrática e entende o objetivo da educação para além do aprendizado conteudista, mas sim do aluno como cidadão atuante.

O contexto em que isso acontece vai depender de características da escola que podem ser mensuradas por indicadores já disponíveis. Primeiramente, ao se tratar das escolas estaduais de Minas Gerais, a **oferta de alguma modalidade específica de ensino** implica em desafios adicionais à equipe docente. De acordo com o SIMADE, também fazem parte da rede estadual de Minas Gerais escolas socioeducativas, prisionais, exclusivas de educação especial, quilombolas e indígenas. Algumas dessas especificidades, como a educação indígena, podem exigir uma formação específica, intercultural e até bilíngue/multilíngue (FUNAI, 2020).

No caso das escolas da educação regular, também há variações consideráveis. O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) calcula para todas as escolas um **Indicador de complexidade de gestão** a partir do Censo Escolar da Educação Básica. Essa medida sintetiza quatro componentes: número de matrículas, número de turnos de funcionamento, complexidade das etapas ofertadas e número de etapas/modalidades oferecidas (INEP, 2014c). Dessa maneira, há uma contextualização do ambiente em que ocorre

a tentativa de implementação de uma gestão democrática ou mesmo meramente gerencial.

O contexto pandêmico vivido pelas escolas desde a substituição imposta pela COVID-19 das aulas presenciais por outras modalidades de ensino foi um desafio adicional a essa complexidade e objeto de avaliações educacionais como o Simave na edição de 2021. No questionário contextual que os estudantes respondem em paralelo aos cadernos das disciplinas avaliadas, foram adicionadas perguntas específicas acerca da opinião do aluno sobre o período de Ensino à Distância (EAD) e de retorno às aulas presenciais. A cada afirmação presente no questionário, os alunos tinham a opção de selecionar alternativas em uma escala desde “Não concordo” a “Concordo muito”.

A partir do bloco de perguntas descritos na Figura 1, referentes ao “Retorno às aulas presenciais”, calculou-se para este estudo um indicador de **avaliação do aluno sobre a adequação da comunidade escolar aos protocolos sanitários e ao novo contexto de ensino**.

O valor atribuído às respostas de cada aluno variou de 1 a 4, sendo o valor máximo correspondente à opção “Concordo muito” do questionário e assim sucessivamente. A nota geral da escola corresponde à média das respostas de todos os alunos que retornaram às aulas presenciais. O objetivo é que esse indicador sirva como uma *proxy* da capacidade de gestão escolar, indicando sua efetividade mesmo em um contexto adverso e inédito.

Figura 1: Bloco sobre retorno às aulas presenciais do questionário contextual do Simave 2021

Bloco 6 - Retorno as aulas presenciais

Pensando no retorno as aulas presenciais, você concorda com as seguintes afirmações:	Não concordo	Concordo pouco	Concordo	Concordo muito	Não houve retorno presencial na minha escola
57. Participei das discussões sobre as regras e o protocolo sanitário de prevenção à Covid-19 para o retorno das aulas presenciais em minha escola.	(A)	(B)	(C)	(D)	(E)
58. As pessoas em minha escola estão se relacionando bem com o retorno das atividades presenciais.	(A)	(B)	(C)	(D)	(E)
59. Os estudantes que retornaram para as atividades presenciais obedecem as regras e protocolo sanitário de prevenção à Covid-19 para o funcionamento da escola.	(A)	(B)	(C)	(D)	(E)
60. Os professores que retornaram para as atividades presenciais obedecem as regras e protocolo sanitário de prevenção à Covid-19 para o funcionamento da escola.	(A)	(B)	(C)	(D)	(E)

Fonte: Secretaria de Estado de Educação de Minas Gerais (SEE-MG)

2.4. Recursos humanos

Expandindo os ativos que a escola dispõe para atuar sobre o aprendizado do estudante, além da Infraestrutura e da Gestão Escolar, soma-se, por fim, o corpo docente da instituição. Para avaliar sua qualificação e condições de trabalho, optou-se por três indicadores calculados pelo Inep.

O primeiro deles é o indicador de **Adequação da Formação Docente**, que busca classificar os profissionais da Educação Básica de acordo com a sua formação acadêmica e a(s) disciplina(s) que lecionam. Com base nisso, são definidas categorias que estão posicionadas em uma escala que vai desde o docente que não possui curso superior completo, em desalinhamento ao Art. 62 da Lei de Diretrizes e Bases da Educação (BRASIL, 1996), até o profissional com formação superior de licenciatura na mesma disciplina que leciona ou bacharelado, desde que com curso de complementação pedagógica concluído.

A gradação dessa escala é útil para os sistemas de ensino elaborarem estratégias de adequação dos profissionais, considerando a realidade e histórico distinto de cada rede, assim como as contingências existentes na formação de professores no Brasil (INEP, 2014a).

Essas limitações dificultam a contratação e permanência de profissionais com formação alinhada ao prescrito nas normas vigentes, tornando indispensável a contribuição de docentes com licenciatura ou bacharelado em áreas distintas das disciplinas que lecionam ou até mesmo sem curso superior. Os desafios expostos não levam em conta nem mesmo a qualidade dessa formação, algo que também tem sido objeto de discussão crescente junto à expansão dos próprios cursos de licenciatura nas últimas décadas, como uma política contínua do MEC (BARRETO, 2015).

Outro indicador do Inep escolhido para caracterizar os recursos humanos das escolas é o **Esforço Docente**. Apesar de não ser uma variável que pode ser mensurada diretamente, esse indicador busca estimar, através da Teoria de Resposta ao Item, o esforço latente do professor segundo as seguintes variáveis do Censo da Educação Básica: número de escolas em que atua, número de turnos de trabalho, número de alunos atendidos e número de etapas nas quais leciona (INEP, 2014b). No maior nível de esforço, o docente teria mais de 400 alunos, atuaria nos três turnos de ensino, em duas ou três escolas e em duas ou três etapas (Ensino

Fundamental Anos Iniciais, Anos Finais ou Ensino Médio). O esforço docente médio de uma escola revelaria o tempo disponível de dedicação dos docentes para cada aluno, o que pode ter efeito no aprendizado e fluxo escolar.

Por fim, o impacto da atuação desse docente vai ser influenciado pela sua permanência ou não na escola por tempo suficiente para que tanto seu esforço, quanto sua formação tenham efeito significativo. Para mensurar isso, optou-se pelo indicador de **Regularidade do Corpo Docente**, também calculado pelo Inep, que atribui para os professores uma pontuação composta de dois fatores: a presença na escola em determinado ano, com pontuação maior para anos recentes, e a presença em anos consecutivos (INEP, 2015). Entende-se que a regularidade facilitaria o estabelecimento de vínculos com a escola e os alunos, assim como a possibilidade de dar continuidade a planejamentos e estratégias pedagógicas.

2.5. Contexto socioeconômico

Até este ponto, o desempenho, o fluxo e a trajetória escolar do aluno foram colocados em perspectiva através de atributos da escola que os condicionariam. Logo, a infraestrutura, a capacidade de gestão escolar e os recursos humanos, especialmente o corpo docente, colaborariam ou competiriam entre si para o sucesso dos estudantes. Contudo, esses atributos não atuam em uma população homogênea e a literatura evidencia o quanto as desigualdades educacionais são também uma expressão das desigualdades sociais no Brasil (GARCIA e HILESHEIM, 2017).

A discussão sobre a eficácia escolar pode ser retomada à publicação de um extenso “*survey*” contratado nos Estados Unidos pelo “*Civil Rights Act*” de 1964 (COLEMAN, 1966), acerca da oferta de oportunidades educacionais a minorias étnicas naquele país. A conclusão final do estudo não conseguiu apontar que as escolas faziam diferença significativa na formação dos estudantes, sendo mais importante o seu contexto socioeconômico. Elas estariam apenas reproduzindo as desigualdades sociais já existentes na sociedade.

Esse debate se estende até tempos presentes em que também se questiona a capacidade do Ideb, ao considerar somente o desempenho dos alunos e o fluxo escolar como medida sintética de qualidade, de traduzir de fato o “resultado” do trabalho feito pela escola. Assume-se, dessa forma, uma premissa de que a escola

seria capaz de suprimir ou superar toda a desigualdade promovida socialmente (ALMEIDA, DALBEN e FREITAS, 2013).

Tal premissa não seria verdadeira, como indicam estudos como o de Silva e Hasenbalg (2000), que decompuseram os fatores explicativos da melhoria educacional, já verificada nas décadas de 80 e 90 do século passado, e encontraram que aproximadamente 60% do sucesso alcançado nos indicadores educacionais podiam ser atribuídos às mudanças nas condições de vida e distribuição geográfica das famílias brasileiras. Somente os 40% restantes estariam ligados ao sistema educacional.

Considerando o exposto, optou-se por incluir na análise deste estudo um **Índice de Nível Socioeconômico** dos alunos da rede estadual de Minas Gerais. Ele é construído a partir da resposta dos estudantes ao questionário contextual do Simave e inclui perguntas sobre características pessoais do aluno (sexo e raça), nível de escolaridade dos responsáveis, condições de moradia, renda e acesso à cultura. A partir desses resultados, é calculado o Nível Socioeconômico de cada escola (MINAS GERAIS, 2021). Esse índice contribui para contextualizar os indicadores de resultado por conterem determinantes extraescolares desse próprio desempenho, tais quais o Capital Econômico, Cultural e Social-Familiar dos estudantes (SILVA e HASENBALG, 2000), os quais atuariam como um filtro sobre o efeito da escola na aprendizagem.

A compreensão dessa interação se mostra importante, não só pela sua magnitude, como já exposto, mas também pelo fato de a educação ser comumente vislumbrada como um meio de superação das condições da pobreza, sendo que a evidência tende a sugerir que é principalmente a superação da pobreza que contribuiu na melhoria dos indicadores educacionais brasileiros nas últimas décadas. Entretanto, o que se verifica na administração pública ainda é uma fragmentação dessas duas frentes, ao invés da articulação de políticas de forma transversal (GARCIA e HILLESHEIM, 2017).

Uma exceção foi o Programa Bolsa Família (PBF), uma política de transferência condicionada de renda, em que as famílias precisavam assumir uma série de compromissos especialmente na área da saúde e da educação. Uma revisão sistemática de estudos sobre o PBF indicou que, apesar da ausência de efeito na proficiência e desempenho, houve uma melhoria significativa nas taxas de abandono e na frequência escolar (SANTOS et al, 2019).

3. METODOLOGIA

3.1. Banco de Dados

Para facilitar a interpretação, os indicadores foram organizados em cinco eixos relevantes para a contextualização da qualidade do ensino: Resultado Educacional, Infraestrutura, Gestão Escolar, Recursos Humanos e Contexto Socioeconômico, sendo cada eixo formado por uma a três variáveis.

Quadro 2: Eixos da análise e tabela resumo das variáveis escolhidas, suas fontes e ano de coleta

Área/Eixo		Sigla	Indicador	Fonte	Ano
Resultado Educacional	1	desemp	Indicador de desempenho	SIMAVE	2021
	2	fluxo	Indicador de fluxo	Inep	2021
	3	dis	Taxa de distorção idade-série	SIMADE	2022
Infraestrutura	4	tx_ocup	Taxa média de ocupação das salas de aula	SIMADE	2022
	5	infra	Diagnóstico de infraestrutura	SEE	2022
Gestão Escolar	6	espec	Oferta de modalidade de ensino específica*	SIMADE	2022
	7	icg	Complexidade de gestão da escola	Inep	2021
	8	avl_ret	Indicador de avaliação do retorno às aulas presenciais	SIMAVE	2021
Recursos Humanos	9	afd	Adequação da formação docente	Inep	2021
	10	ed	Esforço docente	Inep	2021
	11	rd	Regularidade do corpo docente	Inep	2021
Contexto Socioeconômico	12	inse	Nível socioeconômico	SIMAVE	2021

Nota: *Indicador usado apenas para excluir as escolas com modalidade especial.

Fonte: Elaborado pelo autor.

Dos 12 indicadores descritos na Quadro 2, foram utilizados 11 na Análise de Componentes Principais. O indicador “oferta de modalidade de ensino específica” (eixo Gestão Escolar) foi utilizado apenas para excluir as escolas que ofertam alguma modalidade especial de ensino, como Educação Indígena ou Quilombola, pois, conforme detalhado no Capítulo 2, acredita-se que essas escolas possuem

uma realidade muito particular das demais, beneficiando-se de uma análise separada em outro estudo. Logo, apenas as escolas de Ensino Médio e com a modalidade de Ensino Regular serão consideradas.

Do total de 2.387 escolas estaduais de Ensino Médio, 60 escolas foram excluídas da base por ofertarem alguma modalidade específica, restando 2.327 escolas de Ensino Regular. Mais 196 escolas foram retiradas da base por não possuírem informações para os indicadores selecionados, ao fim restando 2.131 unidades de ensino para a análise.

A interpretação dos indicadores selecionados consta no Quadro 3, a seguir.

Quadro 3: Escala e Interpretação das variáveis

Variável	Escala	Interpretação
Indicador de desempenho	0 a 10	Média da nota padronizada em Língua Portuguesa e Matemática em escala que varia de 0 a 10. Quanto mais próximo de 10, maior a proficiência dos alunos e, quanto mais próximo de 0, menor é a proficiência.
Indicador de fluxo	%	Média harmônica da porcentagem de alunos aprovados em todas as etapas do nível de ensino. Quanto maior, melhor.
Taxa de distorção idade-série	%	Porcentagem de alunos com idade superior em 2 anos ou mais à idade recomendada para sua etapa de enturmação. Quanto menor a taxa de distorção idade-série, melhor.
Taxa média de ocupação das salas de aula	Alunos/m ²	Quociente de alunos pela capacidade física das salas de aula, medida em metros quadrados (m ²). Valores maiores indicam salas com menos espaço disponível.
Diagnóstico de infraestrutura	0 a 1	Quanto mais próximo de 0, um número maior de problemas de infraestrutura foi identificado pela Inspeção Escolar. Quanto mais próximo de 1, melhor é a estrutura física do prédio.
Complexidade de gestão da escola	1 a 6	Quanto mais próximo de 1, a escola possui menos matrículas, turnos, etapas e modalidades, e quanto mais próximo de 6, maior é o porte/tamanho da escola.
Indicador de avaliação do retorno às aulas presenciais	1 a 4	Quanto mais próximo de 4, mais alunos marcaram a opção "Concordo muito" no bloco de questões sobre o retorno às aulas presenciais, indicando uma maior adequação da escola aos protocolos sanitários. Quanto mais próximo de 1, mais alunos marcaram a opção "Não concordo".
Adequação da formação docente	%	Porcentagem de professores com formação superior em licenciatura (ou bacharelado com complementação pedagógica) na mesma área da disciplina que leciona. Quanto maior, melhor.
Esforço docente	1 a 6	Quanto mais próximo de 1, menor é o esforço despendido em média pelos docentes da escola, e quanto mais próximo de 6, maior é o esforço.
Regularidade do corpo docente	0 a 5	Quanto mais próximo de 0, mais irregular é o vínculo dos professores com a escola e quanto mais próximo de 5, mais regular é esse vínculo.
Nível socioeconômico	1 a 10	Quanto mais próximo de 10, maior é o nível socioeconômico dos alunos, e quanto mais próximo de 1, menor.

Fonte: Elaborado pelo autor.

Optou-se por dar preferência a informações de âmbito estadual, a mesma esfera de gestão das escolas que compõem a amostra, devido à capacidade do Estado de Minas Gerais ter gerência sobre os indicadores produzidos e sobre as políticas educacionais que afetariam os resultados observados. Por conta disso, a fonte da maioria das informações foi a própria Secretaria de Estado de Educação (SEE-MG) ou sistemas sob sua responsabilidade, SIMADE e Simave. Assim sendo, este estudo poderia ser interpretado como um exercício teórico de categorização das escolas para fins de políticas públicas em âmbito estadual.

Para os demais indicadores, a fonte foi o Inep, em nível federal; contudo, os dados que servem de subsídio para os cálculos realizados pelo instituto provêm do Censo da Educação Básica, cuja fonte primária são as próprias escolas.

Por fim, todos os cálculos foram realizados através da linguagem de programação R e dos pacotes “FactoMineR”, “factoextra” e “MultivariateAnalysis”.

3.2. Análise de Componentes Principais (PCA)

Conhecida mais popularmente como PCA, a Análise de Componentes Principais foi desenvolvida em 1901 por Karl Pearson e aplicada posteriormente em artigo de Hotelling (1933). O seu uso mais comum está associado a problemas que exigem a observação, ao mesmo tempo, de diversas variáveis ou características de um processo (CABRAL, 2018), o que se alinha ao desafio levantado no Capítulo 2 sobre a necessidade de contextualizar o resultado educacional a partir de outras medidas da qualidade ou contexto escolar.

Essa metodologia já foi aplicada com sucesso em problemas semelhantes, como na análise do desempenho das escolas municipais de Ribeirão Preto (RAMOS et al, 2021), assim como em outras áreas, tais quais a pesquisa científica na Educação Superior (TEIXEIRA, TUPY e AMARAL, 2016) e a Saúde Suplementar (SANTOS, 2014).

O objetivo principal da PCA, segundo Mingoti (2005), é descrever a estrutura de variância e covariância de um vetor aleatório, que é composto por p -variáveis aleatórias, através da obtenção de combinações lineares das variáveis iniciais. Essas combinações lineares recebem o nome de Componentes Principais (CPs) e possuem como propriedade serem não correlacionados entre si e de explicarem a maior parte da variação nos primeiros componentes (principais). O número de

componentes mantidos em um sistema determinará a qualidade da aproximação, porém os primeiros tendem a ser já suficientes para explicar a maior parte da variabilidade. Dessa maneira, o sistema de variabilidade do vetor aleatório que contém os k componentes principais ($k < p$) se aproxima da variabilidade do sistema do vetor composto das p -variáveis originais (MINGOTI, 2007).

Os componentes principais são obtidos a partir de um processo de decomposição da matriz de covariâncias de uma população ($\Sigma_{p \times p}$) para o cálculo de seus autovalores e autovetores.

A matriz de variâncias e covariâncias ($\Sigma_{p \times p}$) não é conhecida na prática (MINGOTI e SILVA, 1997) e precisa ser estimada através dos dados amostrais, gerando a matriz de covariâncias amostral $S_{p \times p}$. Denota-se a matriz de covariância amostral como:

$$\widehat{Cov}(X) = S_{p \times p} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1} & \hat{\sigma}_{p2} & \cdots & \hat{\sigma}_{pp} \end{bmatrix} \quad (1)$$

em que $\hat{\sigma}_{ij} = \frac{\sum_{k=1}^n (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}{1-n}$, $i \neq j$, é um estimador não viciado da covariância populacional entre as variáveis X_i e X_j .

Geralmente, os dados são primeiro transformados através de um processo de padronização pelas respectivas médias e desvios padrões, gerando novas variáveis com médias iguais a 0 e variâncias iguais a 1.

Dado um vetor de variáveis originais $X = (X_1, X_2 \dots X_p)$ que contenha as p variáveis de interesse observadas em n elementos amostrais, a padronização é realizada da seguinte maneira:

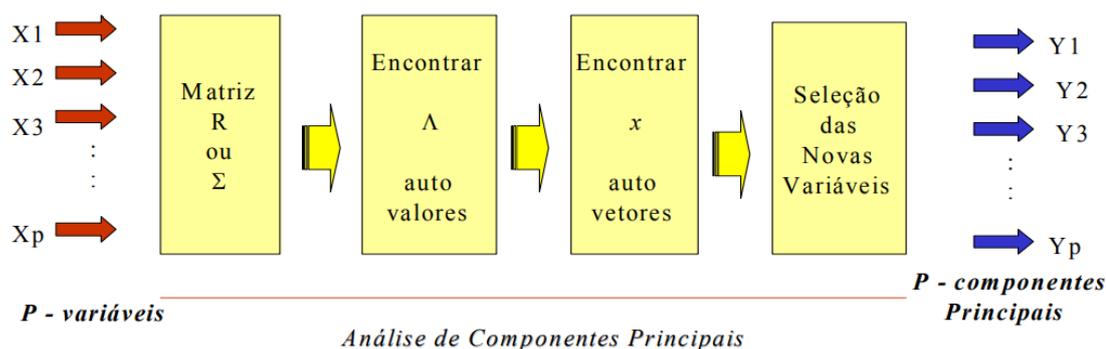
$$Z = (Z_1, Z_2, \dots, Z_p), \text{ em que } Z_i = \frac{X_i - \hat{\mu}_i}{\hat{\sigma}_i} \quad (2)$$

É importante notar que a matriz Z das variáveis padronizadas é igual à matriz de correlação $R_{p \times p}$ das variáveis X_i originais, $i = 1, 2 \dots, p$. Logo, aplicar a técnica de componentes principais à matriz de covariâncias das variáveis padronizadas equivale a obter os componentes principais através da matriz de correlação $R_{p \times p}$ das variáveis originais (MINGOTI, 2007). O procedimento de padronização das variáveis

é recomendado no caso de variáveis em escalas e unidades de medidas diferentes entre si (REGAZZI, 2000).

