



TÉCNICAS ESTATÍSTICAS NA DETECÇÃO DE FRAUDES EM EXAMES COM BASE NAS POLITICAS EDUCACIONAIS

Ana Letícia Alves Cordeiro¹
Rômulo Barbosa Veloso²

RESUMO

O sistema educacional assim como nossa sociedade vive em constantes mudanças que muitas vezes são influenciados por fatores ou agentes externos, dessa forma, os métodos de avaliações utilizadas na educação são relacionados com políticas públicas educacionais. Entretanto, muitas vezes, as provas são assimiladas apenas como um procedimento de classificação ou seleção, criando um ambiente competitivo e ocasionando pressão por parte dos alunos para obter um bom resultado. Atualmente, temos avaliações aplicadas em grande escala, (Avaliações Externas), com intuito de subsidiar e implementar políticas públicas para a área da educação, verificando então o desempenho e qualidade do ensino. Nesse contexto, este trabalho surgiu com o objetivo de detectar indicadores de fraudes em avaliações, com a utilização de técnicas estatísticas. O objetivo desse estudo é comparar resultados obtidos em avaliações matemáticas através de simulações. As técnicas utilizadas são provenientes da Teoria Clássica de Testes (TCT).

Palavras-chave: Políticas Educacionais, Avaliações, Fraudes, TCT.

INTRODUÇÃO

Nossa vida social assim com a escolar vive em constante mudança, a pandemia que estamos vivendo é um exemplo disso, e para isso é necessário que aja possibilidades de da continuidade as tarefas e projetos que vinham sendo desenvolvidos. O papel das políticas pública vem de encontro com o que estamos vivendo e o que possamos enfrentar futuramente, pois possui um grande papel para garantir os nossos direitos, assim como as políticas educacionais, que devem se responsabilizar para que todos tenham direito a educação.

Dessa forma, a política educacional pode ser entendida como um meio de criar valor e conhecimento, permitindo que os alunos se desenvolvam de maneira ampla,

¹Pós Graduando do Programa de Pós Graduação em Modelagem Computacional e Sistemas da Universidade Estadual de Montes Claros – UNIMONTES-MG, lets1466@gmail.com;

² Professor: Doutor, Universidade Estadual de Montes Claros - MG,romulo.veloso@unimontes.br.



incluindo sua capacidade de se comunicar, fundamentar suas idéias e entender o mundo ao seu redor.

De acordo com alguns estudos, percebemos que as políticas de responsabilização educacionais formuladas para promover o progresso educacional geralmente propendem efeitos colaterais imprevisíveis em sua elaboração, pois essas políticas estão relacionadas de forma direta ao uso cada vez mais frequente de sistemas externos de avaliação e à confiança nesses instrumentos.

É notório que as provas são um dos métodos de avaliação mais utilizados no processo de ensino-aprendizagem, e uma importante ferramenta se for bem desenvolvida, analisando e colocando em prática seu verdadeiro significado.

A Prova Brasil, o Sistema Nacional de Avaliação da Educação Básica (Saeb), SARESP e ENEM são exemplos de avaliações diagnósticas que possui o objetivo de avaliar a qualidade do ensino através de testes padronizados, obtendo resultados que são usados para comparação entres redes e escolas. Esses resultados são encontrados em um indicador, o Índice de Desenvolvimento da Educação Básica (Ideb) que foi criado em 2007, calculado a partir dos dados obtidos no Censo Escolar.

Neste contexto e fundamentados nesses resultados divulgados pelo Idep, podemos levantar hipóteses de que essas comparações, de fato podem pressionar tantos os alunos quantos os professores, visto que ambas as partes estarão sendo avaliadas.

Habitualmente presenciamos no cotidiano escolar, que certa pontuação é atribuída a cada aluno por meios de trabalhos e provas, e o que a define é a quantidade de acertos obtidos. O que pode gerar falhas, quando não é considerados indícios de acertos ao acaso, ou padrões suspeitos.

Neste cenário, o objetivo desse trabalho consiste em comparar resultados obtidos em avaliações matemáticas através de uma simulação a partir de um conjunto de dados gerado pela distribuição binomial e compararmos com intuito de analisar se há indicativos de fraudes ou não, baseados nos índices de similaridades.

As técnicas utilizadas neste estudo são provenientes da Teoria Clássica de Testes (TCT) e da Teoria da Resposta ao Item (TRI). A TCT vem de encontro com a ideia apresentada neste trabalho, pois a resposta de um candidato pode ser comparada com a de outro que fez a mesma avaliação, já que a teoria baseia-se no percentual de acertos em relação ao total de itens, através de técnicas estatísticas. A Teoria da Resposta ao Item (TRI) permite analisar o grau de dificuldade dos itens e diferenciar os alunos que possui ou não



habilidades e conhecimentos requisitados além da possibilidade de acerto ao acaso.

