

SIAREM

Seminário Integrado de Avaliação das  
Redes Públicas Estaduais e Municipais

# As medidas resultantes dos questionários contextuais

LUIZ VICENTE FONSECA RIBEIRO

CAEd/UFJF

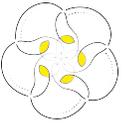


# SIAREM

Seminário Integrado de Avaliação das  
Redes Públicas Estaduais e Municipais



## Considerações Iniciais



## ■ OS INSTRUMENTOS

---

Levando em conta a importância de contextualizar os fatores e práticas que impactam a aprendizagem dos estudantes e, conseqüentemente, seu desempenho, o CAEd utiliza questionários para obter informações sobre aspectos que influenciam o cotidiano escolar. Com o objetivo de compreender de forma abrangente os contextos dos estudantes, professores e diretores e escolas, foram desenvolvidos instrumentos específicos para medir fatores associados à aprendizagem. Discutiremos a seguir alguns aspectos de inovação nos seguintes instrumentos:

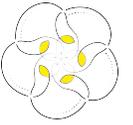
- **Questionário do Estudante**
- **Questionário do Professor**
- **Questionário do Diretor**

# SIAREM

Seminário Integrado de Avaliação das  
Redes Públicas Estaduais e Municipais



## ESTUDANTE



## ■ INSTRUMENTO E MEDIDAS

---

### ❖ PERFIL SOCIODEMOGRÁFICO

- Sexo
- Cor ou raça
- Idade

### ❖ AUTORREGULAÇÃO DA APRENDIZAGEM

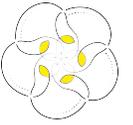
- Engajamento
- Apoio dos pais
- Autonomia
- Expectativas

### ❖ NÍVEL SOCIOECONÔMICO

- Capital social
- Capital econômico
- Capital cultural

### ❖ AMBIENTE ESCOLAR E APRENDIZAGEM

- Clima escolar
- Segurança no contexto escolar
- Práticas pedagógicas



## ■ NOVIDADES PARA O CICLO 2024/2024

---

Ajustes anuais nos instrumentos e nas devolutivas visam melhorar a qualidade da coleta e apresentação dos dados contextuais.

Alguns exemplos:

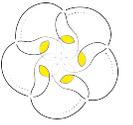
- ❖ Os instrumentos passaram a ter um rodízio de blocos para reduzir o número de itens respondidos, de cerca de 60 itens na série histórica, para aproximadamente 30 itens em 2024.
- ❖ Neste novo ciclo vamos utilizar uma metodologia para a construção da medida de nível socioeconômico (NSE) que permitirá a comparação ajustada entre os resultados de 2023 e 2024.
- ❖ Será revista também a forma de apresentação dos dados sobre a desigualdade de desempenho, buscando melhorar compreensão sobre a equidade nos resultados da avaliação.

# SIAREM

Seminário Integrado de Avaliação das  
Redes Públicas Estaduais e Municipais



## PROFESSOR



## ■ INSTRUMENTO E MEDIDAS

---

### ❖ PERFIL SOCIODEMOGRÁFICO

- Sexo
- Cor ou raça
- Faixa etária

### ❖ TRAJETÓRIA PROFISSIONAL

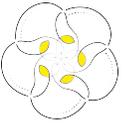
- Formação inicial
- Experiência

### ❖ AMBIENTE ESCOLAR

- Clima escolar
- Segurança no contexto escolar

### ❖ PRÁTICAS PEDAGÓGICAS

- Limitações à prática docente
- Atividades e avaliações
- Atitudes sobre práticas de ensino
- Atitudes sobre avaliações internas e externas
- Reflexões sobre as práticas pedagógicas (itens de resposta aberta)



## ■ A CONSTRUÇÃO DOS ITENS DE RESPOSTA ABERTA

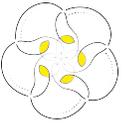
---

### OBJETIVO

Promover a participação ativa e reflexiva dos professores na coleta de informações sobre suas práticas pedagógicas, incentivando o compartilhamento de aspectos relevantes de sua atuação, de modo a enriquecer a análise do ambiente escolar com perspectivas e nuances fundamentais.