A Figura 2 esquematiza o restante do processo para obtenção dos componentes principais.

Figura 2: Fluxo de cálculo da Análise de Componentes Principais



Fonte: VICINI (2005)

Portanto, a partir da matriz de covariâncias das variáveis padronizadas (Σ_{pxp}), conhecida através da sua respectiva matriz amostral (S_{pxp}), é possível encontrar os autovalores por meio da solução da equação característica da matriz (VARELLA, 2008). Sendo $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \dots, \hat{\lambda}_p$ as raízes da equação característica da matriz S_{pxp} , então:

$$\hat{\lambda}_1 > \hat{\lambda}_2 > \hat{\lambda}_3, \dots, \hat{\lambda}_p \quad (3)$$

Cada valor $\hat{\lambda}_j$ possuirá um autovetor \hat{e}_j , de tal forma que:

$$\hat{e}_j = \begin{bmatrix} \hat{e}_{j1} \\ \hat{e}_{j2} \\ \vdots \\ \hat{e}_{jp} \end{bmatrix} \quad (4)$$

Visto isso, sendo $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \dots, \hat{\lambda}_p$ os autovalores da matriz S_{pxp} e sendo $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_p$ os respectivos autovetores normalizados, então o j -ésimo componente principal amostral estimado é definido como:

$$\hat{Y}_j = \hat{e}'_j X = \hat{e}_{j1} X_1 + \hat{e}_{j2} X_2 + \dots + \hat{e}_{jp} X_p \quad j = 1, 2, \dots, p \quad (5)$$

Considere-se também o vetor aleatório $Y = O'X$, no qual $O_{p \times p}$ corresponde à matriz ortogonal de dimensão $p \times p$ formada pelos autovetores da matriz $S_{p \times p}$, isto é:

$$O_{p \times p} = \begin{bmatrix} \hat{e}_{11} & \hat{e}_{21} & \cdots & \hat{e}_{p1} \\ \hat{e}_{12} & \hat{e}_{22} & \cdots & \hat{e}_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{e}_{1p} & \hat{e}_{2p} & \cdots & \hat{e}_{pp} \end{bmatrix} = [\hat{e}_1 \hat{e}_2 \dots \hat{e}_p] \quad (6)$$

Dessa forma, o vetor Y é composto por p combinações lineares das variáveis aleatórias do vetor X , sendo possível obter sua matriz de covariâncias $\Lambda_{p \times p}$, a qual será sempre uma matriz diagonal com elementos iguais a $a_{ii} = \lambda_i$, $i = 1, 2, \dots, p$, ou seja:

$$\Lambda_{p \times p} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \quad (7)$$

Conclui-se, portanto, que as variáveis aleatórias que compõem o vetor Y não são correlacionadas entre si. Por meio dessa propriedade, percebe-se o benefício de utilizar essas combinações lineares de Y como uma forma alternativa de representar as covariâncias do vetor X , reduzindo o número de variáveis de uma dimensão p para uma k menor que p . Como tanto o vetor X , quanto o vetor Y explicam a variabilidade do fenômeno, porém o vetor Y é composto por variáveis não correlacionadas entre si, a interpretação conjunta das suas variáveis se torna mais fácil (MINGOTI, 2007).

Logo, é possível identificar quatro propriedades importantes dos componentes principais, segundo Varella (2008):

1. A variância do componente principal \hat{Y}_j é igual ao valor do autovalor $\hat{\lambda}_j$.

$$Var(\hat{Y}_j) = \hat{\lambda}_j \quad (8)$$

2. Os componentes principais não são correlacionados.

$$Cov(\hat{Y}_i, \hat{Y}_j) = 0, \quad \text{para qualquer } i \neq j \quad (9)$$

3. O primeiro componente é o que apresenta maior variância e assim sucessivamente.

$$\text{Var}(\hat{Y}_1) > \text{Var}(\hat{Y}_2) > \dots > \text{Var}(\hat{Y}_p) \quad (10)$$

4. A variância total das variáveis originais é igual à soma dos autovalores, que por sua vez é igual ao total de variância dos componentes principais.

$$\sum \text{Var}(X_j) = \sum \hat{\lambda}_j = \sum \text{Var}(\hat{Y}_j) \quad (11)$$

Como último passo, é necessário fazer a seleção das novas variáveis que farão parte do sistema de dimensão k ($k < p$). Para isso, é preciso estimar a importância de cada componente a partir do cálculo da proporção da variância total explicada pelo componente, que nada mais é que a razão entre o seu autovalor e a variância total:

$$\frac{\hat{\lambda}_j}{\sum_{j=1}^p \hat{\lambda}_j}, j = 1, 2, \dots, p \quad (12)$$

Busca-se o máximo de variância explicada com o menor número desejável de componentes. Para auxiliar na análise, é possível elaborar um Diagrama de Autovalores ou um Scree Plot, que é a representação gráfica da curva da função que relaciona a ordem e os autovalores, a fim de identificar o ponto em que ocorre a ruptura da queda nos autovalores ordenados (FLECK e BOURDEL, 1998). Dois autovalores que possuem valores muito próximos, e que, portanto, representam uma proporção de variabilidade muito parecida, podem sugerir a opção de conservar somente um dos fatores.

Outro critério, o qual pode ser incluído também na visualização do Diagrama de Autovalores, é o critério de Kaiser (1958), um dos mais utilizados. Ele sugere que somente sejam considerados os componentes com autovalores maiores que 1.

Como o autovalor é a variância captada pelo componente, um autovalor menor que 1 indica um componente que abarca menos que a média das variâncias, que é igual a 1, pois a variância total equivale ao número de variáveis utilizadas e, portanto, de componentes.

Dessa forma, garante-se que eles explicam pelo menos a média da variância dos componentes e que, portanto, são relevantes para a interpretação do modelo.

Estabelecer um alvo de proporção mínima total explicada pelas variáveis mantidas no modelo pode também auxiliar na seleção dos fatores. Convenciona-se, em alguns casos, o mínimo de 70% da variância total sendo explicada pelos componentes principais mantidos (REGAZZI, 2000). Contudo, conjuntos de variáveis com baixa correlação entre si podem aumentar a dificuldade de explicar uma maior parte da variabilidade utilizando somente os primeiros componentes principais. Nesses casos e em determinadas áreas de estudo de maior variabilidade, como as Ciências Sociais, pode-se almejar um mínimo mais baixo de variância explicada pelos fatores no sistema.

Por fim, também deve-se levar em conta a existência de fatores interpretáveis, ou seja, aqueles para os quais claramente é observado um significado e que auxiliam para a explicação do fenômeno estudado, ainda que contribuam menos para a proporção da variância total explicada.

A interpretação de cada componente é realizada a partir do grau de influência que cada variável X_p tem sobre o componente \hat{Y}_j , o qual se dá pela correlação entre os dois (VARELLA 2008). Essa correlação é calculada da seguinte maneira:

$$\text{Corr}(X_p, \hat{Y}_j) = r_{pj} * \hat{Y}_j = \hat{e}_{jp} * \frac{\sqrt{\text{Var}(\hat{Y}_j)}}{\sqrt{\text{Var}(X_p)}} = \sqrt{\hat{\lambda}_j} * \frac{\hat{e}_{jp}}{\sqrt{\text{Var}(X_p)}} \quad (13)$$

Compara-se a influência de diferentes variáveis X_p no componente \hat{Y}_j através do seu peso, cujo cálculo é dado por:

$$w_1 = \frac{\hat{e}_{11}}{\sqrt{\text{Var}(X_1)}}, w_2 = \frac{\hat{e}_{12}}{\sqrt{\text{Var}(X_2)}}, \dots, w_p = \frac{\hat{e}_{jp}}{\sqrt{\text{Var}(X_p)}}, \text{ sendo } w_1 \text{ o peso de } X_1 \quad (14)$$

Ou seja, o peso da variável X_p na combinação linear \hat{Y}_j é o elemento do autovetor \hat{e}_j associado a ela padronizado pelo seu desvio-padrão.

Outra medida utilizada é o cosseno ao quadrado (\cos^2) de um componente. Essa medida corresponde à contribuição do componente para a distância ao quadrado de uma observação em relação a sua posição de origem no plano

cartesiano, uma vez que o vetor das novas variáveis passou por uma transformação linear ortogonal. Essa medida permite avaliar a qualidade do ajuste da PCA referente a determinado componente para cada observação.

O seu cálculo advém do cosseno ao quadrado correspondente ao ângulo do triângulo formado pela origem, a observação e sua projeção no componente. Quanto maior for o valor do cosseno ao quadrado do componente, maior é a proporção da sua contribuição para a distância total e, portanto, para a interpretação da observação (ABDI e WILLIAMS, 2010).

4. RESULTADOS

4.1. Estatísticas Descritivas

Primeiramente, calculou-se a média e os limites mínimo e máximo de cada variável, a fim de conhecer, de forma geral, as características da amostra de escolas estaduais utilizada neste estudo. Os resultados se encontram na Tabela 1, a seguir.

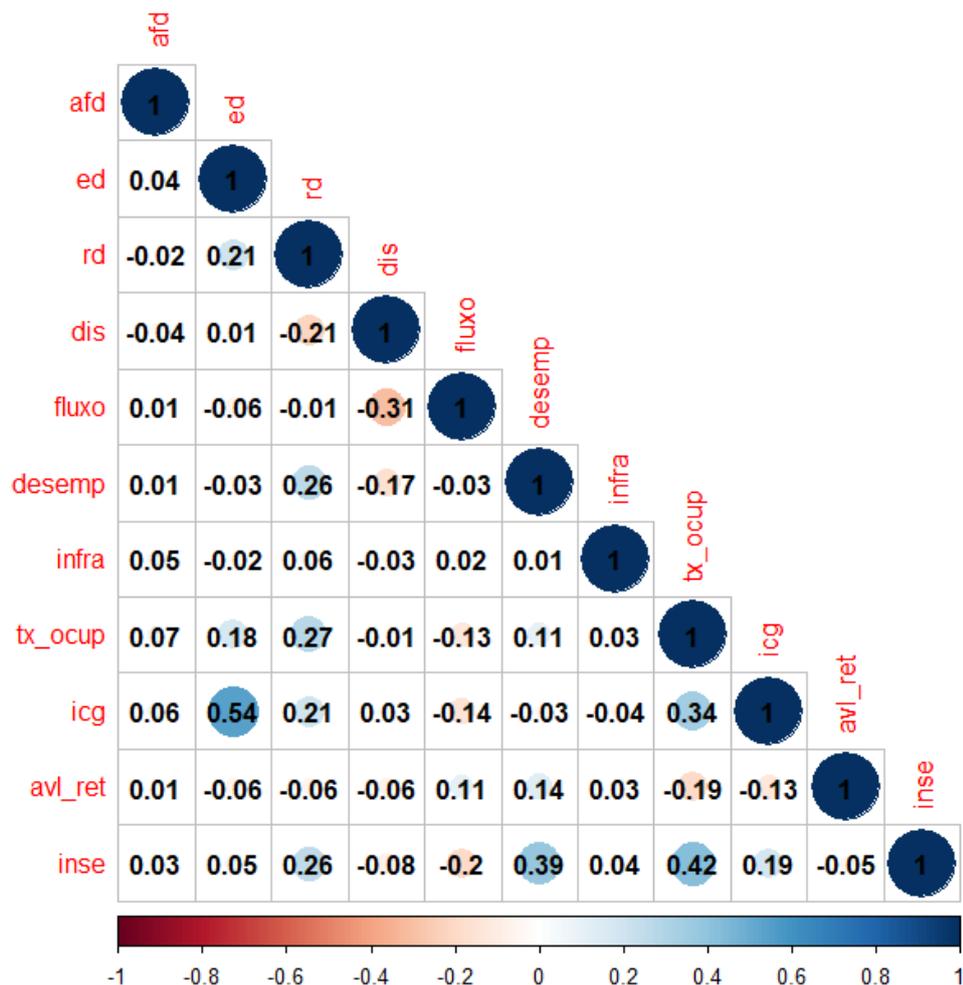
Tabela 1: Análise descritiva das variáveis

Variável	Sigla	Limite Mínimo	Média	Limite Máximo
Indicador de desempenho	desemp	2,20	4,11	6,99
Indicador de fluxo	fluxo	0,00	91,71	100,00
Taxa de distorção idade-série	dis	0,00	15,74	58,49
Taxa média de ocupação das salas de aula	tx_ocup	0,10	0,57	0,92
Diagnóstico de infraestrutura	infra	0,27	0,83	1,00
Complexidade de gestão da escola	icg	1,00	4,50	6,00
Indicador de avaliação do retorno às aulas presenciais	avl_ret	2,19	2,79	3,48
Adequação da formação docente	afd	18,80	71,50	100,00
Esforço docente	ed	2,25	3,93	5,00
Regularidade do corpo docente	rd	1,40	2,97	4,30
Nível socioeconômico	inse	3,13	5,83	8,08

Fonte: Elaborado pelo autor.

As 11 variáveis utilizadas neste trabalho foram padronizadas, gerando um vetor de variáveis com médias iguais a 0 e variâncias iguais a 1, para o qual foi calculada a matriz de correlação (Figura 3).

Figura 3: Matriz de correlação das variáveis padronizadas



Fonte: Elaborado pelo autor.

Conforme mostrado na Figura 3, a correlação das variáveis é baixa na maioria dos casos. As exceções são somente para combinações em que já se esperaria uma correlação mais significativa, como a Complexidade de Gestão da Escola (icg) e o Esforço Docente (ed), uma vez que ambos levam em conta o porte da escola para o seu cálculo; ou o Indicador de Fluxo (fluxo) e a Distorção Idade-Série (dis), ambas relacionadas às taxas de aprovação. Também é interessante notar que se verificou, a partir dos dados das escolas estaduais de Minas Gerais, uma correlação positiva, ainda que não muito forte, entre o Nível Socioeconômico (inse) e o Indicador de Desempenho (desemp).

Para a Análise de Componentes Principais, foi utilizada a matriz de variâncias e covariâncias, porém, conforme exposto na seção anterior, a matriz de correlação

dos dados originais equivale à matriz de variâncias e covariâncias dos dados padronizados, podendo as duas serem utilizadas de forma intercambiável para a obtenção dos autovalores e autovetores.

4.2. Autovalores e autovetores

A Tabela 2 traz a relação desses resultados para os 11 componentes gerados, ou seja, um componente por variável original presente no estudo.

Tabela 2: Resultado numérico da análise por componentes principais (CP)

Variável		CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11
Indicador de desempenho	desemp	-0,233	-0,497	-0,263	-0,115	-0,249	0,070	0,220	0,343	-0,440	0,426	-0,090
Indicador de fluxo	fluxo	0,188	-0,317	0,545	-0,014	0,205	0,066	-0,391	-0,056	-0,571	-0,182	-0,073
Taxa de distorção idade-série	dis	0,062	0,496	-0,410	0,070	-0,231	-0,113	-0,050	-0,298	-0,631	-0,135	-0,051
Taxa média de ocupação das salas de aula	tx_ocup	-0,468	0,032	-0,106	0,101	0,257	0,082	-0,519	-0,341	0,028	0,499	0,223
Diagnóstico de infraestrutura	infra	-0,023	-0,137	0,011	0,652	0,055	-0,726	-0,024	0,148	-0,008	0,040	-0,033
Complexidade de gestão da escola	icg	-0,440	0,312	0,299	-0,034	-0,188	-0,012	-0,136	0,168	0,058	0,032	-0,729
Avaliação do retorno às aulas presenciais	avl_ret	0,156	-0,293	0,052	-0,009	-0,781	-0,075	-0,329	-0,332	0,230	0,006	0,008
Adequação da formação docente	afd	-0,068	-0,014	0,110	0,723	-0,134	0,617	0,223	-0,092	-0,034	-0,048	0,003
Esforço docente	ed	-0,351	0,280	0,446	-0,076	-0,319	-0,112	0,115	0,238	-0,139	-0,043	0,623
Regularidade do corpo docente	rd	-0,392	-0,261	0,143	-0,128	0,094	-0,196	0,499	-0,634	-0,043	-0,187	-0,081
Nível socioeconômico	inse	-0,432	-0,239	-0,361	0,018	0,014	0,081	-0,294	0,215	0,050	-0,690	0,083
Autovalor		2,257	1,585	1,285	1,047	1,014	0,958	0,729	0,635	0,617	0,454	0,417
Variância Explicada (%)		20,52	14,41	11,69	9,52	9,22	8,71	6,63	5,77	5,61	4,13	3,79
Variância Explicada Acumulada (%)		20,52	34,93	46,61	56,14	65,36	74,06	80,69	86,46	92,08	96,21	100,0

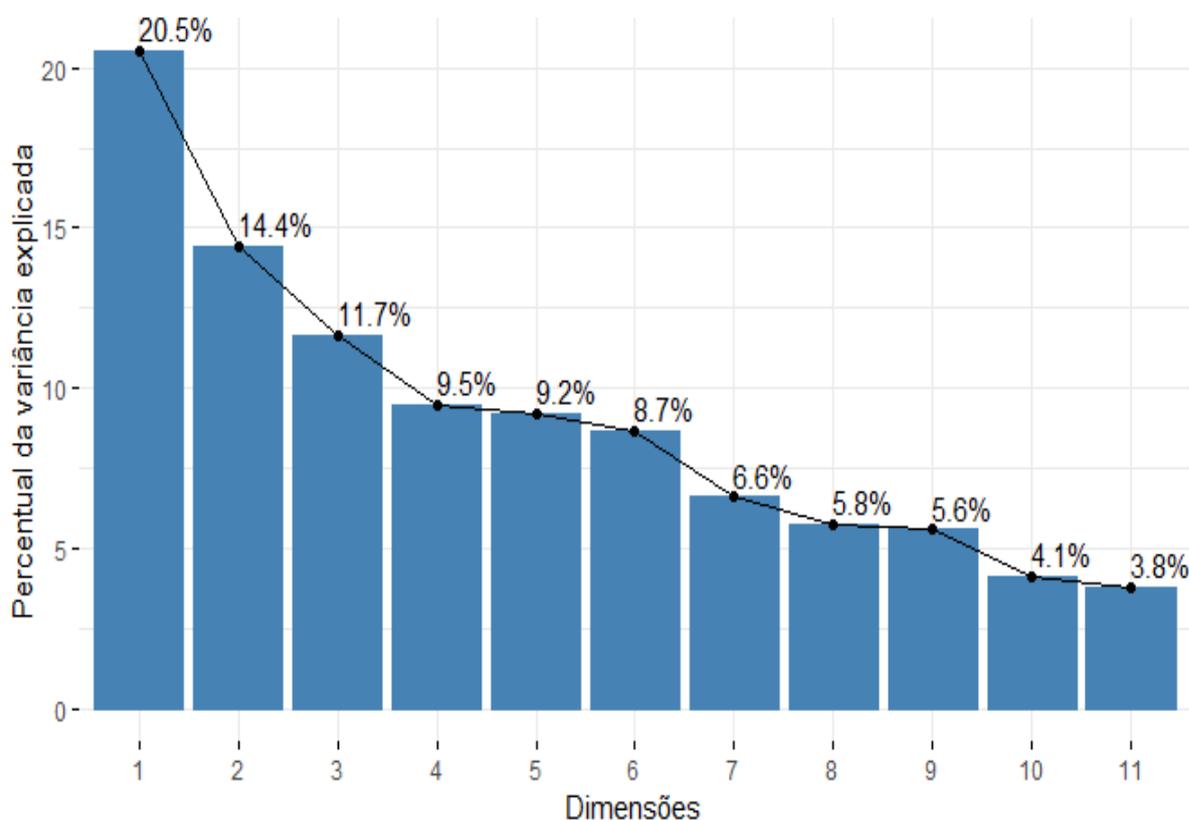
Fonte: Elaborado pelo autor.