Assim, proponho utilizar a TCT com auxílio do software R e da IDE Rstudio, para analisar através de uma simulação, os resultados obtidos e dessa forma detectar se há indicativos de fraudes.

METODOLOGIA

A metodologia deste trabalho consiste em criar uma simulação, onde a partir de um conjunto de dados possamos encontrar indicativos de fraudes em uma avaliação através da análise dos gabaritos dos alunos.

Para esse fim, utilizamos o software RStudio. O conjunto de dados foi gerado a partir da distribuição binomial tomando uma probabilidade de p (sucesso) igual a 0,5. Para gerar os valores aleatoriamente (pseudo-aleatoriamente) foi utilizada uma função `rbinom` da Biblioteca do R: R Core Team (2020). R: Uma linguagem e ambiente para computação estatística. R Foundation for Statistical Computing, Viena, Áustria. URL <https://www.R-project.org/>.

A simulação consiste em comparar os gabaritos de 03alunos, que fizeram a mesma avaliação contendo 20 questões. A primeira comparação é feita entre o candidato 26 e o candidato 30 no qual é possível identificar de início algumas semelhanças nas respostas dadas por ambos, e posteriormente é feita uma comparação entre o candidato 26 e o candidato 33 que foram escolhidos aleatoriamente.

Na próxima etapa, são apresentados alguns resultados onde são exibidos os números exatos de erros, acertos e o valor de estimativa de máxima verossimilhança (MLE). A seguir é dado um Nível Alfa, onde verificamos os índices usados para analisar as comparações feitas e é apresentado se há ou não sinalização de indicativos de fraudes de acordo com cada valor atribuído a P (sucesso): 0.05, 0.01, 0.001.

REFERENCIAL TEÓRICO

Assuntos que envolvam as políticas públicas e políticas de responsabilização vêm sendo cada vez mais debatidas em pesquisas e noticiários, muitas delas com intuito de promover a qualidade dos processos educacionais. Neste estudo, podemos reconhecer uma exemplificação de políticas de responsabilização:



Como exemplo de políticas de responsabilização educacional estão a bonificação docente, em função do atingimento de metas, verificando por meio de avaliações externas, a premiação de alunos ou escolas com destacado desempenho positivo nas avaliações padronizadas, ou a aplicação de sanções e intervenções em escolas que não tenham atingindo metas determinadas, como é o caso em alguns sistemas educacionais norte-americanos. (CAPOCCHI,2017,p. 27)

De fato, o incentivo oferecido por essas políticas pode gerar impactos inesperados no processo educacional. Segundo (CARRASQUEIRA, 2015, p.78) as escolas, os professores e funcionários não são os únicos responsáveis pelos desempenhos dos alunos nessas avaliações em larga escala, mas diversos agentes influenciam no processo de alcançar certas metas e padrões, tendo em vistas os benefícios que poderão ser obtidos ou intervenções caso tenham um rendimento inferior.

No Brasil, conforme o estudo de Brooke (2011, p.46), alguns exemplos de políticas são citados como: Nova Escola (Rio de Janeiro), o sistema de bônus salarial (São Paulo), o prêmio Escola do Novo Milênio (Ceará), e o bônus salarial para professores alfabetizadores (Sobral/Ceará). Um fator em comum dessas políticas é a caracterização de premiação coletiva para profissionais da escola com base no desempenho médio dos alunos no ano anterior, o que nos gera um questionamento: As escolas que apresentaram desempenhos inferiores continuamente nos anos anteriores poderiam apresentar resultados significativamente bons após a criação dessas políticas?

Na pesquisa de Capocchi (2017, p.47) são apresentados alguns efeitos colaterais resultados de políticas educacionais, dentre elas estão à exclusão planejada de alunos com baixa proficiência, fraudes docentes nas quais as regras de aplicação das provas não são realizadas corretamente, e o “ensinar para o teste” termo utilizado pelo autor para se referir quando o professor passa a concentrar maior parte do tempo das aulas em assuntos relacionados aos temas cobrados nestas avaliações, levando em consideração esses efeitos indesejados é de se esperar que o rendimento seja consideravelmente bons em relação as avaliações dos anos anteriores.

Neste âmbito, levando em consideração esses efeitos causados pelas políticas publicas, além da TCT utilizamos um software importante que nos permitiu retirar importantes informações e resultados da simulação feita. O R é um software livre e de



código aberto que nos permitem a partir de programações efetuar testes de uma maneira geral e em especial testes estatísticos enquanto o RStudio é uma IDE (acrônimo em inglês de Interface de Ambiente Integrado) extremamente facilitadora de trabalhos com o R.