### TEMAS

1. As práticas pedagógicas contextualizadas
2. Apropriação dos resultados de avaliações educacionais



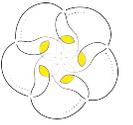
## ■ A CONSTRUÇÃO DOS ITENS DE RESPOSTA ABERTA

---

### FUNDAMENTAÇÃO

A Base Nacional Comum para a Formação de Professores (BNC-Formação) é um documento que apresenta uma matriz com competências e habilidades específicas para a atuação do professor.

Reconhecendo que as discussões em torno da BNC-Formação têm gerado preocupação em educadores e especialistas na área de educação, as habilidades foram cuidadosamente escolhidas levando-se em consideração a literatura especializada da área.



## ■ A CONSTRUÇÃO DOS ITENS DE RESPOSTA ABERTA

---

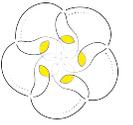
*(habilidade 1.2.4) Articular estratégias e conhecimentos que permitam aos estudantes desenvolver as competências necessárias, bem como favorecer o desenvolvimento de habilidades de níveis cognitivos superiores.*

1) “Quando você precisa trabalhar conteúdos mais complexos e de difícil aprendizagem com os estudantes, que tipo de prática pedagógica você prioriza?”

2) “Por que você considera a prática que você descreveu anteriormente, a mais adequada para trabalhar conteúdos mais complexos e de difícil aprendizagem com os estudantes?”

**A análise das respostas dos professores envolve as possíveis gradações:**

- Respostas mais difusas e generalistas
- Respostas mais estruturadas, que articulam práticas pedagógicas aos desafios do conteúdo
- Respostas mais completas e reflexivas que apresentam exemplos e analisam os impactos das práticas na aprendizagem dos estudantes



## ■ A CONSTRUÇÃO DOS ITENS DE RESPOSTA ABERTA

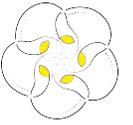
---

*(habilidade 2.3.4) Aplicar os métodos de avaliação para analisar o processo de aprendizagem dos estudantes e utilizar esses resultados para retroalimentar a prática pedagógica.*

1) “Como você interpreta e utiliza o resultado das avaliações realizadas pelo estudante?”

**A análise das respostas dos professores envolve dois aspectos principais.**

- ***Interpretação dos dados:*** espera-se que os professores utilizem indicadores, como a proficiência dos estudantes, a taxa de acertos e erros, o domínio de habilidades e o progresso individual e coletivo, para identificar padrões, lacunas e oportunidades de melhoria.
- ***Aplicação dos dados na prática docente:*** os resultados das avaliações podem fundamentar decisões estratégicas, como planejar intervenções, fornecer *feedback* construtivo, propor atividades remediativas, adaptar o ritmo de ensino e implementar projetos colaborativos, respeitando a singularidade de cada contexto educacional.



## ■ METODOLOGIA: PROCESSAMENTO DE LINGUAGEM NATURAL (PLN)

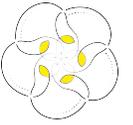
---

Para analisar os **itens de respostas construídas** dos professores, utilizamos o arcabouço do Processamento de Linguagem Natural (PLN), que une conceitos e metodologias de análise do campo da Linguística à computação aplicada. Dessa forma, é possível analisar dados de texto em grandes quantidades de forma eficiente e precisa.

Análises de PLN envolvem algumas etapas básicas:

- Pré-processamento e remoção de pontuação, símbolos e *stopwords* (artigos, preposições e palavras de baixa relevância semântica), transformação para minúsculas etc
- Lematização ou redução à raiz (*stemming*)
- Vetorização (ou transformação “numérica”)
- Análise final

Obs.: As etapas e as estratégias usadas podem variar de acordo com o objetivo de análise.