4.3. Escolha do número de componentes principais

Para seleccionar os componentes mantidos no sistema, foram adotados todos os critérios sugeridos na metodologia: o critério de Kaiser, a análise da curva observada no Scree plot a partir da percentagem da variância explicada por cada componente, o número de componentes necessário para alcançar um total próximo de 70% da variabilidade total explicada no novo sistema e, por fim, a utilidade de interpretação dos componentes pela análise dos pesos de cada variável.

Observando o Scree plot (Figura 4), percebe-se que a percentagem da variância explicada por cada componente tem uma queda sucessiva até por volta do 4ª ao 6ª componente. É nesse ponto que a ruptura da queda da curva da função que relaciona a ordem e os autovalores ocorre pela primeira vez, sugerindo que o número de componentes ideal pode estar entre 4 e 6.

Figura 4: Scree plot - Percentual da variância explicada por Componente Principal



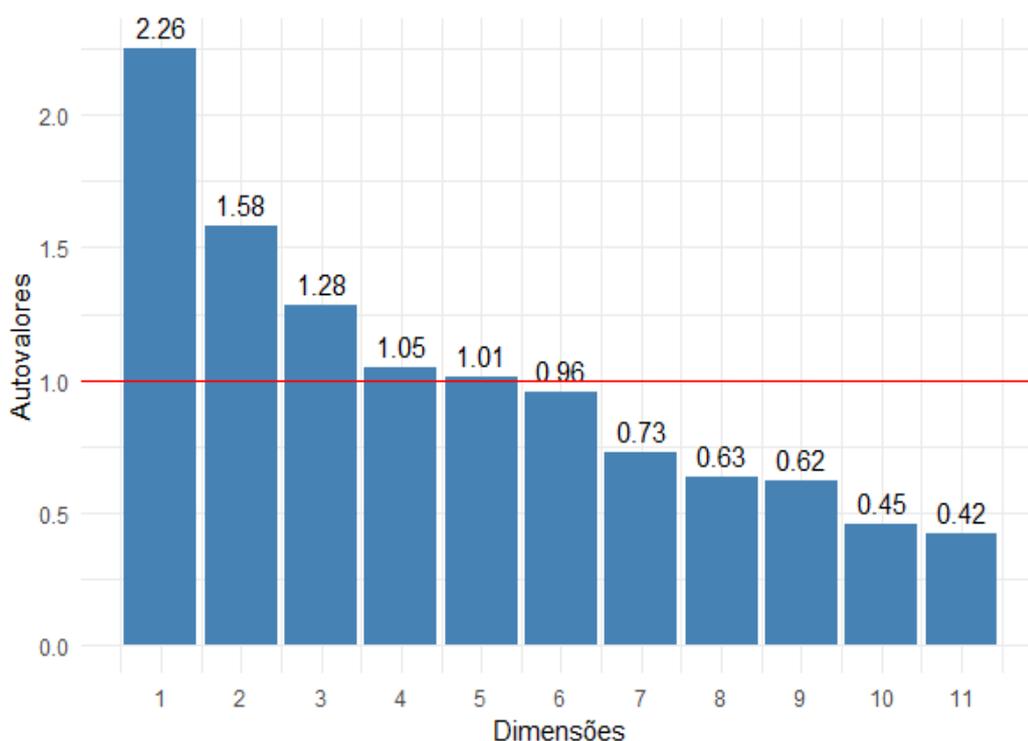
Fonte: Elaborado pelo autor.

Estabelecendo um objetivo de tentar explicar, pelo menos, 70% da variância total com as variáveis mantidas no sistema, o número de componentes ideal seria 6, pois nesse ponto já se observa 74,06% da variância explicada acumulada. O uso de 5 componentes reduziria essa porcentagem para 65,36%, o que não atende exatamente o critério da variância total visto em Regazzi (2000). Ainda assim, esse número pode ser suficiente, considerando a área de estudo.

As Ciências Sociais tendem a tratar de fenômenos em que há baixa correlação entre as variáveis, o que dificulta explicar muito da variância com poucos componentes. Por esse motivo, entende-se que, havendo outras justificativas corroborando essa decisão, 5 componentes explicando 65,36% da variância total podem ser admitidos.

É o que se verifica no Scree plot da Figura 5, o qual agora traz o resultado dos autovalores para cada componente. Somente os 5 Componentes Principais (CP) passam no critério de Kaiser, sugerindo que o total de 6 componentes pode ser desnecessário, até pelo fato do 6^a componente explicar uma porcentagem da variância muito próxima do seu antecessor.

Figura 5 - Scree plot - Seleção das novas variáveis pelo critério de Kaiser



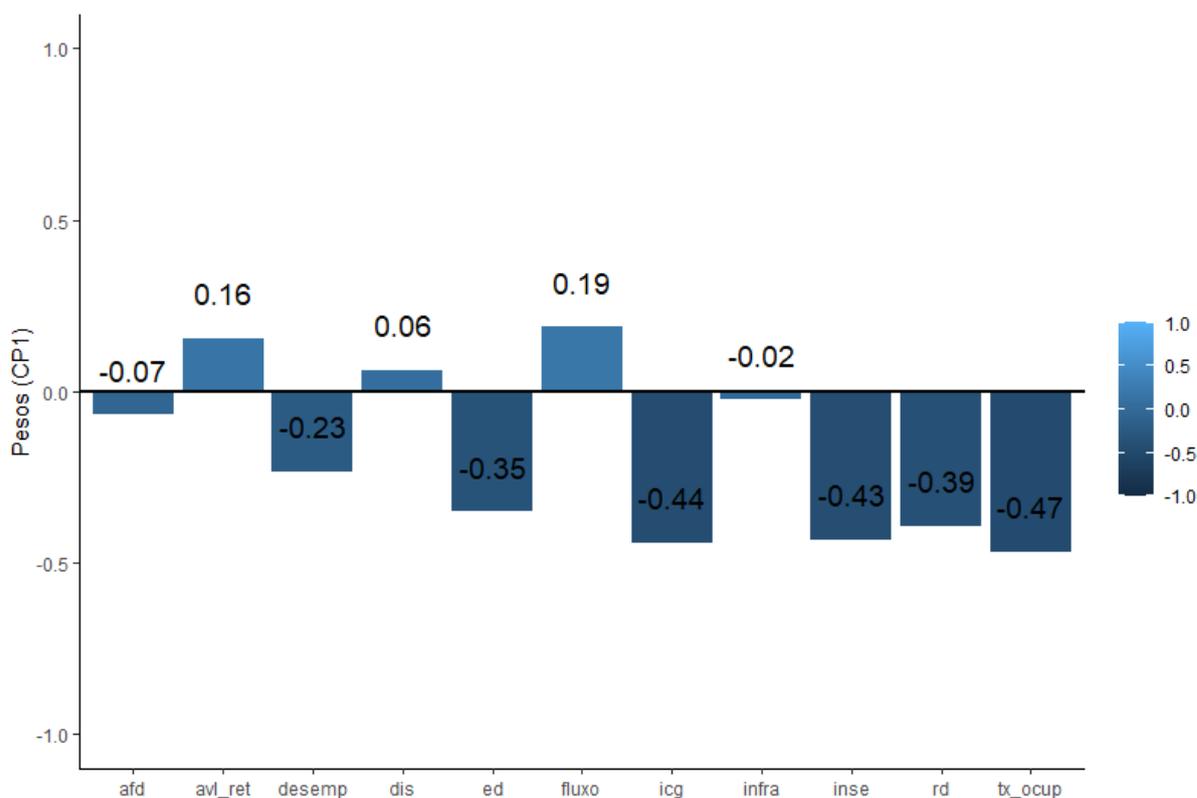
Fonte: Elaborado pelo autor.

O critério final para determinar se somente serão utilizados os 5 componentes principais foi sua capacidade de interpretação a partir dos pesos de cada variável no componente.

4.4. Interpretação dos componentes principais

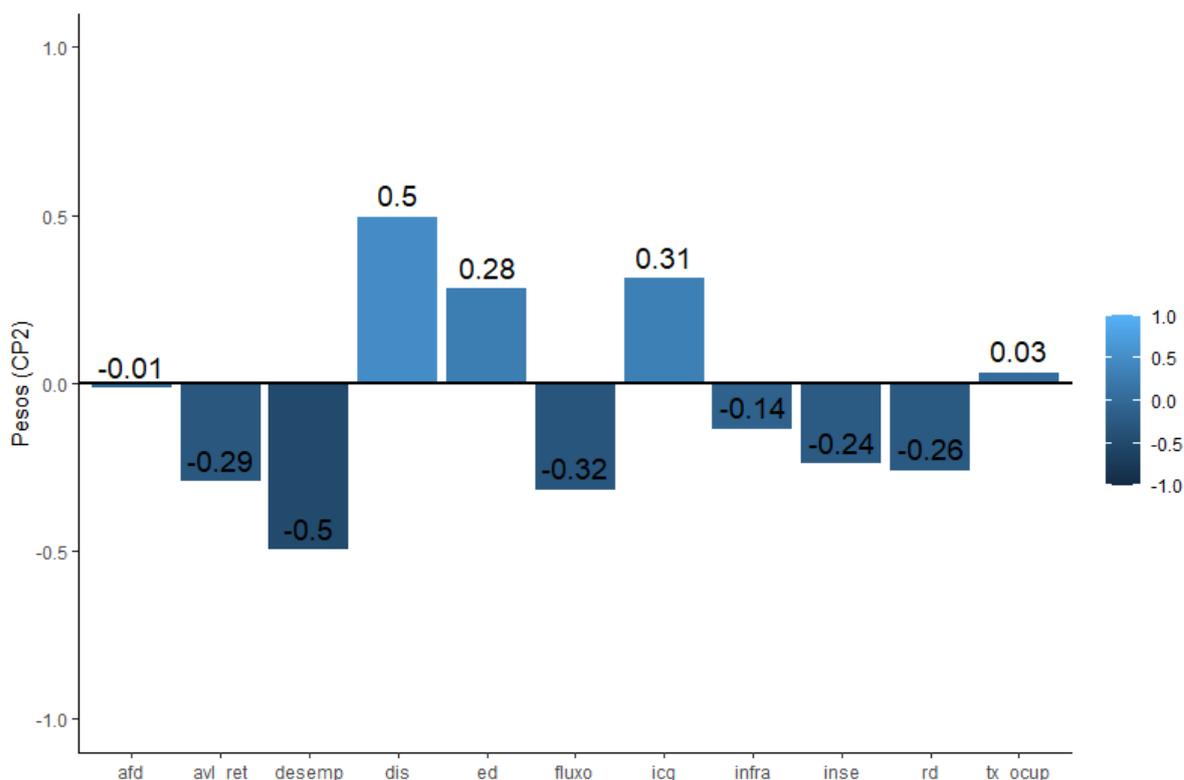
Analisando os pesos do CP1 (Figura 6), nota-se que as variáveis com maior peso dizem respeito principalmente ao porte da escola e a sua condição socioeconômica. São eles a Complexidade de Gestão da Escola (icg), a Taxa Média de Ocupação das Salas de Aula (tx_ocup) e o próprio Nível Socioeconômico (inse).

Figura 6: Pesos CP1 (20,5% da variância explicada)



Fonte: Elaborado pelo autor.

Logo, devido ao sinal das variáveis no componente, escolas com escores altos no CP1 indicam escolas pequenas e pobres, enquanto escores baixos representam escolas de grande porte e mais ricas. Por sua vez, o CP2 (Figura 7) agrega em maior medida variáveis relativas aos resultados educacionais da escola.

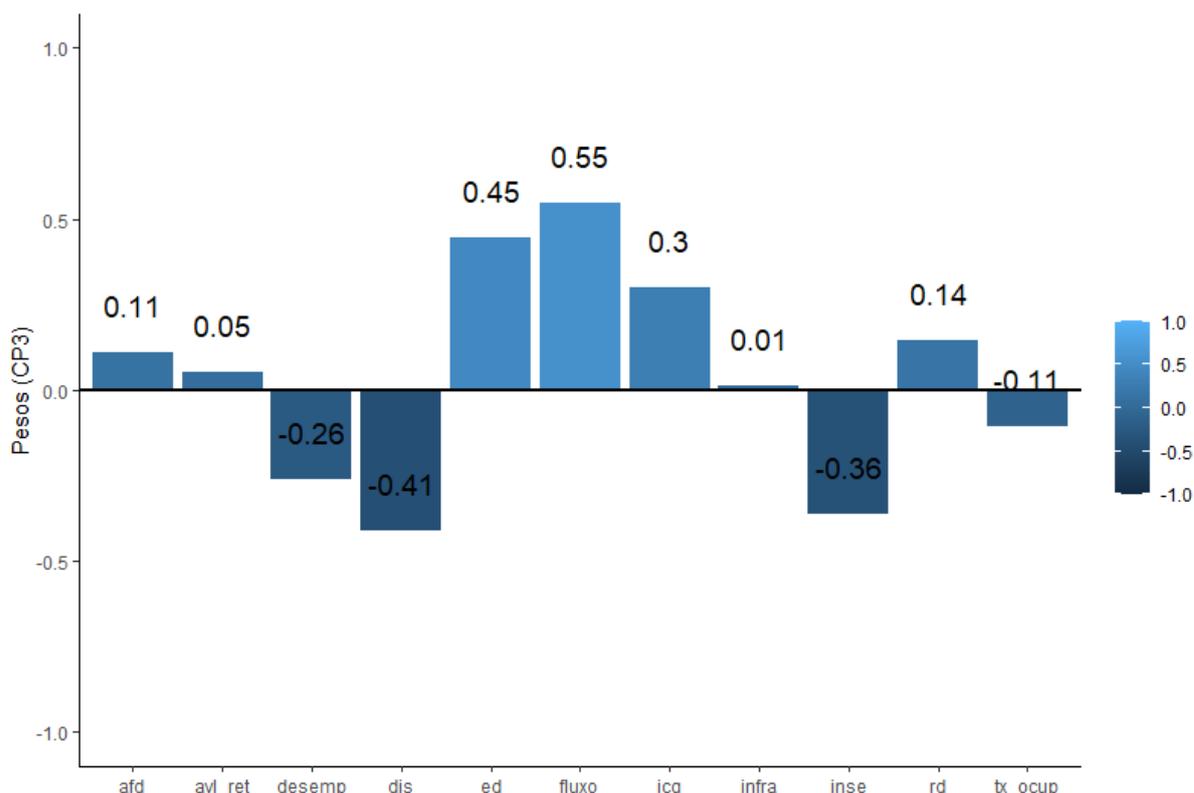
Figura 7: Pesos CP2 (14,4% da variância explicada)

Fonte: Elaborado pelo autor.

Destacam-se variáveis como o Indicador de Desempenho (desemp), o Indicador de Fluxo (fluxo) e a Distorção Idade-Série (dis), a qual, como se espera, tem sinal oposto ao Fluxo. Em segundo plano, estão algumas variáveis que contextualizariam esse resultado, como o Esforço Docente (ed), que parece se associar com a Distorção Idade-Série (dis), e a Regularidade Docente (rd) e o Nível Socioeconômico (inse), que como já foi dito, se relacionam mais com o desempenho.

Portanto, esse componente poderia ser interpretado como uma síntese dos indicadores de resultado educacionais de uma escola. Devido aos sinais das variáveis, escolas com escores maiores tem resultados piores e escolas com escores menores tem resultados melhores.

O terceiro componente (Figura 8) também apresenta um peso forte dos indicadores de resultado, mas com um comportamento distinto do CP2.

Figura 8: Pesos CP3 (11,7% da variância explicada)

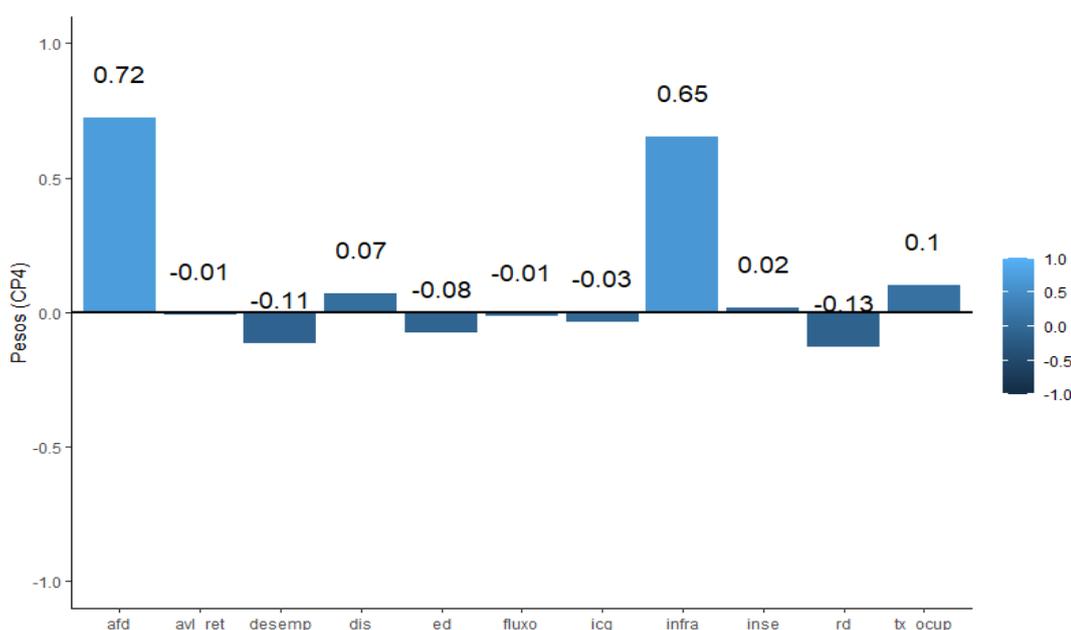
Fonte: Elaborado pelo autor.

Primeiramente, o Indicador de Fluxo (fluxo) tem maior proeminência e, diferente do caso anterior, com sinal oposto ao do Indicador de Desempenho. As demais variáveis com relevância nesse componente estão fortemente associadas com o Fluxo, tais como a Distorção Idade-Série (dis), também com sinal contrário, o Esforço Docente (ed), o Nível Socioeconômico (inse) e a Complexidade de Gestão da Escola (icg). Esses dois também estão correlacionados com o desempenho e o Esforço Docente, respectivamente, como já foi discutido. Pode-se interpretar dessa configuração que o CP3 não é uma medida síntese dos resultados educacionais da escola como é o CP2, mas sim uma medida do possível descompasso entre o desempenho e o fluxo de uma escola. Como foi levantado no Capítulo 2, com o intuito de melhorar superficialmente os resultados, uma escola poderia focar no seu Fluxo por meio de uma política de não-retenção, descuidando do aprendizado dos estudantes. O CP3 parece captar mais o Fluxo e ser capaz de categorizar casos em que esse tipo de desajuste possa ser significativo.

Logo, interpreta-se que escolas com escores maiores nesse componente possuem um fluxo melhor, porém com desempenho ruim, e escolas com escores menores possuem um fluxo pior, porém com desempenho satisfatório.

No CP4 (Figura 9), observa-se a proeminência de praticamente só duas variáveis: a Adequação da Formação Docente (afd) e o Diagnóstico de Infraestrutura (infra).