O nível alfa apresentado nos resultados é composto pelos índices: W, GBT e S₂ que são baseados nos números de respostas idênticas, K, K₁, K₂ e S₁ que são baseados nos números de respostas incorretas idênticas.

RESULTADOS E DISCUSSÃO

Através do código construído no R, simulamos com apoio do EXEL uma amostra de alunos que fizeram uma avaliação contendo 20 questões. Os gabaritos dos candidatos foram os seguintes: (onde 1 significa resposta correta e 0 significa resposta errada).

Tabela 1 Resposta do candidato 26

1	1	0	1	1	1	0	0	1	0	1	1	1	0	1	1	1	1	0	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Tabela 2 Resposta do candidato 30

1	1	0	1	1	1	0	0	1	0	1	1	1	0	1	1	1	1	0	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Tabela 3 Resposta do candidato 33

0	1	1	1	1	0	0	0	1	0	1	0	0	0	1	0	1	0	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Em seguida foram analisados os gabaritos de dois candidatos que apresentaram respostas suspeitas (candidatos 26 e 30), e dois candidatos escolhido aleatoriamente (candidatos 26 e 33) e obtemos os seguintes valores para os diversos índices:

Tabela 4: Comparação entre os candidatos 26 e 30

Nível Alfa			
Índice	P= 0,05	P= 0,01	P= 0,001
W	SIGNIFICANTE	SIGNIFICANTE	SIGNIFICANTE
GBT	SIGNIFICANTE	SIGNIFICANTE	SIGNIFICANTE
K	SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE
K1	SIGNIFICANTE	SIGNIFICANTE	NÃO SIGNIFICANTE
K2	SIGNIFICANTE	SIGNIFICANTE	NÃO SIGNIFICANTE
S1	SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE
S2	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE

Tabela 5: Comparação entre os candidatos 26 e 33

Nível Alfa			
Índice	P= 0,05	P= 0,01	P= 0,001
W	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE
GBT	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE
K	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE
K1	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE
K2	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE
S1	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE
S2	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE	NÃO SIGNIFICANTE

O nível alfa é composto por alguns índices que de acordo com a variação do valor de P apresenta se é significativo ou não indícios de fraudes.

O índice W exibe as probabilidades estimadas de resposta correta de cada item. Na comparação dos resultados dos candidatos 26 e 30, os candidatos apresentaram o mesmo de número de respostas corretas e incorretas idênticas. O número esperado de



respostas idênticas é igual a 11,472 e desvio padrão do número esperado de respostas idênticas igual a 2,188, sendo assim o valor de W é dado por 3,8967 e a probabilidade dos gabaritos serem iguais existe num nível de significância igual a $5e-05$ de acordo com os valores P , sendo assim, através do índice W é apresentada sinalização de fraudes.

Já na comparação dos candidatos 26 e 33, o número esperado de respostas idênticas é igual 10,806, o desvio padrão do número esperado de respostas idênticas igual a 2,188, e o valor do índice W é dado por 0,5458, logo a probabilidade de indicativo de fraudes existe em um nível de significância de 0,2926, portanto não podemos afirmar que seja suspeita, logo não é significativo indícios de fraudes.

O índice GBT que representa o Teste da Binomial Generalizada sinaliza indicativos de fraudes quando feita a comparação entre os candidatos 26 e 30, este teste reconhece se os gabaritos de resposta dados pelos candidatos apresentam semelhanças ou não, através da distribuição binomial composta. Em seus resultados é possível observar a probabilidade de correspondência em cada item da avaliação e a sua distribuição de probabilidade exata para o número de correspondências. Sendo assim, entre os candidatos 26 e 30 é dada uma probabilidade de observar 20 correspondências, que permite identificar sinalização de fraude. Prontamente, quando feita a análise entre os candidatos 26 e 33, é dada uma probabilidade de 31,58% de observar 12 correspondências, não sinalizando indicativos de fraudes.