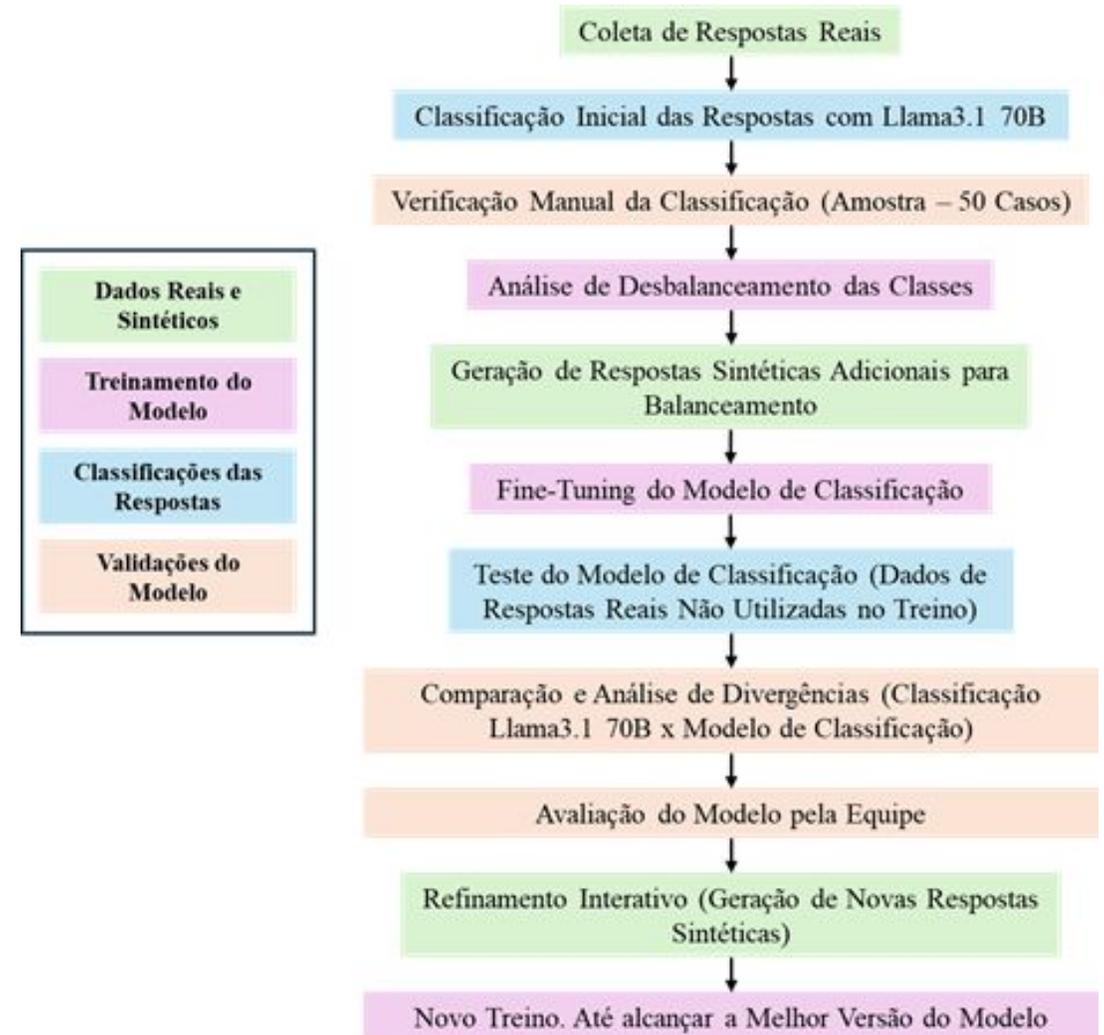


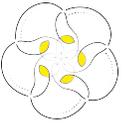
## ■ CLASSIFICAÇÃO DAS RESPOSTAS DOS PROFESSORES

Para análise dos dois primeiros itens (relacionados às práticas pedagógicas para tratar conteúdos complexos), a metodologia utilizada foi desenvolvida com o objetivo de **treinar e validar** um **modelo de classificação** que analisa o **grau de reflexão** presente nas **práticas pedagógicas** relatadas por professores.

A abordagem proposta envolveu um conjunto de etapas, desde a classificação inicial das respostas, a geração e balanceamento do conjunto de dados de treinamento, até o ajuste fino (*fine-tuning*) de um modelo pré-treinado em português (BERTimbau), bem como a subsequente validação e refinamento contínuo do modelo.

Figura: Processo de treinamento e validação do modelo de classificação





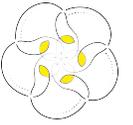
## ■ UTILIZAÇÃO DE MODELO DE LINGUAGEM PARA CLASSIFICAÇÃO

### Modelo de Linguagem Generativo (Llama3.1 70B) no Processo Metodológico

- Utilizado na **classificação inicial** das respostas (baseado em uma chave de correção com valores entre 0 e 3).
- Aplicado na **geração de respostas sintéticas** para balanceamento do banco de dados.
- Empregado para **análises de divergências** entre as classificações iniciais e as classificações realizadas pelo modelo treinado.
- **Auxiliou na estruturação e validação** do modelo de classificação, assegurando resultados mais consistentes e confiáveis.