Figura 9: Pesos CP4 (9,5% da variância explicada)

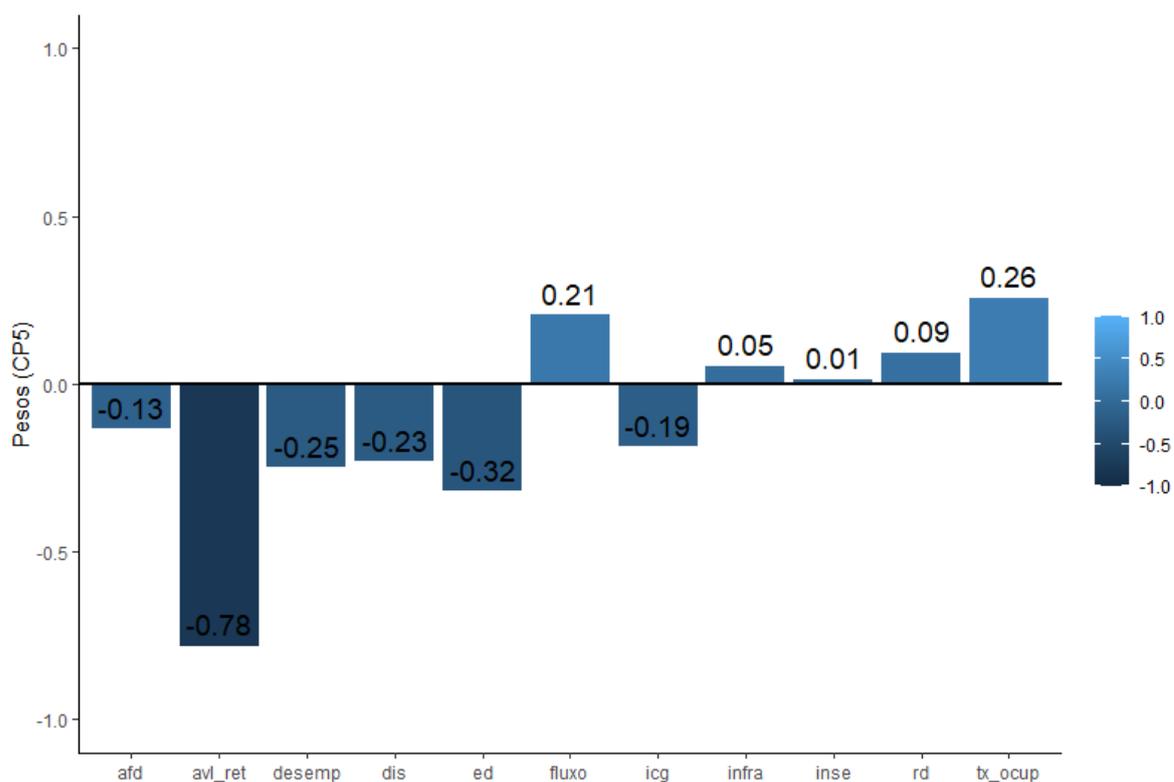


Fonte: Elaborado pelo autor.

Chama atenção essas variáveis não terem tido um peso maior em outros componentes, especialmente aquelas que lidavam com os resultados educacionais, e pelo fato de aparecerem juntas e quase sozinhas com a mesma polaridade em um componente apenas. Pode-se interpretar esse comportamento como um sinal da capacidade de os docentes mais bem formados conseguirem selecionar justamente as escolas com melhores condições de infraestrutura, independente de outros fatores que poderiam conduzir uma política de alocação e movimentação desses professores: como a necessidade de profissionais mais qualificados em escolas de menor nível socioeconômico ou com resultados piores. Portanto, conclui-se que essa variável traduz a qualidade da infraestrutura e da formação dos profissionais de determinada escola.

Por fim, o CP5 (Figura 10) apresenta um peso maior do Indicador de avaliação do retorno às aulas presenciais (avl_ret), sendo interessante ter esse componente no sistema justamente para ajudar a interpretar essa variável que não apareceu com muito peso nos demais componentes.

Figura 10: Pesos CP5 (9,2% da variância explicada)



Fonte: Elaborado pelo autor.

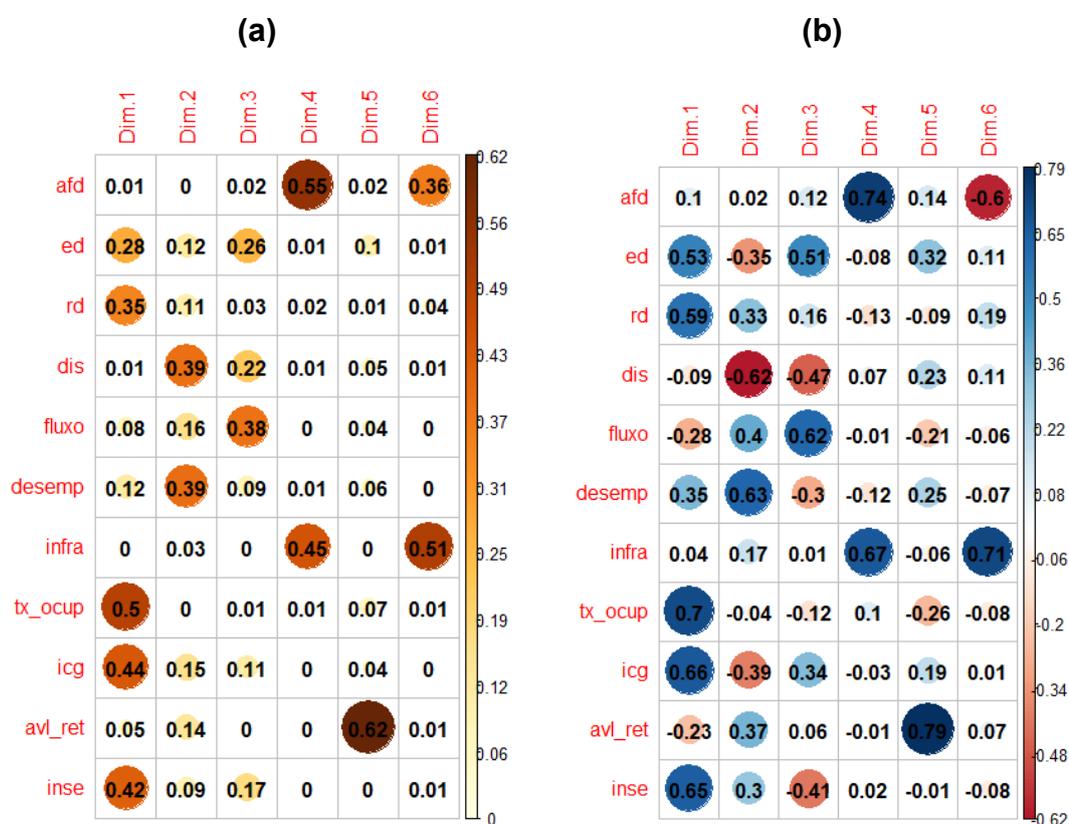
Percebe-se que a avaliação que o aluno faz do retorno às aulas presenciais está variando junto a fatores relacionados à própria organização desse retorno, como a taxa de ocupação das salas de aula (tx_ocup) e ao Esforço Docente (ed), assim como a indicadores dos resultados educacionais, como o desempenho (desemp), fluxo (fluxo) e distorção idade-série (dis).

Logo, esse componente poderia ser interpretado como primeiramente um reflexo da avaliação do aluno à organização da escola e da comunidade escolar ao retorno às aulas presenciais, por meio também da ocupação das salas de aula e do Esforço Docente, e o reflexo potencial disso no desempenho dos estudantes. Devido ao sinal das variáveis, escores menores representam resultados melhores na avaliação dos alunos.

4.5. Análise do ajuste das variáveis aos componentes principais

Na Figura 11 são apresentadas duas matrizes para complementar a interpretação e seleção das novas variáveis: a primeira traz o valor de \cos^2 , uma medida do ajuste das variáveis pelos componentes principais, e a segunda é a matriz de correlação das variáveis com os componentes.

Figura 11: Qualidade do ajustamento das variáveis (a) e matriz de correlação entre variáveis e componentes (b)



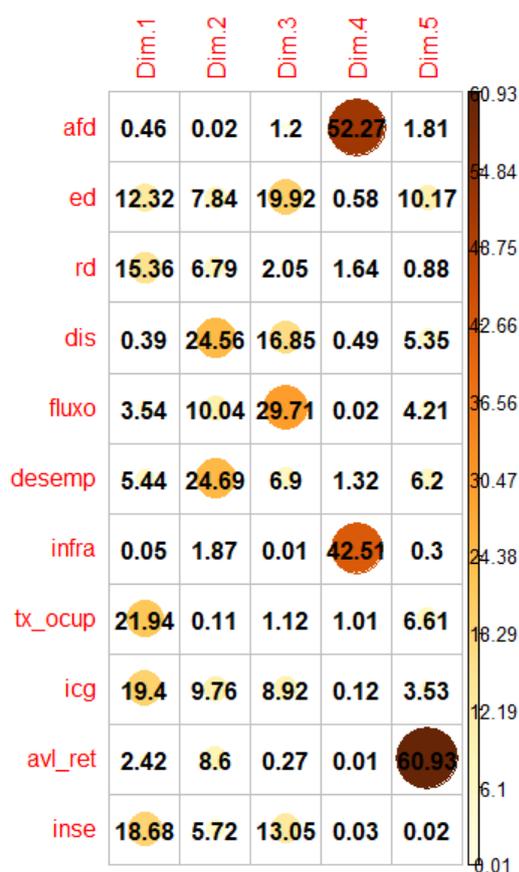
Fonte: Elaborado pelo autor.

Pelo nível de ajuste, percebe-se que todas as variáveis foram bem ajustadas em pelo menos um dos componentes. A matriz de correlação traz um quadro parecido, pelo qual é possível concluir de forma mais definitiva que o CP6 agrega pouco ao sistema. Além de não passar no critério de Kaiser e da análise por meio do Scree Plot, percebe-se que sua interpretação seria muito semelhante ao que já é explicado pelo CP5, no que tange à Adequação da Formação Docente (afd) e ao Diagnóstico de Infraestrutura (infra) da escola. Definiu-se assim que serão utilizados

os primeiros cinco Componentes Principais neste estudo, sendo aceitável uma variância explicada acumulada de 65,4%, que se aproxima de 70%, e condiz com as variâncias típicas observadas nesta área de estudo.

A matriz do percentual de contribuição de cada variável para os Componentes Principais é mostrada na Figura 12.

Figura 12: Matriz de contribuição percentual das variáveis para os CPs

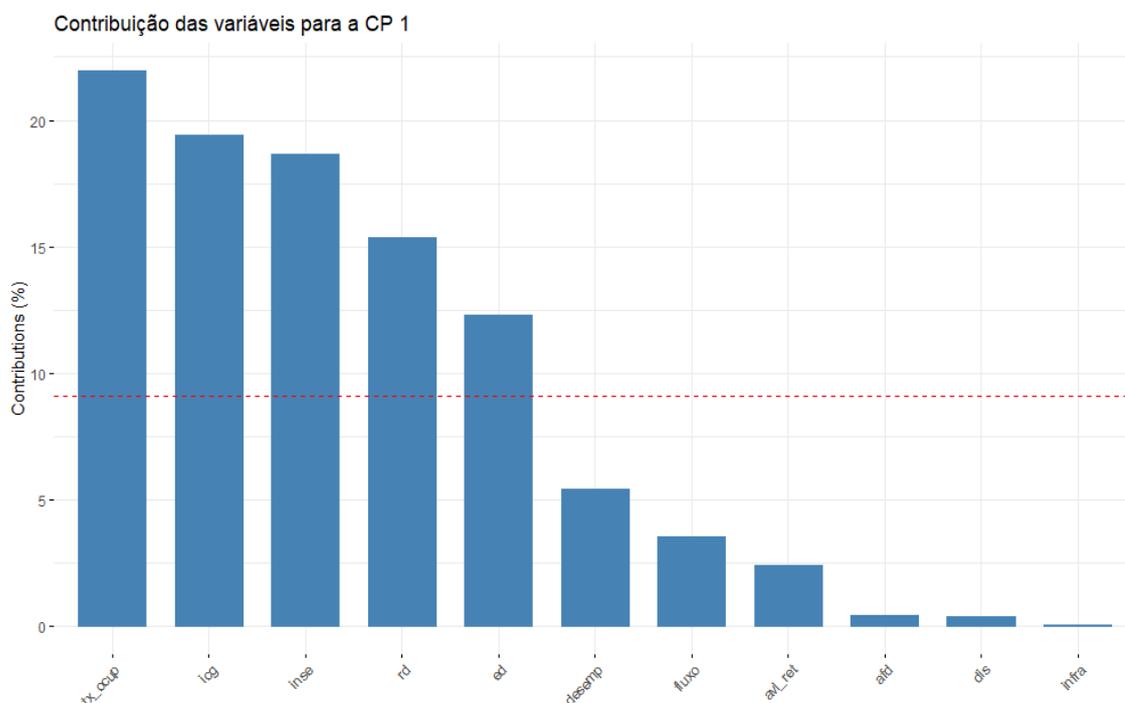


Fonte: Elaborado pelo autor.

Nessa visualização, foi apresentada tanto uma matriz das contribuições percentuais, como também um gráfico por CP com o ordenamento por tamanho da contribuição. Nesses gráficos, as linhas vermelhas representam a contribuição média esperada por cada variável, caso essa distribuição fosse uniforme, ou seja, se cada variável contribuísse na mesma proporção para o CP. Neste caso, esse valor médio é aproximadamente 9%, de forma que as variáveis que superam esse limite contribuem para o componente acima do esperado numa distribuição normal, sendo mais importantes na interpretação dos escores calculados a partir dos autovetores.

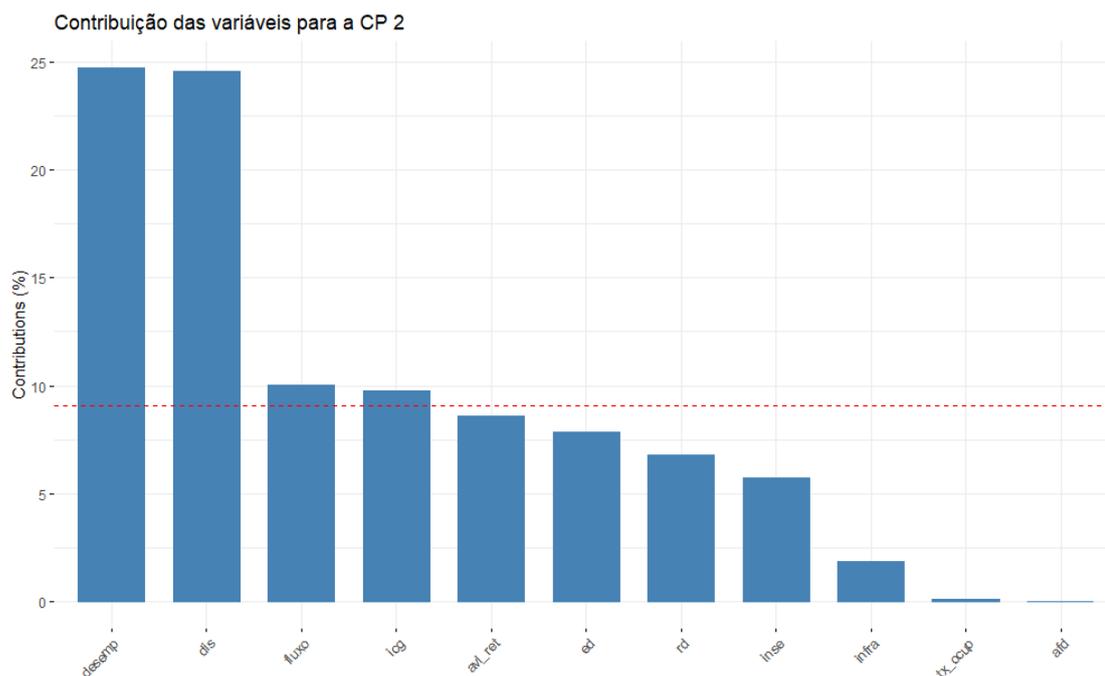
A visualização dos pesos em ordem decrescente separadamente por Componente Principal facilita a interpretação (Figuras 13 a 17).

Figura 13: Ordenamento das variáveis por percentual de contribuição no CP1

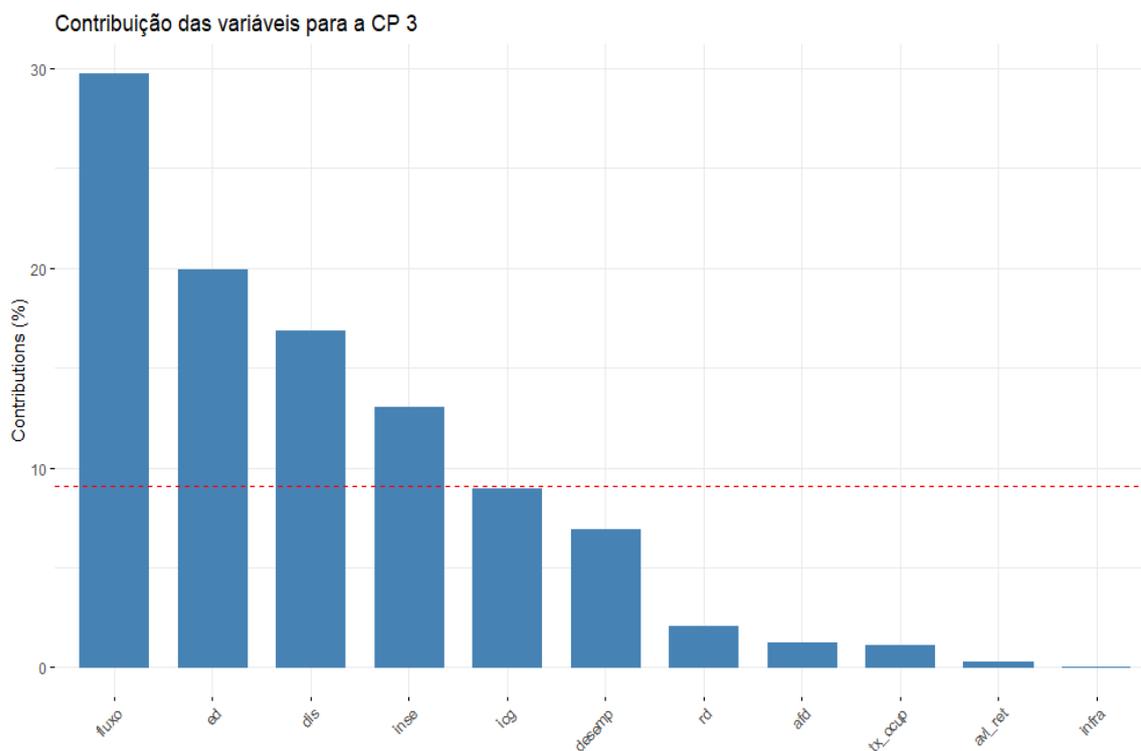


Fonte: Elaborado pelo autor.