O índice S_2 é baseado no número de respostas idênticas, por exemplo, os candidatos 26 e 30 apresentaram o número de respostas incorretas idênticas igual a 6 e o número de respostas corretas idênticas igual a 14, ou seja o total de número de respostas idênticas foram 20 questões. Já os candidatos 26 e 33 tiveram o número de respostas incorretas idênticas igual a 4 e o número de respostas corretas idênticas igual a 8, somando um total de 12 respostas idênticas. O índice S_2 pode ser entendido como a amplificação do índice S_1 , que conta o número de respostas idênticas entre o candidato que realmente fez a avaliação e o que copiou, baseado em uma variável aleatória. Porém ao verificar apenas as respostas incorretas, é eliminado a chance de que as respostas corretas também possam não serem respondidas pelos próprios candidatos. Desta maneira os índices K , K_1 e S_1 podem não sinalizar quando um candidato copia as respostas corretas. Sendo assim, de acordo com o índice S_2 , tanto na análise entre as



respostas dos candidatos 26 e 30 quanto às respostas dos candidatos 26 e 33, não é SIGNIFICANTE indicativos de fraudes.

CONSIDERAÇÕES FINAIS

Visto que as fraudes em exames e provas externas existem, não vemos muito material ou estudos que abordem esse tema de uma forma detalhada com ferramentas que permitem de fato analisar esses indícios, desta forma, este trabalho surgiu com interesse na detecção de indicativos de fraudes em exames e avaliações, levando em conta os efeitos colaterais surgido muitas vezes pelas políticas de responsabilização. A metodologia se deu através de uma simulação feita, onde foram analisados três gabaritos de uma amostra de candidatos que fizeram uma avaliação de 20 questões, depois de obtidos alguns resultados, com a ajuda do software R, a Teoria Clássica dos Testes e Teoria de Resposta ao Item, foi feita uma análise fundamentadas em índices baseados nas respostas dos candidatos, onde um candidato é suspeito de ser o copiador.

Os candidatos 26 e 30 apresentaram de fatos respostas idênticas e os índices W, e GBT sinalizaram uma probabilidade de indicativos de fraudes para todos os valores testados de P. Os índices K, K1, K2 e S1 também sinalizaram indicativos de fraudes quando $P=0,05$ e o índice S2 não sinalizou indícios de fraudes, para nenhum valor de P. Para os demais valores de P de acordo com cada índice, podemos verificar mais detalhadamente na Tabela 4, os índices que apresentarem significâncias ou não.

Logo, a respeito da análise feita entre os candidatos 26 e 33, que de início não apresentam semelhanças nas respostas, não foi possível identificar qualquer tipo de indicativos de fraudes, já que para todos os índices abordados e para todos os valores de P utilizados não apresentaram uma probabilidade significativa de que exista fraudes.

Este trabalho apresenta os primeiros resultados obtido com intuito na detecção de fraudes, exibindo alguns índices que permitiram essa análise.

De acordo com Capocchi (2017, p.168) as respostas para algumas questões quando se dizem a respeito de políticas de responsabilização, podem contribuir para que os efeitos colaterais causados na educação sejam minimizados. De fato, as bonificações oferecidas por essas políticas incentivam á um bom resultado e melhorias no ensino, porem como já discutido são encontrados implicações indesejáveis que nos deixam em duvida se houve melhorias ou se esse alcance de metas pode ser suspeito.



Com isso, abre a discussão sobre trabalhos e projetos que possam identificar quando essas implicações acontecem, e contribuir para diminuir esses efeitos colaterais.

REFERÊNCIAS

LOBO, G. M. O. A POLÍTICA DE AVALIAÇÃO E RESPONSABILIZAÇÃO EDUCACIONAL: EFEITOS NAS PRÁTICAS ESCOLARES. Didática e Prática de Ensino na relação com a Sociedade. EdUECE - Livro 3. Disponível em: <http://uece.br/endipe2014/ebooks/livro3/76%20A%20POL%C3%8DTICA%20DE%20AVALIA%C3%87%C3%83O%20E%20RESPONSABILIZA%C3%87%C3%83O%20EDUCACIONAL%20EFEITOS%20NAS%20PR%C3%81TICAS%20ESCOLARES.pdf>

Carrasqueira, K. *et al.* Políticas de responsabilização escolar: um estudo exploratório sobre as estratégias de instâncias intermediárias na rede municipal do Rio de Janeiro. JORNAL DE POLÍTICAS EDUCACIONAIS v.9, n.17 e 18 | Janeiro-Junho e Agosto-Dezembro de 2015 | PP. 75–89. Disponível em: <https://revistas.ufpr.br/jpe/article/view/41299>

SOARES, D. J. M. TEORIA CLÁSSICA DOS TESTES E TEORIA DE RESPOSTA AO ITEM APLICADAS EM UMA AVALIAÇÃO DE MATEMÁTICA BÁSICA. TESE (Dissertação de Mestrado) – Universidade Federal de Viçosa, VIÇOSA, MG, 2018. Disponível em : <https://www.locus.ufv.br/bitstream/handle/123456789/18404/texto%20completo.pdf?sequence=1&isAllowed=y>