### Resultados e Processo de Validação do Modelo

- **Validade do Modelo:** Modelo ajustado apresentou **resultados consistentes** em dados de teste e na avaliação qualitativa realizada pela equipe.
- **Robustez e Consistência:** Foram observadas melhorias na **generalização** do modelo e **redução de vieses**, garantindo desempenho confiável em diferentes cenários e conjuntos de dados.
- **Avaliação e Refinamento:** O desenvolvimento seguiu uma abordagem iterativa, com **validação contínua** realizada pela equipe e suporte de um modelo generativo. Esse processo possibilitou o **refinamento constante** do modelo por meio de ajustes nos hiperparâmetros e reavaliações baseadas em métricas de desempenho.



## ■ EXEMPLOS DE RESPOSTAS E SUAS CLASSIFICAÇÕES

---

### Nível 0

**Prática pedagógica:** *"A internet."*

**Justificativa:** *"Devido ao acesso de todos."*

### Nível 1

**Prática pedagógica:** *"Trabalho em grupo com descritores."*

**Justificativa:** *"Os alunos com facilidade ajudam os alunos com muita dificuldade."*

### Nível 2

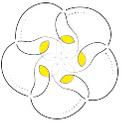
**Prática pedagógica:** *"Priorizo a exposição com interação."*

**Justificativa:** *"Priorizo a exposição com recursos como slides, que apresento de forma dialógica, buscando sempre o feedback dos estudantes."*

### Nível 3

**Prática pedagógica:** *"Priorizo sempre metodologias ativas para engajar os alunos. Resolução de questões envolvendo situações problemas, trazendo conexões com a vida real."*

**Justificativa:** *"As metodologias ativas, como a Aprendizagem Baseada em Problemas e o Ensino Investigativo, mudam o foco da educação para o estudante, fazendo com que ele assuma um papel central no processo de aprendizagem."*