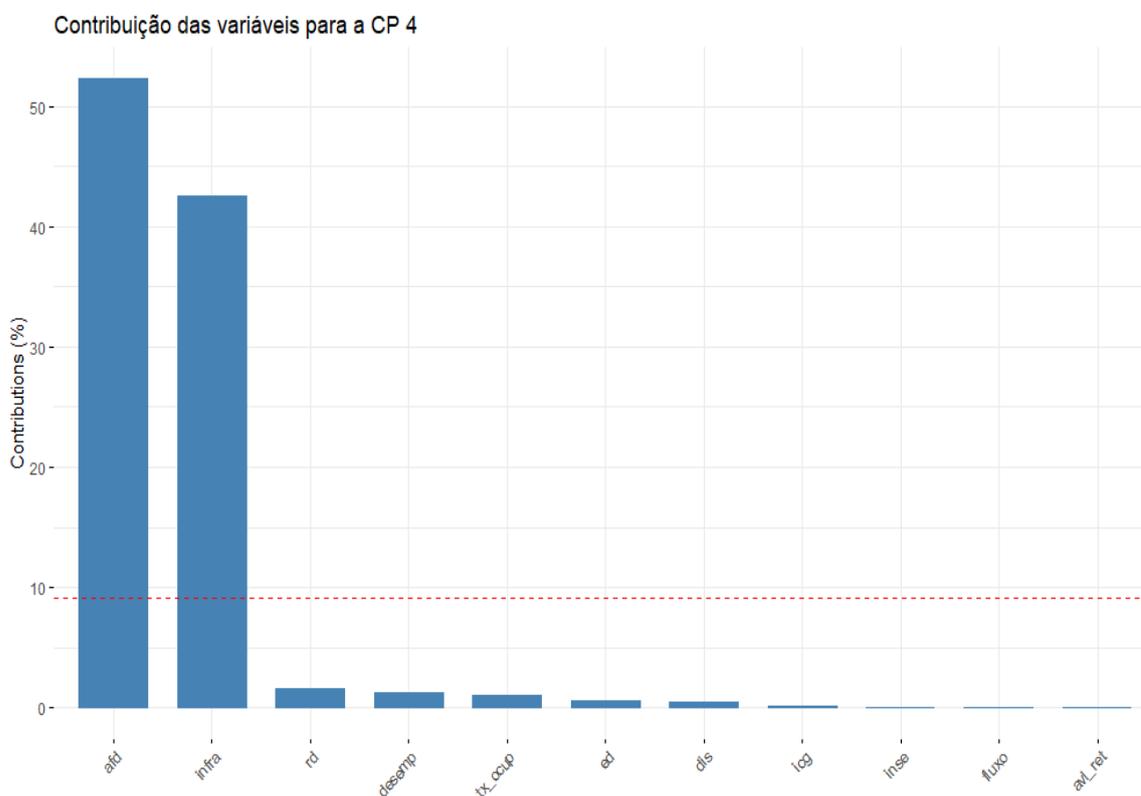
Figura 14: Ordenamento das variáveis por percentual de contribuição no CP2



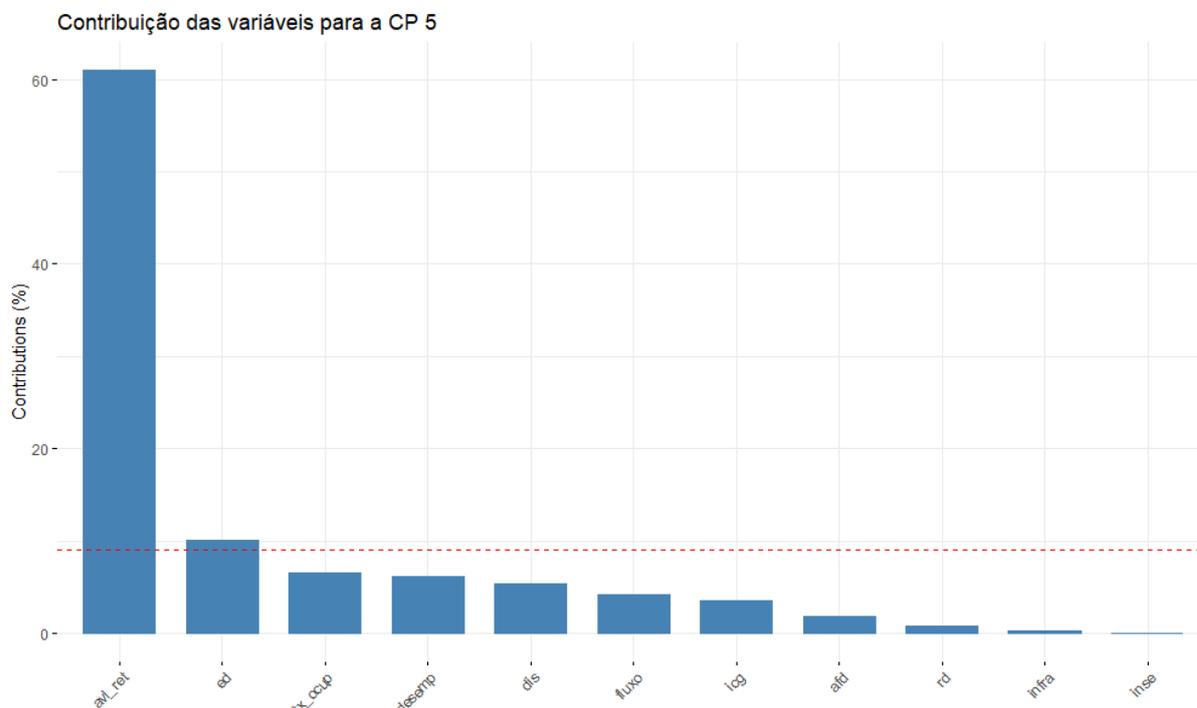
Fonte: Elaborado pelo autor.

Figura 15: Ordenamento das variáveis por percentual de contribuição no CP3

Fonte: Elaborado pelo autor.

Figura 16: Ordenamento das variáveis por percentual de contribuição no CP4

Fonte: Elaborado pelo autor.

Figura 17: Ordenamento das variáveis por percentual de contribuição no CP5

Fonte: Elaborado pelo autor.

Tendo em vista os pesos de cada variável e as análises feitas até o momento, é possível resumir a interpretação dos componentes principais conforme o exposto no Quadro 4.

Quadro 4: Resumo da interpretação dos Componentes Principais selecionados

CP	Nome	Interpretação
1	Contexto socio-administrativo	Escolas com escores baixos nesse componente indicam escolas de grande porte e mais ricas e escores altos representam escolas pequenas e mais pobres
2	Síntese dos resultados educacionais	Síntese dos indicadores de resultado educacionais de uma escola, sendo escores menores representativos de resultados melhores e escores maiores representativos de resultados piores.
3	Desajuste entre desempenho-fluxo escolar	Medida de desajuste entre o desempenho e o fluxo de uma escola, sendo escores maiores indicativos de um fluxo melhor, porém com desempenho ruim, e escores menores indicativos de um fluxo pior, porém com desempenho satisfatório.
4	Infraestrutura-Formação docente	Escolas com escores altos nesse componente possuem condições melhores de infraestrutura e de formação docente, enquanto escolas com escores baixo apresentam condições piores nesses dois quesitos.
5	Retorno às aulas presenciais	Avaliação dos alunos acerca do retorno às aulas presenciais, sendo escores menores indicativos de uma avaliação positiva e escores maiores indicativos de uma avaliação negativa.

Fonte: Elaborado pelo autor.

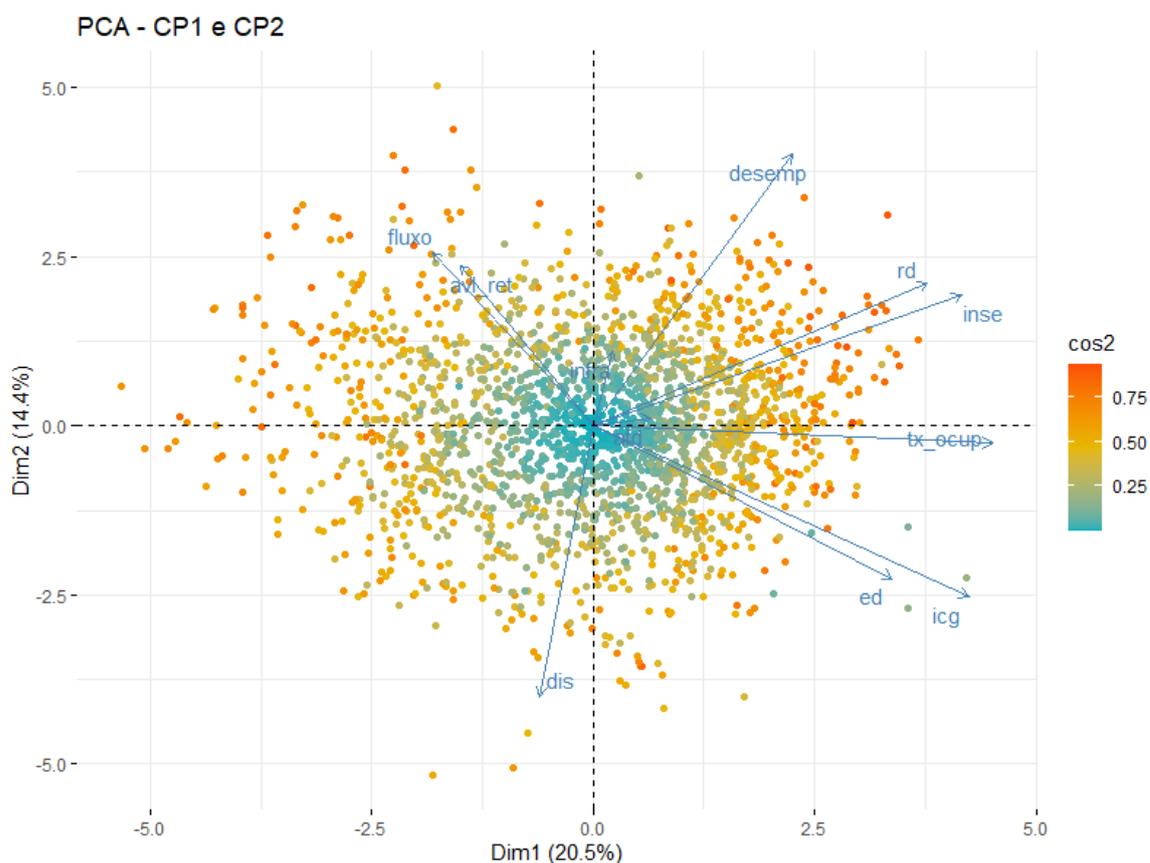
4.6. Análise dos grupos de escolas caracterizados pelos componentes principais

Nesta seção, são apresentadas cinco análises cruzando os valores dos componentes principais selecionados e identificando os grupos de variáveis formados. A partir dos cruzamentos de Componentes Principais, serão apresentados os resultados com os valores desses componentes para as escolas.

Análise 1: Porte e nível socioeconômico da escola (CP1) por resultado educacional (CP2)

A Figura 18 apresenta o diagrama dos escores de cada escola no CP1 (eixo horizontal) contra seus escores no CP2 (eixo vertical). A cor de cada ponto correspondente a uma escola indica a qualidade do seu ajustamento aos componentes (\cos^2).

Figura 18: Distribuição dos indivíduos na análise CP1 x CP2



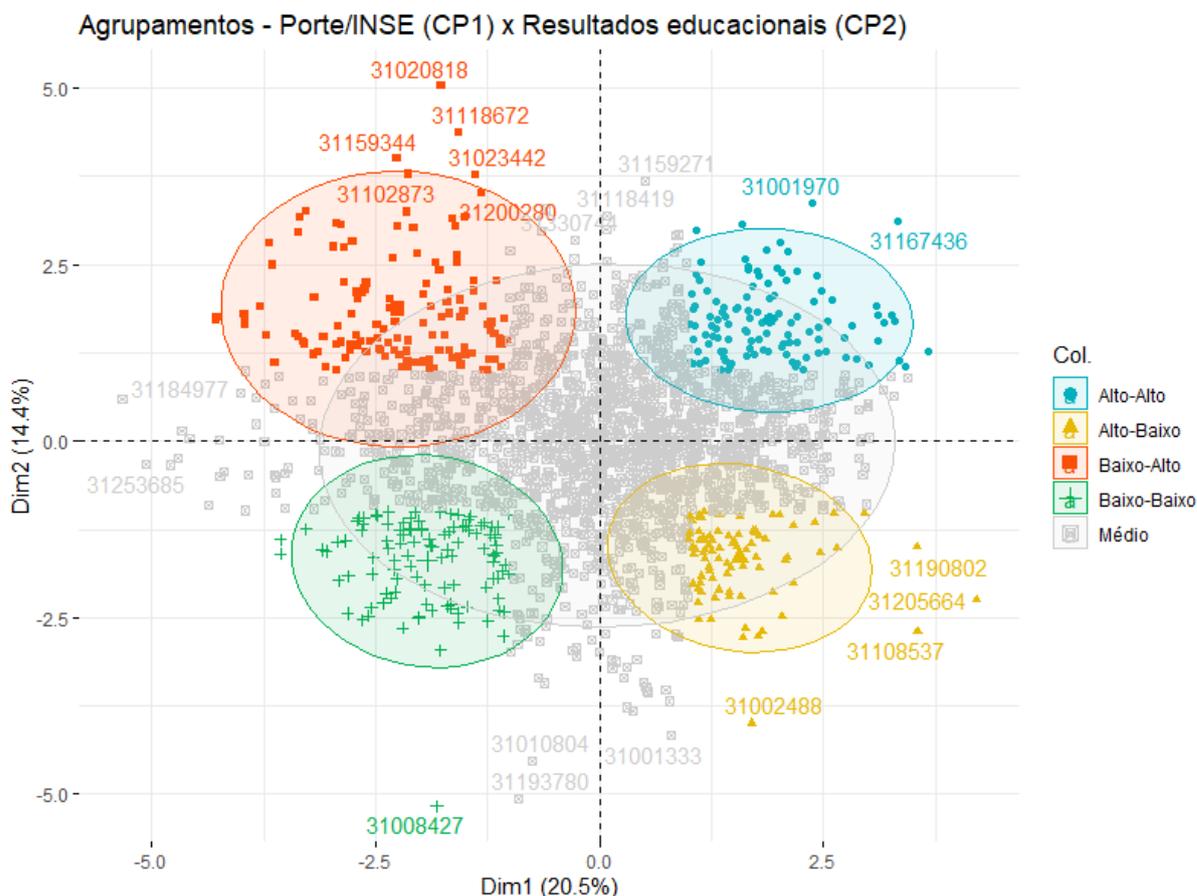
Fonte: Elaborado pelo autor.

As escolas de porte menor (icg menor) e mais pobres (inse menor) estão à esquerda do gráfico e, à direita, estão as escolas maiores e mais ricas. Por sua vez, quanto mais acima no eixo vertical a escola estiver, melhores tendem a ser seus resultados educacionais, a saber: um desempenho (desemp) maior, um fluxo maior e uma distorção idade-série (dis) menor.

A partir dessa configuração, foram formados os grupos de escolas que possuísssem um escore padronizado no CP igual ou maior que 1 (polaridade positiva) ou igual ou menor que -1 (polaridade negativa), ou seja, que variam um desvio-padrão ou mais da média observada para os escores padronizados daquele CP.

Formam-se, assim, até quatro grupos de escolas (Figura 19) que se destacam nos dois componentes (Alto-Alto ou Baixo-Baixo) ou somente em um dos componentes (Alto-Baixo ou Baixo-Alto). O primeiro adjetivo do nome do grupo sempre se referirá ao componente no eixo horizontal e o segundo, ao componente no eixo vertical.

Figura 19: Agrupamentos na análise CP1 x CP2



Fonte: Elaborado pelo autor.

A Tabela 2 mostra as estatísticas dos indicadores para os quatro agrupamentos, estando destacadas somente aquelas variáveis com percentual de contribuição para o CP maior do que aquela contribuição mínima esperada em uma distribuição uniforme.

Tabela 2: Resumo descritivo dos agrupamentos CP1-CP2

Variável		Contribuição CP1	Contribuição CP2	Média amostral	Grupo (CP1-CP2)			
					Alto-Alto	Baixo-Alto	Alto-Baixo	Baixo-Baixo
Taxa média de ocupação das salas de aula	tx_ocup	21,94	0,11	0,57	0,70	0,37	0,66	0,44
Complexidade de gestão da escola	icg	19,40	9,76	4,50	4,72	2,95	5,50	4,10
Nível socioeconômico	inse	18,68	5,72	5,83	7,03	5,48	6,19	4,77
Regularidade do corpo docente	rd	15,36	6,79	2,97	3,52	2,74	3,07	2,46
Esforço docente	ed	12,32	7,84	3,93	3,95	3,40	4,24	3,88
Indicador de desempenho	desemp	5,44	24,69	4,11	4,68	4,41	4,01	3,45
Taxa de distorção idade-série	dis	0,39	24,56	0,16	0,09	0,10	0,23	0,24
Indicador de fluxo	fluxo	3,54	10,04	91,71	93,83	97,26	77,35	90,37
Avaliação do retorno às aulas presenciais	avl_ret	2,42	8,60	2,79	2,81	2,95	2,70	2,77
Adequação da formação docente	afd	0,46	0,02	71,50	73,44	69,47	70,13	69,46
Diagnóstico de infraestrutura	infra	0,05	1,87	0,83	0,86	0,83	0,81	0,81
Número de escolas no grupo				2131	114	128	88	108

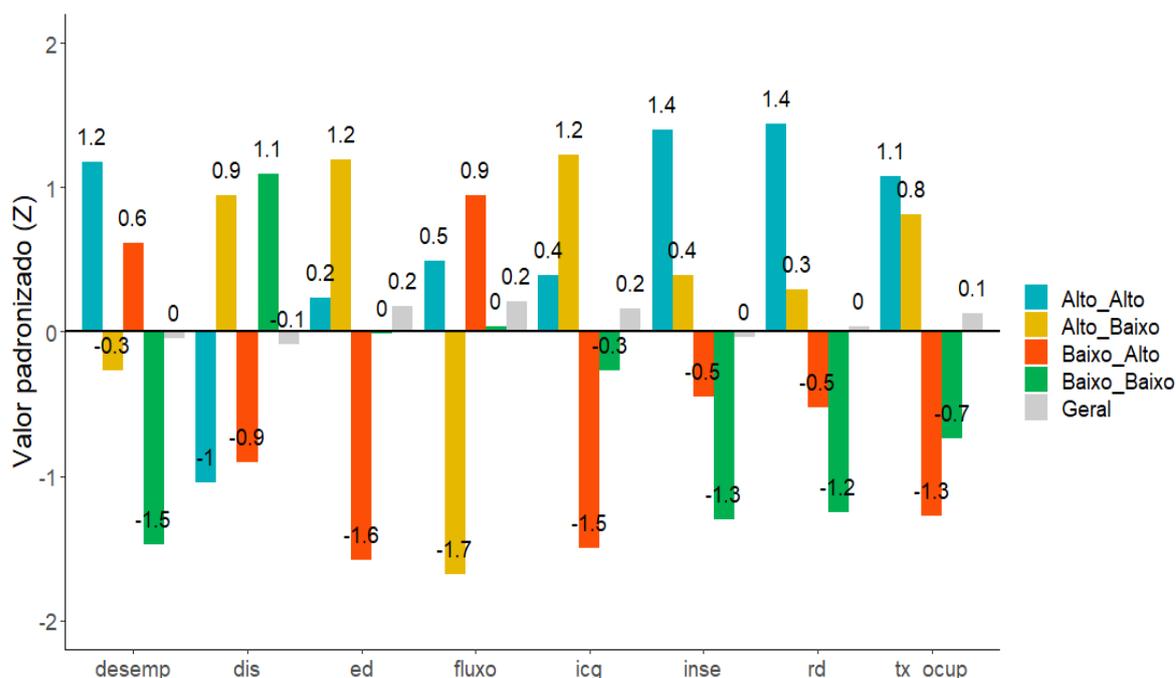
Fonte: Elaborado pelo autor.

Percebe-se que foi possível identificar escolas de um “porte” muito semelhante, grandes e ricas ou pequenas e pobres, com resultados totalmente distintos.

O grupo Baixo-Alto representa aquelas escolas com um tamanho menor e nível socioeconômico geralmente inferior à média, mas com resultados comparados às maiores escolas. Por outro lado, o grupo Alto-Baixo traz escolas que possuem resultado aquém do esperado.

A Figura 20 apresenta um resumo dos valores padronizados para as variáveis que mais contribuem para os componentes analisados. Nela fica ainda mais evidente a distinção entre os grupos.