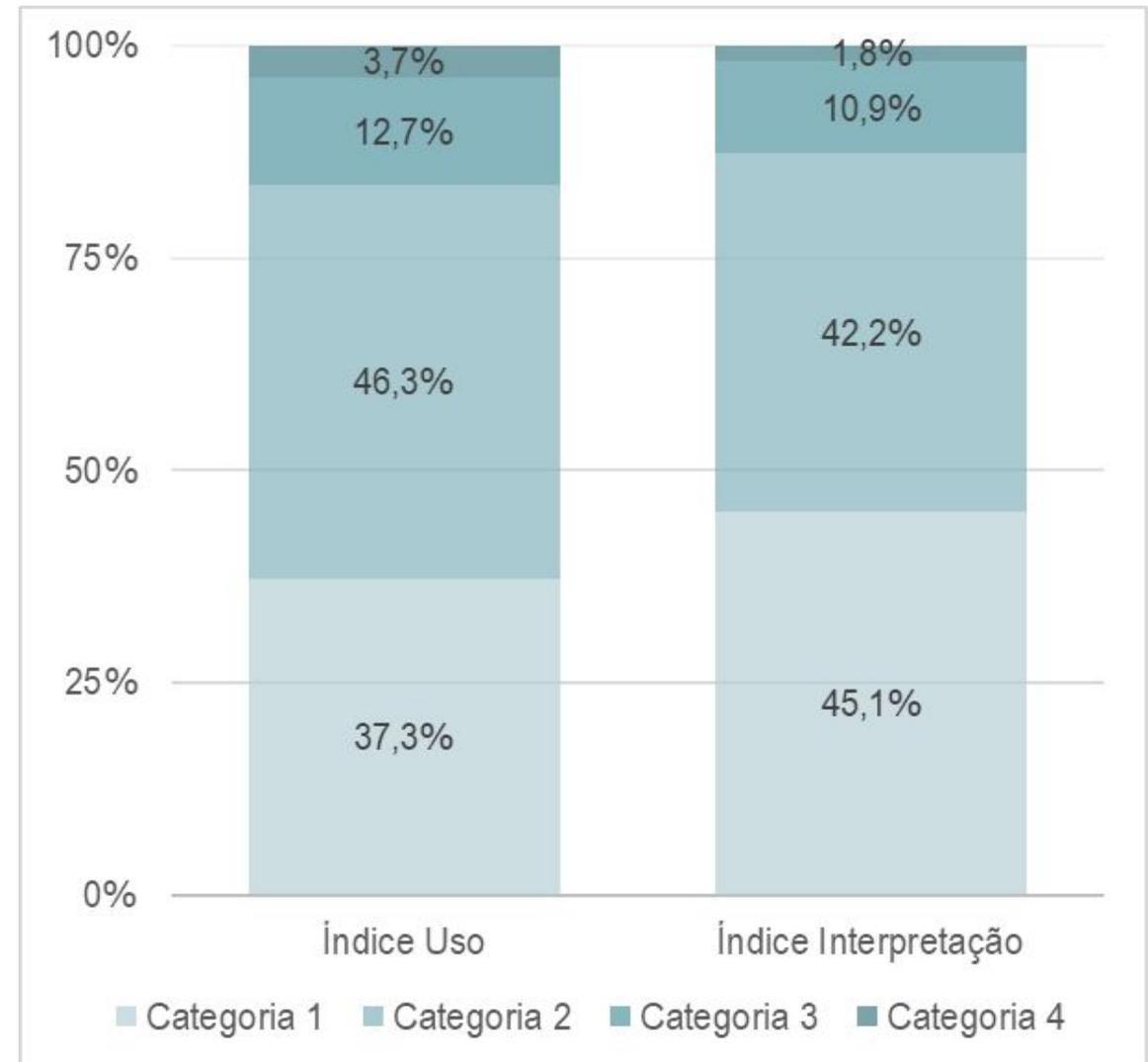
## ■ INTERPRETAÇÃO E UTILIZAÇÃO DE RESULTADOS DE AVALIAÇÕES

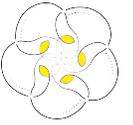
Para o item sobre a **interpretação** e **utilização** de resultados de avaliações, a estratégia de análise foi utilizar algoritmos de **similaridade textual** para identificar o alinhamento semântico das respostas abertas com base em palavras-chave de "utilização" e "interpretação":

1. A vetorização foi feita por *embeddings* de um modelo pré-treinado para o português
2. Definimos um conjunto de palavras-chave para o cálculo da semelhança
3. Utilizamos mais uma camada de semelhança, com conceitos-chave como filtro semântico

Tratamos os resultados de forma que cada resposta recebesse um índice com valor entre 0 e 1. Depois, dividimos essa escala em 4 categorias com intervalos 0,25

Gráfico: distribuição dos índices de uso e interpretação





## ■ EXEMPLOS DE RESPOSTAS

Respostas com escores próximos de zero são respostas consideradas inválidas, como:

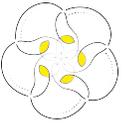
- *“Primordiais.”* (0,005) ou *“Regular.”* (0,007)
- *“Eles têm dificuldade.”* (0,02)
- *“Fazendo a correção das atividades e interagindo com eles.”* (0,15)

Respostas com algum conteúdo já recebem um escore maior:

- *“Vejo seus pontos fracos e busco tratar deles nos conteúdos futuros.”* (0,26)
- *“De forma qualitativa, a fim de identificar os avanços e as lacunas de aprendizagem existentes no processo de ensino aprendizagem.”* (0,40)

Respostas mais refinadas alcançam melhores escores:

- *“Utilizo para ver os conteúdos nos quais os alunos não tiveram um bom desempenho para reforçar adotando outras práticas pedagógicas.”* (0,53)
- *“Eu interpreto como parte do desenvolvimento deles, sempre levo em conta o desenvolvimento e participação em sala. Entendo as avaliações como um compilado dos conteúdos formais com as ações dos alunos em sala.”* (0,64)
- *“As avaliações têm também a função de auxiliar o professor a traçar um plano que contribua para a aprendizagem. Se certos conteúdos trabalhados tiveram um alto índice de erros, significa que não ficou tão claro para os alunos e que esse assunto deve ser reforçado usando novas estratégias.”* (0,87)

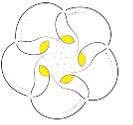


## ■ EXEMPLOS DE RESPOSTAS

---

Por fim, a resposta com maior índice em ambas as dimensões (valor 1,0) foi a seguinte:

*“Os resultados das avaliações que os estudantes realizam funcionam, para mim, como um mapa. Eles me mostram onde cada aluno está na jornada de aprendizagem. Não vejo a avaliação apenas como uma forma de dar uma nota; ela é uma ferramenta para entender o que está funcionando e o que ainda precisa de ajustes, tanto para mim quanto para os próprios alunos. Quando olho os resultados, procuro identificar padrões: quem tá com mais dificuldade, em que tipo de questão ou conceito isso tá acontecendo, e como posso ajudar. Se eu percebo que muitos alunos erraram a mesma coisa, isso é um sinal claro de que eu preciso revisar aquele conteúdo de um jeito diferente, talvez com uma abordagem mais prática ou com mais exemplos. Além disso, as avaliações me ajudam a traçar caminhos personalizados. Se vejo que um aluno domina uma parte do conteúdo, posso desafiar-lo com algo mais complexo. Por outro lado, quem tá enfrentando mais dificuldade, eu ofereço atividades mais direcionadas, reforço, ou até mesmo explicações individuais, se necessário. Resumindo, os resultados são mais do que números; são bússolas que me ajudam a ajustar o rumo do aprendizado de cada um e da turma como um todo, garantindo que ninguém fique para trás.”*



## ■ ASPECTOS DESTACADOS

---

O Processamento de Linguagem Natural permite a extração de análises mais refinadas a partir dos **itens de respostas construídas** dos questionários dos professores.

Para o próximo ciclo de avaliações, seremos capazes de capturar uma ampla gama de respostas diferentes dos professores e produzir medidas específicas com relação às seguintes dimensões de sua atuação profissional:

- Quais práticas pedagógicas são consideradas mais eficientes pelos respondentes
- O motivo pelo qual o profissional considera essa prática mais eficiente
- A capacidade de reflexão do profissional com relação às suas metodologias de ensino
- O grau de utilização e interpretação dos resultados de avaliações pelos professores

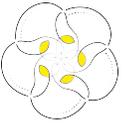
Esses dados nos fornecem insumos ricos para a tomada de decisão e formulação de políticas públicas que visam à melhoria do ensino nas Redes públicas brasileiras.

# SIAREM

Seminário Integrado de Avaliação das  
Redes Públicas Estaduais e Municipais



## DIRETOR



## ■ INSTRUMENTO E MEDIDAS

---

### ❖ PERFIL SOCIODEMOGRÁFICO

- Sexo
- Cor ou raça
- Faixa etária

### ❖ TRAJETÓRIA PROFISSIONAL

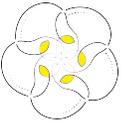
- Formação inicial
- Experiência
- Forma de ingresso na gestão

### ❖ AMBIENTE ESCOLAR

- Clima escolar

### ❖ PERFIL DE GESTÃO

- Político-Institucional
- Pedagógica
- Administrativo-Financeira
- Pessoal e Relacional

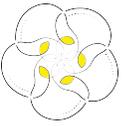


## ■ A CONSTRUÇÃO DO PERFIL DE GESTÃO DE DIRETORES

---

### OBJETIVO

Tendo como referência a Base Nacional Comum de Competências do Diretor Escolar (BNC-Diretor Escolar), as questões apresentadas no questionário permitem identificar diferentes perfis de liderança escolar, os quais se conformam por processos de combinação e recombinação de prioridades em face: (i) de qualidades pessoais e profissionais dos diretores e membros da equipe; (ii) de pressões exercidas pelas demandas próprias às diferentes frentes de atuação existentes na escola; (iii) das condições materiais da escola e do nível socioeconômico do público atendido; (iv) do tipo de rede de ensino em que se inserem; e (v) da cultura institucional estabelecida.



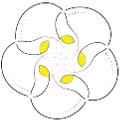
## ■ A CONSTRUÇÃO DO PERFIL DE GESTÃO DE DIRETORES

---

### FUNDAMENTAÇÃO

As assertivas dos itens foram elaboradas com base na matriz de competências do diretor, prevista na Base Nacional Comum de Competências do Diretor Escolar (BNC-Diretor Escolar), aprovada pelo CNE em 2021. Essa matriz define dez competências gerais, desdobradas em 17 competências específicas, distribuídas em quatro dimensões: Político-Institucional, Pedagógica, Administrativo-Financeira e Pessoal e Relacional.