Figura 20: Comparação das escolas por agrupamento CP1-CP2



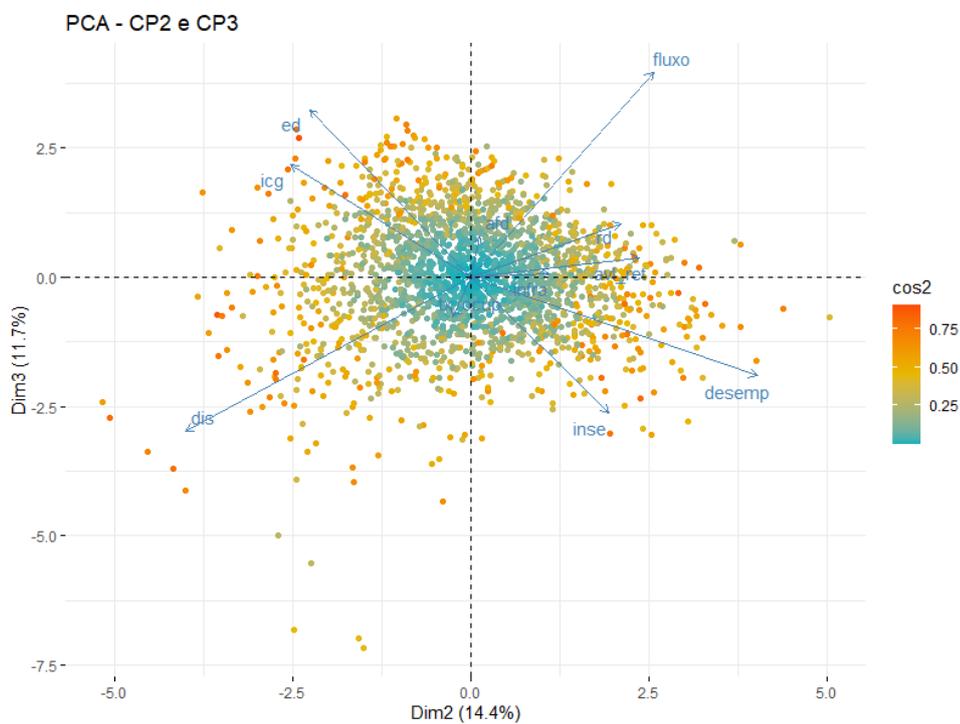
Fonte: Elaborado pelo autor.

Análise 2: Ajuste ou desajuste entre o desempenho (CP2) e o fluxo escolar (CP3)

Nos gráficos das figuras 21 e 22, o CP2, que é uma síntese dos resultados educacionais, porém com um peso maior do desempenho, está disposto no eixo horizontal, de maneira que à esquerda estão escolas com desempenho inferior e à direita, os melhores desempenhos.

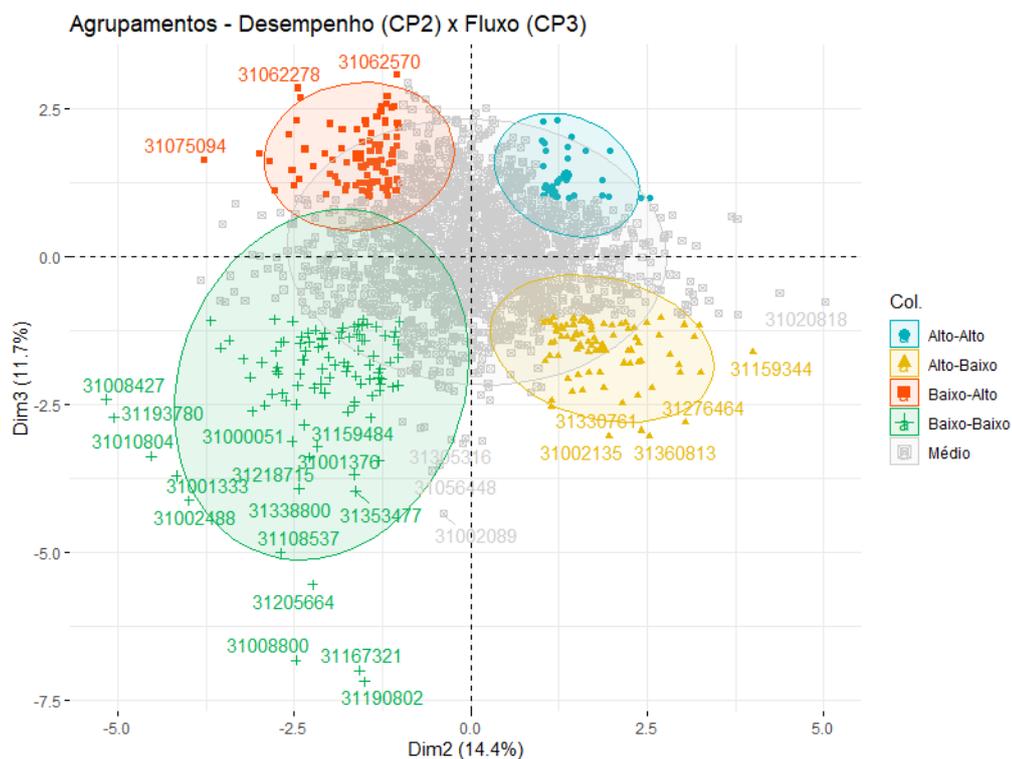
Já no eixo vertical, está disposto o CP3, associado mais fortemente ao fluxo e, na polaridade inversa, à distorção idade-série. Logo, escolas na parte inferior do gráfico tenderão a ter um fluxo pior e, por isso, maior distorção idade-série, e escolas na parte superior tenderão a ter melhor sucesso na aprovação dos alunos e adequação da sua trajetória escolar.

Figura 21: Distribuição dos indivíduos na análise CP2 x CP3



Fonte: Elaborado pelo autor.

Figura 22: Agrupamentos na análise CP2 x CP3



Fonte: Elaborado pelo autor.

Nos agrupamentos da Figura 22, chamam atenção os grupos que apresentam um desajuste entre os dois CPs. A saber, o grupo Alto-Baixo, em que as escolas estão conseguindo um desempenho satisfatório, porém provavelmente à custa de uma retenção maior dos estudantes, o que pode resultar também no aumento da evasão escolar, e o grupo Baixo-Alto, no qual as escolas estão conseguindo manter um fluxo elevado, por meio de taxas de aprovação maiores, porém sem garantir o mesmo nível de aprendizagem que se observa em média para as demais escolas.

Outro grupo que pode ser objeto de interesse é o Baixo-Baixo. Pela Figura 22, é possível constatar que este grupo possui maior dispersão dos indivíduos e valores mais extremos. Esse agrupamento traz aquelas escolas que não estão conseguindo sucesso nem na aprendizagem (desempenho) dos estudantes, nem na manutenção do fluxo, fazendo com que esses alunos abandonem a escola ou não concluam sua vida escolar dentro do tempo ideal. Essa categoria com 103 escolas, conforme apresenta a Tabela 3, poderia ser objeto de uma política de priorização, por meio da qual essas escolas seriam acompanhadas mais de perto pelos Órgãos Públicos e pelas avaliações de larga escala.

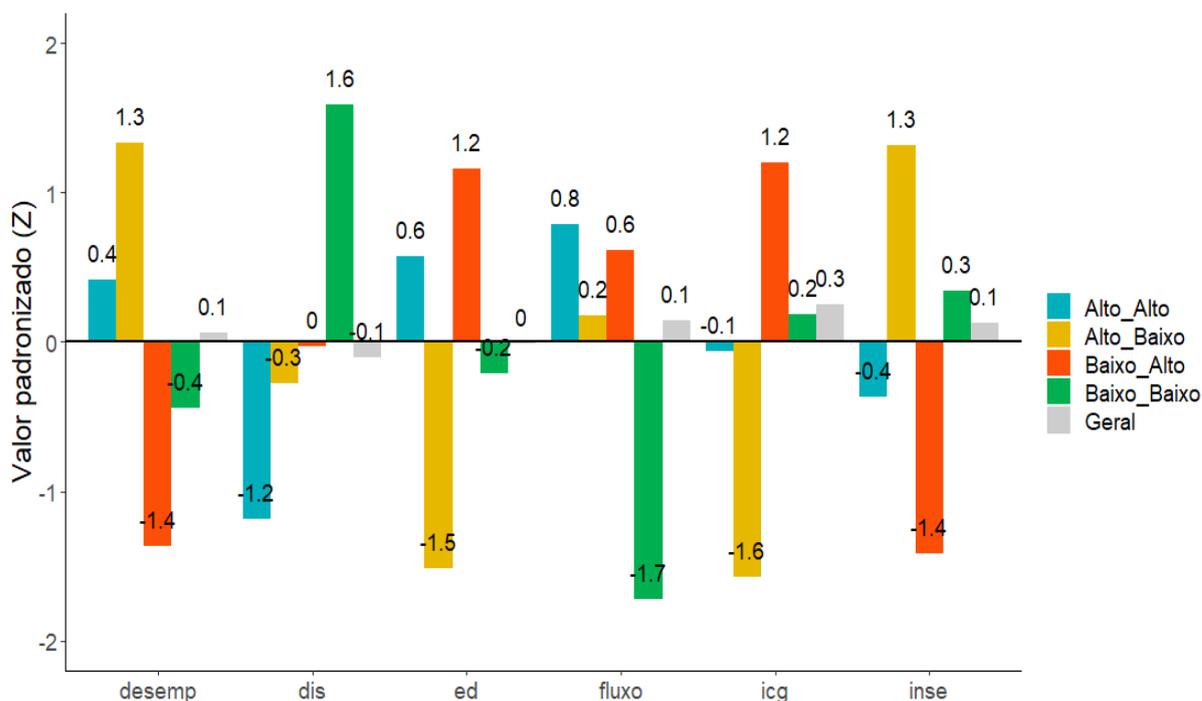
Tabela 3: Resumo descritivo dos agrupamentos CP2-CP3

Variável	Contribuição CP2	Contribuição CP3	Média amostral	Grupo (CP2-CP3)			
				Alto-Alto	Baixo-Alto	Alto-Baixo	Baixo-Baixo
Indicador de desempenho desemp	24,69	6,90	4,11	4,26	3,46	4,68	3,88
Taxa de distorção idade-série dis	24,56	16,85	0,16	0,06	0,16	0,14	0,31
Complexidade de gestão da escola icg	9,76	8,92	4,50	4,27	5,18	3,19	4,45
Indicador de fluxo fluxo	10,04	29,71	91,71	98,29	96,48	92,00	72,50
Esforço docente ed	7,84	19,92	3,93	4,14	4,35	3,40	3,86
Nível socioeconômico inse	5,72	13,05	5,83	5,52	4,86	6,58	5,97
Regularidade do corpo docente rd	6,79	2,05	2,97	3,23	2,77	2,99	2,64
Adequação da formação docente afd	0,02	1,20	71,50	76,19	73,81	68,48	69,29
Taxa média de ocupação das salas de aula tx_ocup	0,11	1,12	0,57	0,47	0,50	0,57	0,57
Avaliação do retorno às aulas presenciais avl_ret	8,60	0,27	2,79	2,94	2,72	2,84	2,72
Diagnóstico de infraestrutura infra	1,87	0,01	0,83	0,87	0,80	0,85	0,81
Número de escolas no grupo			2131	33	85	85	103

Fonte: Elaborado pelo autor.

Logo, verifica-se a utilidade da Análise de Componentes Principais também como uma ferramenta de categorização e até de priorização de indivíduos de um grupo com variação semelhante, como demonstrado também pela Figura 23, que compara os grupos segundo as variáveis mais relevantes nessa análise específica.

Figura 23: Comparação das escolas por agrupamento CP2-CP3



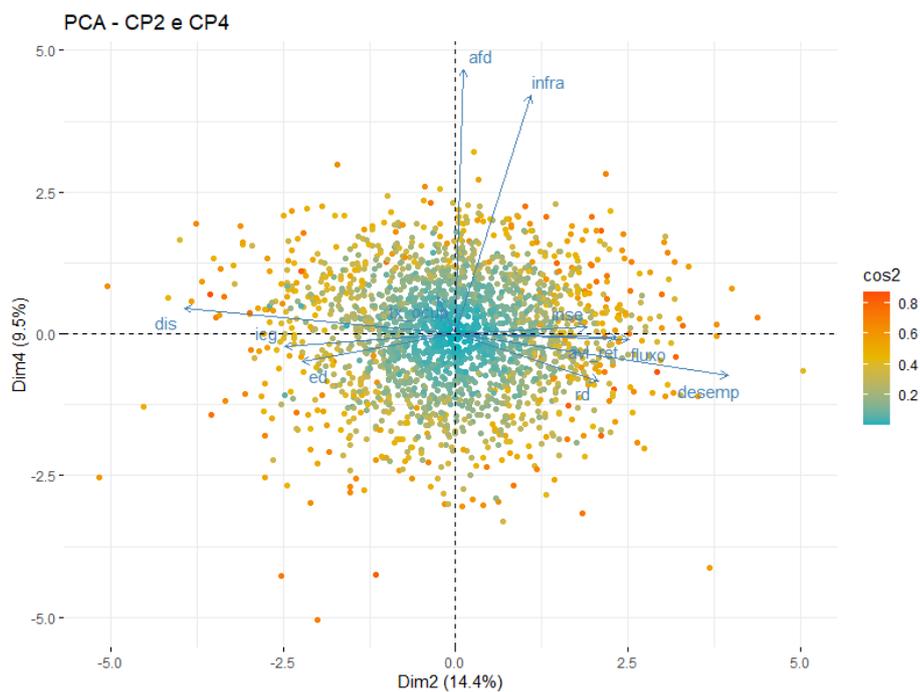
Fonte: Elaborado pelo autor.

Análise 3: Resultados educacionais (CP2) e condições de infraestrutura e corpo docente (CP4)

A Figura 24 traz no eixo horizontal, novamente, o CP2, a fim de sintetizar as variáveis relacionadas aos resultados educacionais, enquanto no eixo vertical está posicionado o CP4, cujas variáveis com maior peso são a Adequação da Formação Docente (afd) e o Diagnóstico de Infraestrutura (infra).

Escolas posicionadas na parte superior do gráfico tendem a ter condições físicas melhores e professores mais bem formados, enquanto as posicionadas na parte inferior terão características opostas. A interpretação do eixo horizontal para o CP2 permanece a mesma da última análise.

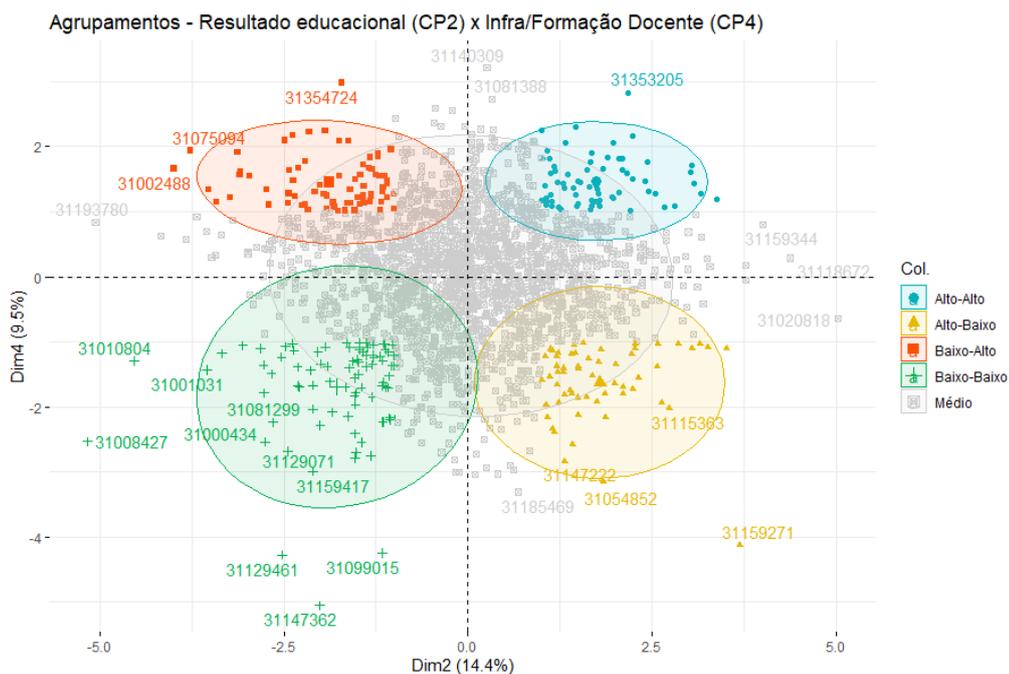
Figura 24: Distribuição dos indivíduos na análise CP2 x CP4



Fonte: Elaborado pelo autor.

Novamente, os agrupamentos ilustrados na Figura 25 permitem comparar escolas com realidades supostamente parecidas, mas com resultados distintos.

Figura 25: Agrupamentos na análise CP2 x CP4



Fonte: Elaborado pelo autor.

Nos grupos Alto-Alto e Baixo-Alto, as escolas possuem infraestrutura e formação docente acima da média geral, porém o último grupo tem um desempenho 20% inferior. Caberia uma exploração, inclusive qualitativa dos grupos, para entender a origem dessa ineficiência. Conforme a discussão levantada na literatura indica, o efeito da escola sobre esse resultado pode não ser tão significativo, uma vez que é possível observar na tabela de resumo que o grupo Baixo-Alto apresenta um Nível Socioeconômico menor do que o grupo Alto-Alto, apesar de terem valores comparáveis de Formação Docente e Infraestrutura.

Analisando o grupo Alto-Baixo, que seriam aquelas escolas com infraestrutura e corpo docente precários, porém bons resultados, percebe-se que a média do Nível Socioeconômico é igual à média do melhor grupo, o Alto-Alto, conforme demonstrado na Tabela 4. Ou seja, talvez esses resultados estariam indicando que o nível socioeconômico dos alunos nessas escolas precárias estaria sendo suficiente para anular o efeito negativo da escola sobre a sua aprendizagem. Ele também explicaria as escolas no Baixo-Alto, onde o contrário ocorreria: o nível socioeconômico dos alunos limitando o efeito positivo que as condições de infraestrutura e a formação docente poderiam ter sobre os resultados educacionais.

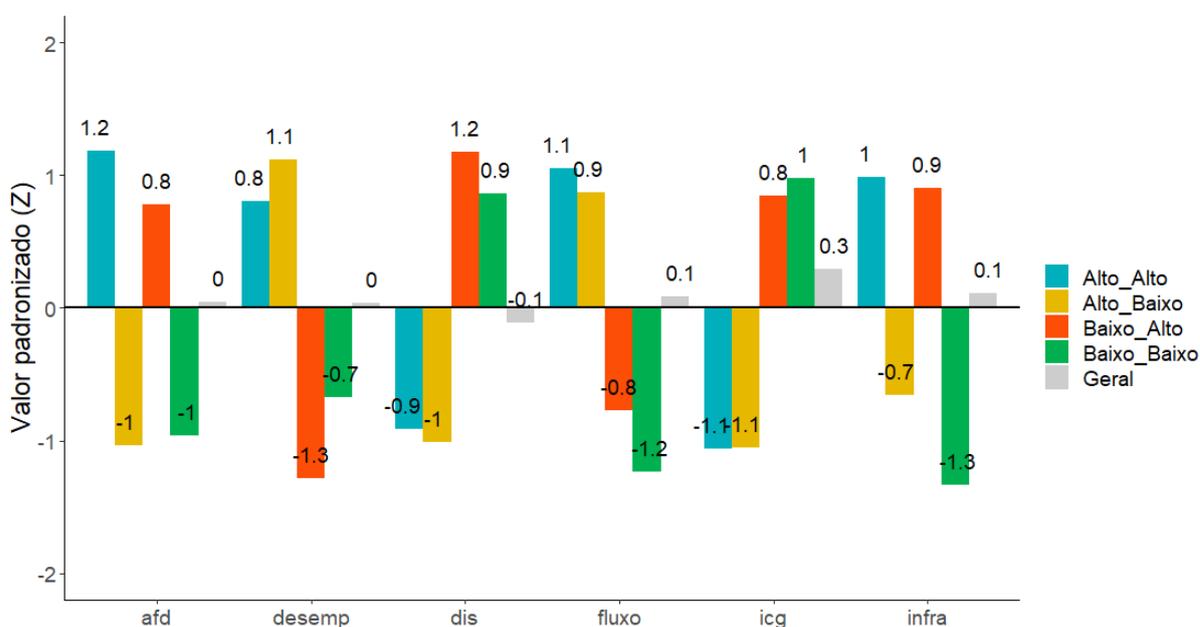
Tabela 4: Resumo descritivo dos agrupamentos CP2-CP4

Variável		Contribuição CP2	Contribuição CP4	Média amostral	Grupo (CP2-CP4)			
					Alto-Alto	Baixo-Alto	Alto-Baixo	Baixo-Baixo
Indicador de desempenho	desemp	24,69	1,32	4,11	4,44	3,53	4,58	3,79
Taxa de distorção idade-série	dis	24,56	0,49	0,16	0,10	0,25	0,09	0,23
Indicador de fluxo	fluxo	10,04	0,02	91,71	96,56	87,45	95,63	85,15
Complexidade de gestão da escola	icg	9,76	0,12	4,50	3,65	4,85	3,66	4,93
Adequação da formação docente	afd	0,02	52,27	71,50	86,41	81,15	57,37	58,37
Diagnóstico de infraestrutura	infra	1,87	42,51	0,83	0,92	0,91	0,75	0,68
Avaliação do retorno às aulas presenciais	avl_ret	8,60	0,01	2,79	2,84	2,72	2,91	2,71
Esforço docente	ed	7,84	0,58	3,93	3,66	4,03	3,65	4,07
Regularidade do corpo docente	rd	6,79	1,64	2,97	3,12	2,50	3,13	2,81
Nível socioeconômico	inse	5,72	0,03	5,83	6,12	5,35	6,12	5,43
Taxa média de ocupação das salas de aula	tx_ocup	0,11	1,01	0,57	0,59	0,54	0,47	0,55
Número de escolas no grupo				2131	69	73	58	86

Fonte: Elaborado pelo autor.