Para cada uma das dimensões foram desenvolvidas 7 assertivas baseadas nas descrições das respectivas competências. Essas 28 assertivas foram organizadas em pares, de forma que todas as assertivas de uma das dimensões fossem confrontadas com ao menos uma alternativa das outras 3 dimensões. Assim, o bloco da escolha forçada foi constituído por 42 itens com duas alternativas cada.

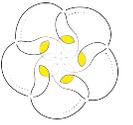


## ■ A CONSTRUÇÃO DO PERFIL DE GESTÃO DE DIRETORES

Bloco de “escolha forçada”: está estruturado com base em um conjunto de frases descritivas de comportamento, em pares, e exige que o respondente escolha a frase que melhor descreve seu desempenho. Exemplo:

**Dentre o par de atribuições abaixo escolha aquela que, durante a sua gestão na instituição, você busca favorecer com maior frequência:**

- |         |  |     |  |
|---------|--|-----|--|
| 20. (A) | Garantir à comunidade escolar acesso às informações sobre as atividades desenvolvidas na escola.                                   | (B) | Focalizar o trabalho na construção de consensos sobre expectativas altas e equânimes da aprendizagem para toda a escola. |
| 21. (A) | Fomentar o debate sobre a construção das políticas educacionais com as famílias e a comunidade.                                    | (B) | Ter conhecimento das características pedagógicas próprias de cada etapas e disciplina ofertada pela escola.              |
| 22. (A) | Criar estratégias e assegurar que toda a comunidade escolar tenha opiniões e crenças respeitadas e sejam respeitosos com os pares. | (B) | Garantir, na rotina da escola, momentos de troca, planejamento e avaliação entre os professores.                         |



## ■ ANÁLISE E COMPOSIÇÃO DA MEDIDA

---

### Criação dos escores:

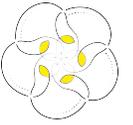
- Foram criados quatro escores, um para cada dimensão da BNC-Diretor Escolar
- Cada escore foi calculado pelo somatório das alternativas assinaladas pelo diretor em itens relacionados a uma dimensão específica

### Intervalo de valores dos escores:

- O valor máximo de um escore é **15**, quando o diretor escolhe todas as alternativas relacionadas a uma dimensão
- O valor mínimo é **0**, quando nenhuma alternativa da dimensão é assinalada

### Pressuposto de equilíbrio nas dimensões:

- Supõe-se que os diretores, em média, priorizam as quatro dimensões de forma equilibrada, por serem igualmente desejáveis
- A probabilidade de escolha de uma dimensão em detrimento de outra é **50%** para cada item
- A média esperada para cada dimensão é **7,5** pontos



## ■ ANÁLISE E COMPOSIÇÃO DA MEDIDA

---

### Análise de priorização estatisticamente relevante:

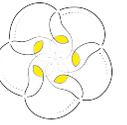
- A tendência ao equilíbrio permite verificar se um diretor prioriza alguma dimensão de forma significativa
- A distribuição binomial é usada para calcular a probabilidade de um diretor obter uma pontuação específica em uma dimensão

### Teste de hipóteses:

- **Hipótese nula ( $H_0$ ):** O diretor prioriza a dimensão de forma esperada ( $X = 7,5$ )
- **Hipótese alternativa ( $H_1$ ):** O diretor prioriza muito ou pouco a dimensão ( $X \neq 7,5$ )