Por fim, novamente, chama-se atenção para o grupo de 86 escolas no Baixo-Baixo, no qual estão algumas das piores infraestruturas, com professores menos qualificados e que, acima de tudo, não apresentam resultados satisfatórios, conforme verifica-se na Figura 26. Tais escolas podem ser objeto de uma política de focalização ou priorização, uma vez que todas também se encontram no pior grupo da análise anterior, focada especificamente nos resultados educacionais.

Figura 26: Comparação das escolas por agrupamento CP2-CP4



Fonte: Elaborado pelo autor.

Igualmente ao que foi observado na última análise, o grupo mais vulnerável é onde se verifica a maior dispersão dos resultados, indicando que existe uma heterogeneidade relevante interna ao próprio grupo Baixo-Baixo.

Análise 4: Relação entre o porte e nível socioeconômico da escola (CP1) e a adequação da comunidade escolar aos protocolos sanitários (COVID) (CP5)

Na análise a seguir, a Figura 27 retoma o CP1 no eixo horizontal, a fim de contextualizar o porte socio-administrativo das escolas, enquanto no eixo vertical é apresentado o CP5, o qual permite encontrar escolas cuja percepção dos alunos

acerca do retorno presencial é positiva ou negativa, um potencial *proxy* da capacidade de gestão escolar.

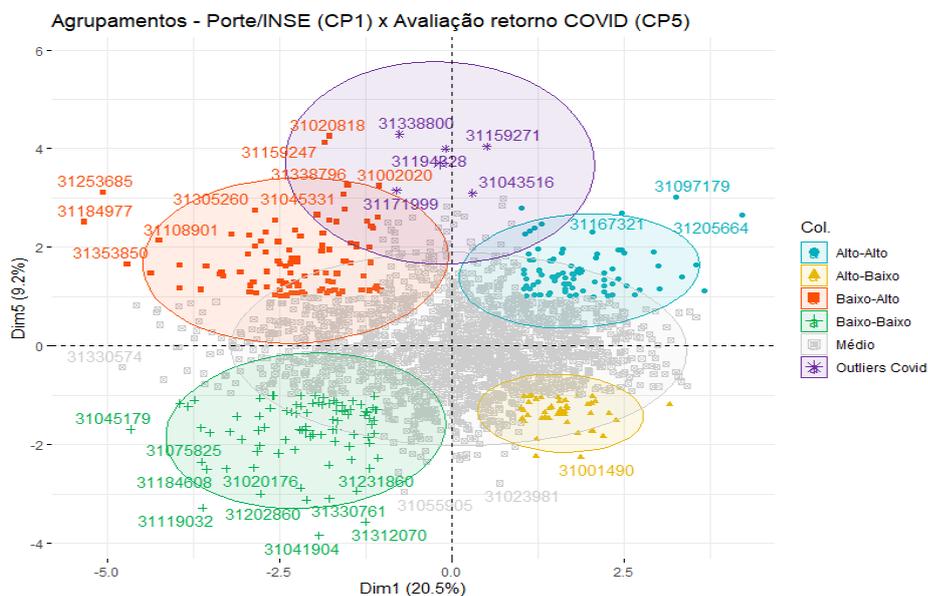
Figura 27: Distribuição dos indivíduos na análise CP1 x CP5



Fonte: Elaborado pelo autor.

Por sua vez, a Figura 28 traz os agrupamentos realizados a partir da combinação dos componentes citados.

Figura 28: Agrupamentos na análise CP1 x CP5



Fonte: Elaborado pelo autor.

Nota-se que, nesta combinação, o menor dos quatro agrupamentos é o Alto-Baixo, indicando que existem menos escolas ricas e de porte grande para o qual a avaliação sobre o retorno presencial dos alunos foi consideravelmente ruim. Logo, conclui-se que a capacidade de gestão escolar, se adequadamente for possível aproximá-la por esse indicador de retorno presencial, é condicionada em certa medida pelo contexto socio-administrativo explicado pelo CP1, o que também poderia explicar a maior dispersão nos agrupamentos cujo valor é menor para esse componente (Baixo-Alto e Baixo-Baixo).

Uma atenção mais detalhada sobre os resultados, apresentados na Tabela 5, pode ser de interesse administrativo para a Secretaria de Estado de Educação, principalmente aqueles de maior amplitude no CP5, já que podem conter a experiência das escolas que melhor e pior conseguiram receber os estudantes após o período prolongado de suspensão da modalidade de ensino presencial.

Tabela 5: Resumo descritivo dos agrupamentos CP1-CP5

Variável	Contribuição CP1	Contribuição CP5	Média amostral	Grupo (CP1-CP5)					
				Alto-Alto	Baixo-Alto	Alto-Baixo	Baixo-Baixo	Outliers CP5	
Taxa média de ocupação das salas de aula	tx_ocup	21,94	6,61	0,57	0,62	0,34	0,73	0,45	0,37
Complexidade de gestão da escola	icg	19,40	3,53	4,50	5,60	3,57	4,60	3,26	4,00
Nível socioeconômico	inse	18,68	0,02	5,83	6,52	5,05	6,76	5,10	5,56
Regularidade do corpo docente	rd	15,36	0,88	2,97	3,23	2,54	3,36	2,69	2,81
Esforço docente	ed	12,32	10,17	3,93	4,29	3,85	3,91	3,43	4,33
Avaliação do retorno às aulas presenciais	avl_ret	2,42	60,93	2,79	2,90	3,06	2,62	2,60	3,24
Indicador de desempenho	desemp	5,44	6,20	4,11	4,49	4,13	4,22	3,63	4,82
Taxa de distorção idade-série	dis	0,39	5,35	0,16	0,19	0,19	0,12	0,16	0,22
Indicador de fluxo	fluxo	3,54	4,21	91,71	81,27	93,94	91,77	95,60	87,81
Adequação da formação docente	afd	0,46	1,81	71,50	75,50	71,65	68,06	67,15	64,58
Diagnóstico de infraestrutura	infra	0,05	0,30	0,83	0,82	0,82	0,82	0,84	0,80
Número de escolas no grupo				2131	80	100	42	86	12

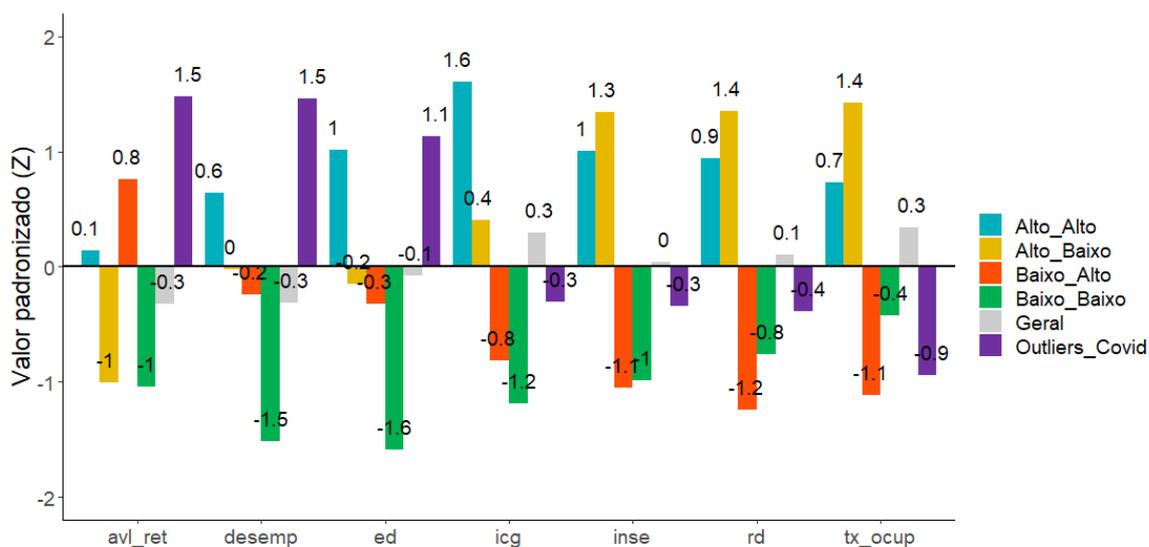
Fonte: Elaborado pelo autor.

O grupo Baixo-Baixo traz uma subcategoria especialmente vulnerável, uma vez que abarca escolas mais pobres e que, apesar de terem um porte menor, tiveram talvez a maior dificuldade apresentada na recepção a esses alunos, demonstrada pela amplitude dos escores para o CP1 e a dispersão dos resultados. Estudos de caso que comparem esse desempenho com escolas do grupo Baixo-Alto, ou seja, de um mesmo contexto socio-administrativo, mas com avaliação bem mais positiva, podem indicar práticas de gestão escolar potencialmente compartilháveis com essas escolas mais vulneráveis.

Com esse intuito, também se identificou um agrupamento adicional na análise focada na avaliação do retorno pós COVID-19: o de outliers, neste caso positivos, no CP5. Os outliers nesse contexto formam um grupo de 12 escolas que tiveram escore padronizado menor que -3 para o CP5, ou seja fora da distribuição de 99,8% dos valores de menor desvio em relação à média. É relevante notar também que essas escolas não possuem escores igualmente relevantes para o CP1, na verdade estando na média geral para todos os indicadores de contexto socio-administrativo.

Portanto, esse agrupamento outlier tem o potencial de conter escolas cuja capacidade de gestão escolar e cujas práticas e estratégias pedagógicas mais surtiram efeito na avaliação dos alunos sobre o retorno das aulas presenciais. Na Figura 29, é possível observar como o grupo de outliers se diferencia dos demais para as principais variáveis desta análise em questão.

Figura 29: Comparação das escolas por agrupamento CP1-CP5



Fonte: Elaborado pelo autor.

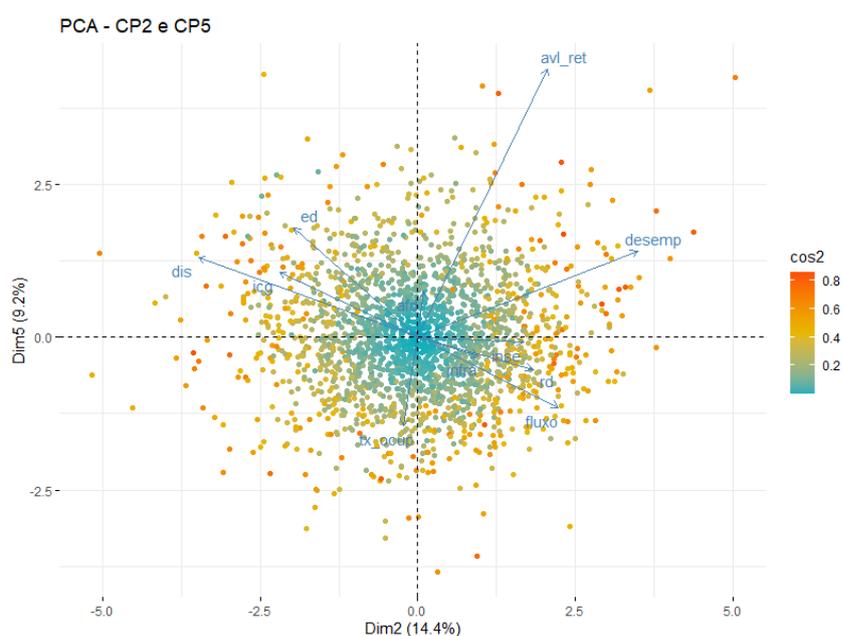
Cabe ressaltar que, em média, essas escolas apresentaram o maior nível de Esforço Docente, apesar de não serem escolas particularmente grandes, como demonstra o Indicador de Complexidade de Gestão da Escola (icg), podendo haver uma relação do sucesso dessas práticas pedagógicas com a maior alocação do tempo disponível dos docentes.

Ainda mais promissor, pensando em uma análise mais qualitativa das práticas pedagógicas adotadas nessas escolas, é a observação de que, de todos os grupos, são também as escolas outliers que apresentaram a maior média do Indicador de desempenho. Isso não foi verificado para o Indicador de Fluxo, contudo é possível que a capacidade de gestão escolar em receber bem os estudantes no retorno das aulas presenciais foi influente para um nível maior de proficiência verificado nas avaliações de rede aplicadas no período.

Análise 5: Relação entre os resultados educacionais (CP2) e a adequação da comunidade escolar aos protocolos sanitárias (COVID) (CP5)

A última análise mantém o foco no CP5, ou seja, na avaliação dos alunos sobre o retorno às aulas presenciais, representada no eixo vertical das Figuras 30 e 31, porém agora comparando com o CP2, que sintetiza os resultados educacionais.

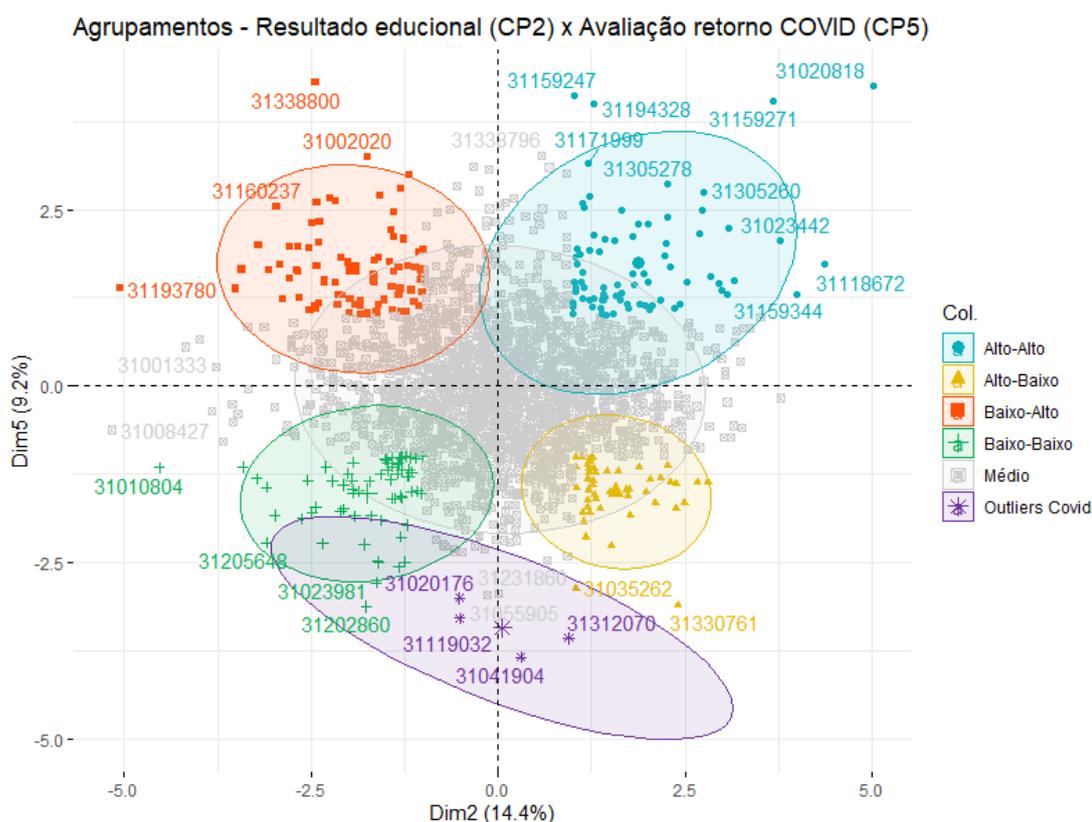
Figura 30: Distribuição dos indivíduos na análise CP2 x CP5



Fonte: Elaborado pelo autor.

Os agrupamentos das escolas (Figura 31) evidenciam uma maior dispersão dos dados no hemisfério superior, associado a uma avaliação positiva do retorno às aulas presenciais. Conclui-se, a partir disso, que uma avaliação positiva, pensando nesta como uma *proxy* da capacidade de gestão escolar, influenciou, mas não foi determinante sozinha para garantir resultados positivos, levando em conta a variação de escores.

Figura 31: Agrupamentos na análise CP2 x CP5



Fonte: Elaborado pelo autor.

Entretanto, uma avaliação negativa tendeu a achatar o topo da amplitude dos resultados educacionais, haja vista a diferença na dispersão presente no grupo Alto-Baixo comparado a dispersão do grupo Alto-Alto. Em média, a avaliação 13% inferior do retorno às aulas presenciais por parte do aluno foi acompanhada também por um indicador de desempenho 7% menor, ainda que acima da média amostral.

Por fim, destacou-se na Figura 31 os outliers negativos que tiveram escore padronizado maior que 3 para o CP5, ou seja fora da distribuição de 99,8% dos valores. Esse total de 6 escolas representa o conjunto das piores avaliações do

retorno às aulas presenciais, sendo também promissor uma análise qualitativa das práticas e estratégias pedagógicas adotadas nesse contexto pós-COVID que podem ter resultado em avaliações mais negativas.

Contudo, diferente dos outliers positivos da última análise, aqui não é possível afirmar que esse grupo destoou significativamente das demais escolas para nenhum indicador dos mais relevantes para a análise dos CP2 e CP5, exceto a própria avaliação do retorno presencial, conforme mostra a Tabela 6. Verifica-se um desempenho abaixo da média, em alinhamento à constatação que escolas com avaliações negativas do retorno tiveram maior dificuldade de ter resultados melhores nesse indicador.