### Limites para rejeição da hipótese nula:

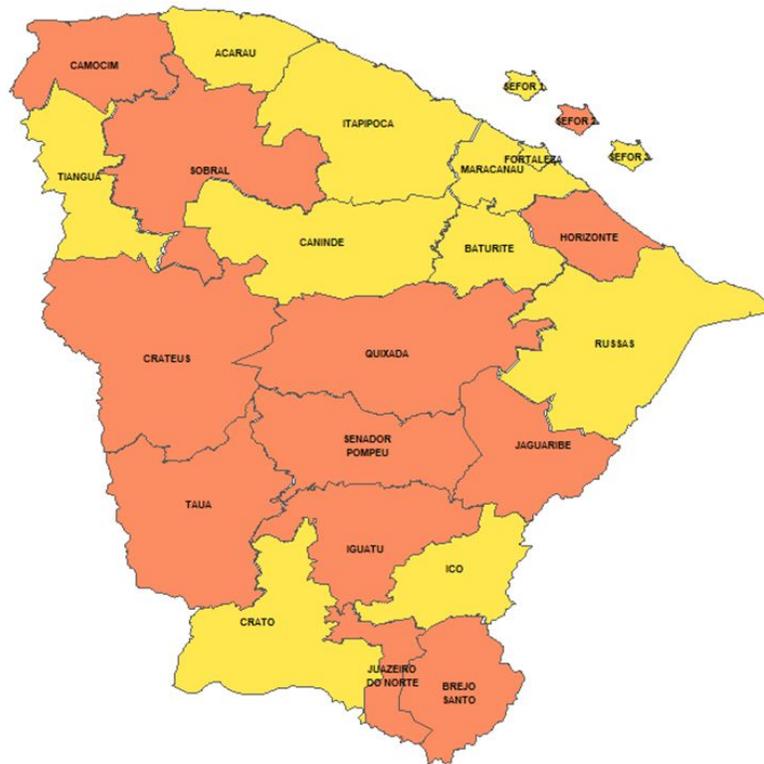
- Pontuação **menor que 4,3**: Indica que o diretor **não prioriza** a dimensão
- Pontuação **maior que 10,7**: Indica que o diretor **prioriza muito** a dimensão
- Esses valores consideram um nível de significância de **5%**



# EXEMPLO DE RESULTADOS POR REGIONAL

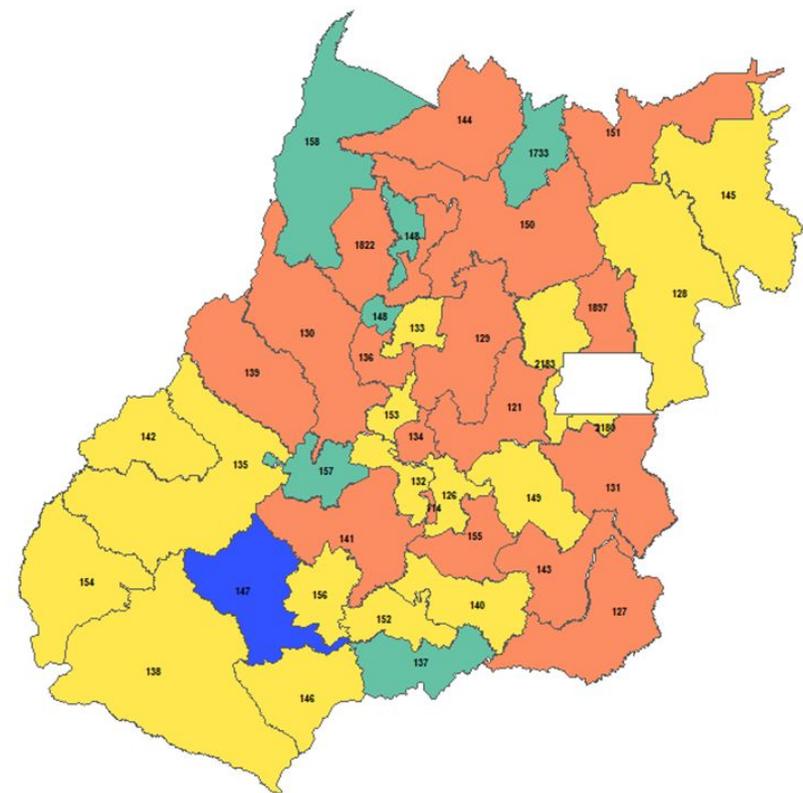
## Ceará

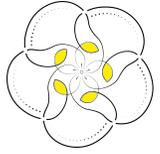
Perfil ■ Pessoal e Relacional ■ Pedagógica ■ Administrativo-Financeira ■ Político-Institucional



## Goiás

Perfil ■ Pessoal e Relacional ■ Pedagógica ■ Administrativo-Financeira ■ Político-Institucional





# SIAREM

Seminário Integrado de Avaliação das  
Redes Públicas Estaduais e Municipais

# Obrigado!

JUNTOS PELO DIREITO DE APRENDER



[siarem.caeddigital.net](http://siarem.caeddigital.net)