Tabela 6: Resumo descritivo dos agrupamentos CP2-CP5

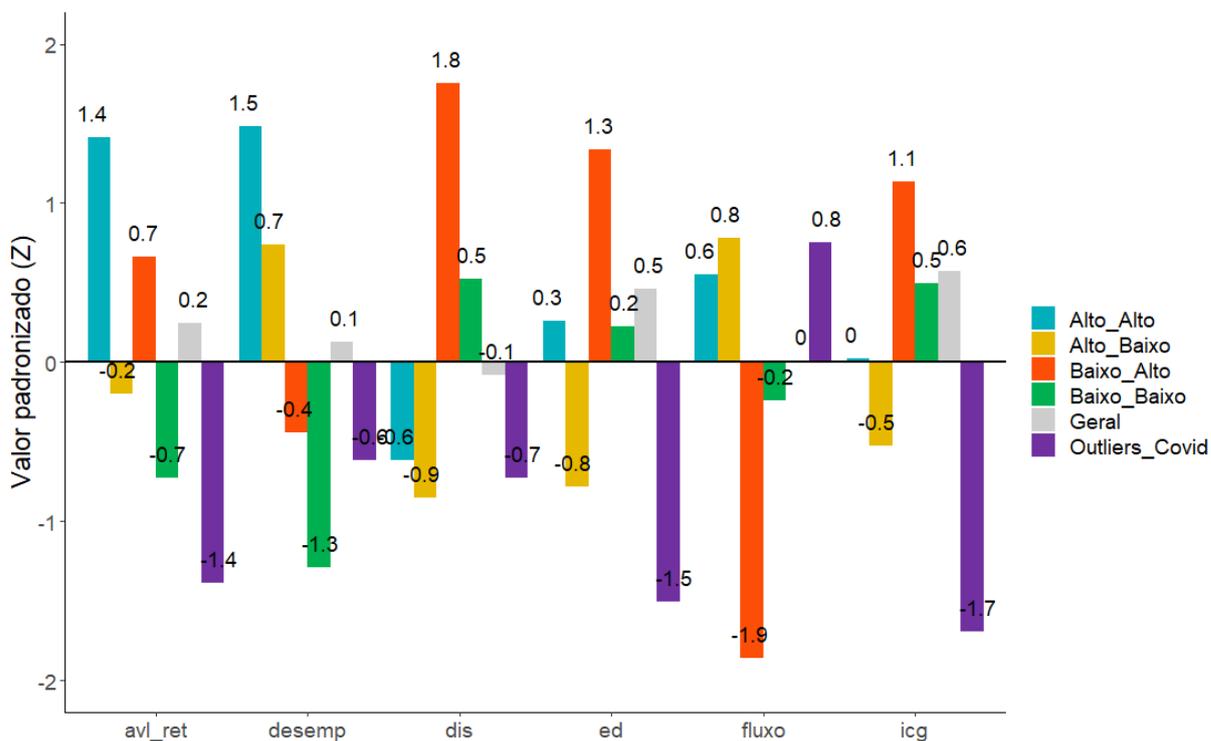
Variável	Contribuição CP2	Contribuição CP5	Média amostral	Grupo (CP2-CP5)					
				Alto-Alto	Baixo-Alto	Alto-Baixo	Baixo-Baixo	Outliers CP5	
Indicador de desempenho	desemp	24,69	6,20	4,11	4,70	3,86	4,37	3,49	3,78
Taxa de distorção idade-série	dis	24,56	5,35	0,16	0,12	0,30	0,10	0,20	0,11
Indicador de fluxo	fluxo	10,04	4,21	91,71	95,42	78,90	96,99	89,96	96,79
Complexidade de gestão da escola	icg	9,76	3,53	4,50	3,97	5,04	3,45	4,42	2,33
Esforço docente	ed	7,84	10,17	3,93	3,87	4,22	3,53	3,86	3,30
Avaliação do retorno às aulas presenciais	avl_ret	8,60	60,93	2,79	3,09	2,89	2,67	2,53	2,36
Regularidade do corpo docente	rd	6,79	0,88	2,97	2,99	2,70	3,29	2,67	2,58
Nível socioeconômico	inse	5,72	0,02	5,83	5,91	5,56	6,47	5,29	5,63
Diagnóstico de infraestrutura	infra	1,87	0,30	0,83	0,84	0,82	0,85	0,81	0,80
Taxa média de ocupação das salas de aula	tx_ocup	0,11	6,61	0,57	0,43	0,51	0,62	0,59	0,42
Adequação da formação docente	afd	0,02	1,81	71,50	71,76	74,70	68,14	69,06	61,45
Número de escolas no grupo				2131	79	82	47	66	6

Fonte: Elaborado pelo autor.

Ainda assim, ao comparar esse grupo de outliers com o agrupamento Baixo-Alto, que por sua vez teve em média uma avaliação 18% melhor, observa-se um Indicador de desempenho médio muito próximo, como é possível verificar na Figura

32. Sugere-se assim, que ainda que a capacidade de gestão da escola possa contribuir para o resultado educacional do aluno, ela é menos influente que outras variáveis, especialmente o Nível Socioeconômico, como já observado em análises anteriores.

Figura 32: Comparação das escolas por agrupamento CP2-CP5



Fonte: Elaborado pelo autor.

5. CONCLUSÃO

Este trabalho teve o objetivo de categorizar as escolas estaduais de Ensino Médio de Minas Gerais por meio da aplicação da técnica de Análise de Componente Principais, buscando principalmente contextualizar indicadores de resultados educacionais, por meio de características da escola relacionadas a seu contexto social, administrativo e de insumos, tais quais corpo docente e infraestrutura.

A partir das análises realizadas, foi possível categorizar as escolas em agrupamentos que levam em conta mais de um aspecto da realidade escolar simultaneamente. Com os componentes principais de “Contexto socio-administrativo” (CP1) e de “Síntese dos resultados educacionais” (CP2), que explicam em conjunto cerca de 35% da variância total, se obteve um resumo dos principais indicadores de resultado educacional contextualizados por características relevantes da escola, principalmente seu nível socioeconômico e nível de complexidade da gestão escolar.

Dessa forma, obteve-se grupos com contextos socio-administrativos comparáveis, porém com resultados distintos, chamando atenção para os grupos Baixo-Alto (128 escolas) e Baixo-Baixo (108 escolas). Em média, os dois grupos são formados por escolas menores e mais pobres, porém há uma diferença relevante no resultado educacional observado para cada grupo. Uma comparação mais qualitativa entre escolas desses agrupamentos pode revelar estratégias pedagógicas ou outros fatores não incluídos no modelo que estão favorecendo o desempenho do grupo Baixo-Alto, já que as variáveis de contexto já estão consideradas nessa comparação.

Outros agrupamentos que chamam a atenção foram obtidos pela combinação do componente referente à “Síntese dos resultados educacionais” (CP2) com o “Desajuste entre desempenho-fluxo escolar” (CP3) e com a “Infraestrutura-Formação docente” (CP4). Em ambas, o grupo Baixo-Baixo correspondeu às escolas com os piores indicadores, totalizando, na primeira combinação, 103 escolas com todos os indicadores de resultado abaixo da média geral e, na segunda combinação, 86 escolas que além do resultado insatisfatório, também possuem indicadores ruins para infraestrutura e formação docente. Em uma perspectiva administrativa, a identificação dessas escolas permitiria a focalização de ações pedagógicas ou de

investimento em pessoal e rede física, já que representam os casos de maior vulnerabilidade no Ensino Médio da rede estadual de Minas Gerais.

A combinação entre o componente de “Síntese dos resultados educacionais” (CP2) e o de “Infraestrutura-Formação docente” (CP4) também evidenciou o grupo Alto-Baixo, que são escolas com resultados positivos, apesar de terem uma infraestrutura e formação docente precária, e o grupo Baixo-Alto, que apresentam resultados ruins, mesmo com professores qualificados e uma rede física satisfatória. Essa categorização poderia indicar uma relação de eficiência no uso desses “insumos” para a obtenção dos resultados educacionais, contudo o resumo descritivo das variáveis para cada agrupamento mostrou que existe uma diferença relevante no nível socioeconômico entre os grupos. O grupo Alto-Baixo, de melhor resultado, tem um nível socioeconômico, em média, 14% maior do que o Baixo-Alto.

Essa constatação, à luz da literatura, coloca em dúvida a dimensão do efeito dos fatores de infraestrutura e corpo docente nos resultados educacionais quando também há o efeito do contexto social. Na análise dos Componentes Principais deste estudo, verificou-se que os dois indicadores citados não contribuíram de forma determinante para outros componentes além do CP4, referente à “Infraestrutura-Formação docente”, e nele não foram acompanhados de outras variáveis. Concluiu-se dessa observação que o resultado educacional esteve mais relacionado com fatores como o nível socioeconômico e o porte da escola. Ademais, professores mais qualificados tenderiam a ter mais facilidade de optar pela escola em que são alocados, não levando em conta necessariamente a demanda da rede de ensino, explicando em parte a relação entre os dois indicadores.

Por fim, o componente relativo à avaliação dos estudantes quanto ao retorno às aulas presenciais (CP5) permitiu concluir que o porte e o nível socioeconômico da escola tiveram baixa influência na avaliação dos estudantes sobre o retorno, entretanto verificou-se menos escolas ricas e de maior nível de complexidade com avaliações negativas. Depreende-se que esse contexto socio-administrativo, mesmo que não preponderante, auxiliou em certa medida na capacidade de gestão escolar.

Por sua vez, um retorno avaliado como positivo foi acompanhado, em média, por resultados educacionais melhores. Na comparação entre os grupos Alto-Alto e Alto-Baixo, o último teve um indicador de desempenho em média 7% pior, mesmo tendo um nível socioeconômico maior do que o primeiro grupo. A dispersão verificada na “Síntese dos resultados educacionais” (CP2) em combinação com a

“Avaliação do retorno às aulas presenciais” (CP5) indicou também que: escolas com avaliações positivas sobre o retorno variaram mais nos resultados educacionais e as escolas com avaliações negativas tenderam a ter uma dispersão menor, porém com resultados, em média, inferiores.

Logo, a Análise de Componentes Principais permitiu tanto fortalecer as constatações já levantadas na literatura, a respeito dos fatores mais influentes nos resultados educacionais – em que ficou claro aqui também a importância do contexto socio-administrativo – como também auxiliou na classificação e priorização daqueles casos mais vulneráveis. Contudo, fica evidente que existe uma parcela considerável da variabilidade que não foi possível explicar somente com os componentes mantidas no modelo, que explicam 65% da variância total. Portanto, existe um espaço para trabalhos futuros que incluam outros indicadores que ajudem a explicar essa variabilidade extra ou também análises qualitativas que aprofundem experiências específicas de escolas nos agrupamentos sugeridos.

6. REFERÊNCIAS

- ABDI, Hervé; WILLIAMS, Lynne J. Principal componente Analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, Fairfax, v. 34, n. 2, p. 433-459, jul./agos. 2010. Disponível em:
<<https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.101> > Acesso em: 1 dez. 2022.
- ALMEIDA, Luana Costa; DALBEN, Adilson; FREITAS, Luiz Carlos de. O Ideb: limites e ilusões de uma política educacional. **Educação & Sociedade**, Campinas, v. 34, n. 125, p. 1153-1174, out./dez. 2013. Disponível em:
<<https://www.scielo.br/j/es/a/FGHLWhm47PZpFShWNxs5GLM/abstract/?lang=pt#>> Acesso em: 1 dez. 2022.
- BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil: promulgada em 5 de outubro de 1988**. Brasília: 1988. Disponível em:
<http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm> Acesso em: 3 dez. 2022.
- BRASIL. Lei nº. 9394, de 25 de junho de 1996. **Lei de Diretrizes Básicas da Educação**. Brasília, DF. 1996. Disponível em:
<https://www.planalto.gov.br/ccivil_03/leis/l9394.htm>. Acesso em 12 dez. 2022.
- BARRETTO, Elba Siqueira de Sá. Políticas de formação docente para a educação básica no Brasil: embates contemporâneos. **Revista Brasileira de Educação**. Rio de Janeiro, v. 20, n. 62, p. 679-701, jul./set. 2015. Disponível em:
<<https://www.scielo.br/j/rbedu/a/6dBCYcmPwf9BM447tNpYpgR/abstract/?lang=pt>>. Acesso em: 3 dez. 2022.
- CABRAL, Pedro Henrique Trindade Dias. **Controle de Processos em Mineração baseado em Análise por Componentes Principais: Um Estudo de Caso Real**. Monografia (especialização) - Universidade Federal de Minas Gerais, ICX - Departamento de Estatística, 2018. Disponível em:
<https://repositorio.ufmg.br/bitstream/1843/33538/1/Artigo_Pedro%20Dias_Especializa%C3%A7ao_Estatistica.pdf>. Acesso em 3 dez. 2022.
- CHIRINÉA, Andréia Melanda. **O Índice de Desenvolvimento da Educação Básica (IDEB) e as dimensões associadas à qualidade da educação na escola pública**

municipal. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências de Marília, 2010. Disponível em:

<<https://repositorio.unesp.br/handle/11449/96363>>. Acesso em 2 dez. 2022.

CHIRINÉA, Andréia Melanda; BRANDÃO, Carlos da Fonseca. O IDEB como política de regulação do Estado e legitimação da qualidade: em busca de significados.

Avaliação e Políticas Públicas em Educação. Rio de Janeiro, v. 23, n. 87, p. 461-484, abr./jun. 2015. Disponível em:

<<https://www.scielo.br/j/ensaio/a/L3FwTBZ79fLPRRwHFfVgmkP/abstract/?lang=pt>>. Acesso em: 3 dez. 2022.

COLEMAN, James S. et al. **Equality of educational opportunity**. Office of Education and Welfare, Washington, DC. 1966.

FLECK, Marcelo P.A. e BOURDEL, Marie C. Método de simulação e escolha de fatores na análise dos principais componentes. **Revista de Saúde Pública**. São Paulo, v. 32, n. 3, p. 267-272, jun. 1998. Disponível em:

<<https://www.scielo.br/j/rsp/a/zWdnX5bLGcwXGkcyjQjSG6C/abstract/?lang=pt>>. Acesso em 4 dez. 2022.

FUNAI. Fundação Nacional do Índio. **Educação Escolar Indígena**. Disponível em: <<https://www.gov.br/funai/pt-br/atuacao/povos-indigenas/cidadania/educacao-escolar-indigena>>. 2020. Acesso em: 2 de dez. 2022.

GARCIA, Adir Valdemar; HILLESHEIM, Jaime. Pobreza e desigualdades educacionais: uma análise com base nos Planos Nacionais de Educação e nos Planos Plurianuais Federais. **Educar em Revista**. Curitiba, Edição Especial n. 2, p. 131-147, set. 2017. Disponível em:

<<https://www.scielo.br/j/er/a/s4Z4xXszc389JhTJKvr7kXv/abstract/?lang=pt>>. Acesso em 3 dez. 2022.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, n. 24, p. 417-441, 498-520, 1933

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira; Ministério da Educação. **Nota Técnica: Índice de Desenvolvimento da Educação Básica - Ideb**. Brasília. 2007. Disponível em:

<https://download.inep.gov.br/educacao_basica/portaI_ideb/o_que_e_o_ideb/Nota_Tecnica_n1_concepcaoIDEB.pdf>. Acesso em 2 dez. 2022.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira; Ministério da Educação. **NOTA TÉCNICA No 11/2015: Indicador de regularidade do docente da educação básica**. Brasília, 2015. Disponível em:

<https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_regularidade_vinculo/nota_tecnica_indicador_regularidade_2015.pdf>.

Acesso em 2 dez. 2022.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira; Ministério da Educação. **NOTA TÉCNICA No 020/2014: Indicador de adequação da formação do docente da educação básica**. Brasília, 2014a. Disponível em:

<https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_formacao_legal/nota_tecnica_indicador_docente_formacao_legal.pdf>.

Acesso em 2 dez. 2022.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira; Ministério da Educação. **NOTA TÉCNICA No 039/2014: Indicador de Esforço Docente**. Brasília, 2014b. Disponível em:

<https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_esforco/nota_tecnica_indicador_docente_esforco.pdf>. Acesso em 2

dez. 2022.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira; Ministério da Educação. **NOTA TÉCNICA No 040/2014: Indicador para mensurar a complexidade da gestão nas escolas a partir dos dados do Censo Escolar da Educação Básica**. Brasília, 2014c. Disponível em:

<https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/escola_complexidade_gestao/nota_tecnica_indicador_escola_complexidade_gestao.pdf>. Acesso em 2 dez. 2022

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira; Ministério da Educação. **Resumo Técnico: Resultados do Índice de Desenvolvimento da Educação Básica**. Brasília. 2020. Disponível em:

<https://download.inep.gov.br/educacao_basica/portaI_ideb/planilhas_para_download/2019/resumo_tecnico_ideb_2019_versao_preliminar.pdf>. Acesso em 1 dez. 2022.

KAISER, H.F. The varimax criteria for analytical rotation in factor analysis. *Psychometrika*. n. 23, p. 141-51, 1958.

MINAS GERAIS. Secretaria de Estado de Educação; Sistema Mineiro de Avaliação e Equidade da Educação Pública. **Revista da Rede**. Juiz de Fora, v. 2, 2021. Disponível em: <<https://simave.educacao.mg.gov.br/#!/colecoes>>. Acesso em 2 dez. 2022.

MINAS GERAIS. Secretaria de Estado de Educação de Minas Gerais, 2022. **Governo de Minas anuncia novas etapas do Mãos à Obra na Escola, com investimento de R\$ 270 milhões nas unidades da rede estadual**: Diagnóstico de Infraestrutura direciona ações do programa, que soma investimento recorde de R\$689 milhões em 1.706 escolas. Disponível em: <<https://www2.educacao.mg.gov.br/leis/story/12036-governo-de-minas-anuncia-novas-etapas-do-maos-a-obra-na-escola-com-investimento-de-r-270-milhoes-nas-unidades-da-rede-estadual>>.2022. Acesso em: 2 dez. 2022.

MINGOTI, Sueli Aparecida. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Ed. UFMG, 2005. 295p.

MINGOTI, Sueli Aparecida; SILVA, Aldy Fernandes da. Um exemplo de aplicação de técnicas de estatística multivariada na construção de índices de preços. **Nova Economia**. Belo Horizonte, v. 7, n. 2, p. 203-212, 2013. Disponível em: <<https://revistas.face.ufmg.br/index.php/novaeconomia/article/view/2249>>. Acesso em: 1 dez. 2022.

PASSADOR, Cláudia Souza; SALVETTI, Thales Silveira. Gestão escolar democrática e estudos organizacionais críticos: convergências teóricas. **Educação & Sociedade**. Campinas, v. 34, n. 123, pp. 477-492, abr./jun. 2013. Disponível em: <<https://www.scielo.br/j/es/a/BPVZ9Q4dq99vtBxtJKYwVNB/?lang=pt>>. Acesso em 3 dez. 2022.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, Series 6, 2, p. 559-572, 1901.

PORTELA, Alysson Lorenzon; BUSSMANN, Tanise Brandão; OLIVEIRA, Ana Maria Hermeto de. A relação de fatores individuais, familiares e escolares com a distorção idade-série no ensino público brasileiro. **Nova Economia**. Belo Horizonte, v. 27, n. 3,

p. 477-509, 2018. Disponível em:

<<https://revistas.face.ufmg.br/index.php/novaeconomia/article/view/3138>>. Acesso em 2 dez. 2022.

RAMOS, Mozart Neves et al. Uma análise estatística multivariada do desempenho das escolas municipais de Ribeirão Preto. **Avaliação e Políticas Públicas em Educação**. Rio de Janeiro, v. 29, n. 113, p. 857-873, out./dez. 2021. Disponível em: <<https://www.scielo.br/j/ensaio/a/txyjBh67gQp4dNKrHmPCc5P/?lang=pt>>. Acesso em 2 dez. 2022.

REGAZZI, A. J. **INF 766 - Análise multivariada**. Viçosa: Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas. Departamento de Informática, 2001. 166p. Apostila de disciplina.

SANTOS, Mariana Cristina Silva et al. Programa Bolsa Família e indicadores educacionais em crianças, adolescentes e escolas no Brasil: revisão sistemática. **Ciência & Saúde Coletiva**. Brasília, v. 24, n. 6, pp. 2233-2247, 2019. Disponível em: <<https://www.scielo.br/j/csc/a/yqq93sK7XtqR5MYb4GQJMsC/?lang=pt>>. Acesso em 3 dez. 2022.

SANTOS, Renata Costa. **Identificação das Variáveis que mais Influenciam no Custo Assistencial da Saúde Suplementar: Estudo de caso de uma operadora de plano de saúde**. Monografia (especialização) - Universidade Federal de Minas Gerais, ICX - Departamento de Estatística, 2014.

SILVA, Nelson do Valle; HASENBALG, Carlos. Tendências da desigualdade educacional no Brasil. **Dados**. Rio de Janeiro, v. 43, n. 3, p. 423-445, 2000. Disponível em: <<https://www.scielo.br/j/dados/a/pjz8g84tSdvrM6p3GbwdNtk/?lang=pt>>. Acesso em 2 dez. 2022.

SOARES, J. F.; RIGOTTI, J. I. R.; ANDRADE, L. T. As desigualdades socioespaciais e o efeito das escolas públicas de Belo Horizonte. *In*: RIBEIRO, L. C. D. Q.; KATZMAN, R. **A Cidade contra a Escola?** 1ª. ed. Rio de Janeiro: Letra Capital, 2008, p. 367.

TEIXEIRA, André Luiz da Silva; TUPY, Igor Santos; AMARAL, Pedro Vasconcelos Maia do. **A percepção dos benefícios e dificuldades na interação universidade-**

empresa: o caso dos grupos de pesquisa mineiros. Artigo de Periódico - Universidade Federal de Minas Gerais, FCE - DEPARTAMENTO DE CIÊNCIAS ECONÔMICAS, 2016.

VARELLA, Carlos Alberto Alves. **Análise Multivariada Aplicada as Ciências Agrárias: Análise de Componentes Principais.** Universidade Federal Rural do Rio de Janeiro, 2008.

VICINI, Lorena. **Análise multivariada da teoria à prática.** Monografia (especialização) – Universidade Federal de Santa Maria, CCNE, 2005.