



Um estudo de multitorreção com professores de matemática

Rafael Filipe Novôa **Vaz**
Instituto Federal do Rio de Janeiro
Brasil

rafael.vaz@ifrj.edu.br

Lilian **Nasser**

Universidade Federal do Rio de Janeiro
Brasil

lnasser.mat@gmail.com

Resumo

As provas ou exames se constituem como o principal instrumento avaliativo em matemática nos diferentes níveis de ensino. Os alunos são submetidos a exames para “aferir” o conhecimento, sendo que essa aferição está associada à crença de que o conhecimento de um indivíduo pode ser medido que os testes são instrumentos precisos, justos e imparciais, concepções positivistas. Os primeiros estudos de multitorreção realizados com licenciandos em matemática indicaram que a atribuição de notas não é tão objetiva. A mesma prova gerou notas distintas e com grande amplitude de resultados, ao ser corrigida por pessoas diferentes. Na fase seguinte, a investigação foi realizada com professores a fim de verificar se a experiência poderia de algum modo reduzir a amplitude de notas. Os resultados forneceram novos indícios do grau de subjetividade existente na correção de provas de matemática e colocaram em “xeque” a crença na precisão, justiça e imparcialidade.

Palavras chave: avaliação escolar, matemática, multitorreção, subjetividade.

Introdução

A avaliação está naturalmente presente em nossas ações diárias, avaliamos um produto em uma loja, avaliamos uma roupa, um carro e uma casa, avaliamos professores, pais, amigos e alunos. Avaliar algo ou alguém pode ser considerada uma característica intrínseca ao ser humano. No entanto, a avaliação escolar possui um caráter diferenciado ao assumir um papel social. Ao avaliar e emitir um resultado, representado por uma nota ou um conceito, o professor legitima um processo que “mede o conhecimento” dos estudantes. Classifica, julga, atribui um parecer de apto ou não apto a passar para o próximo nível de escolaridade. A avaliação escolar é programada, ocorre em um espaço artificial, está associada ao sentimento de dever cumprido, de obrigação e julgamento.

Nas aulas de matemática, é muito comum, pelo menos no Brasil, a utilização de testes (ou exames) como o principal e, às vezes, o único instrumento avaliativo. Buriasco, Ferreira & Ciani (2009) defendem que a avaliação escolar é composta por um rito e um mito. De acordo com essas pesquisadoras, “o rito de avaliar – aplicar uma prova ou um teste escrito e converter as resoluções e respostas de cada estudante a um valor numérico” (p.70) está associado ao mito de “medir e classificar de maneira precisa os alunos” (p.71).

Os exames escolares possuem, predominantemente, as seguintes características: são individuais, escritos, sem consulta, com tempo delimitado, esporádicos, intermitentes e breves, possibilitam a ausência de convivência como exigência para avaliar, oferecem tratamento genérico a todos os alunos com a pretensão de ser neutra (Raphael, 1998). Esse modelo avaliativo é fundamentado em uma visão positivista, em que é desprezada a subjetividade do professor que corrige a prova, que tem seus pilares na neutralidade e imparcialidade. Segundo Morgan (2000, p. 225-226), tais concepções estão presentes nos três pressupostos básicos que sustentam o pensamento geral na avaliação:

- presume-se que os indivíduos possuem atributos (como conhecimento, compreensão, habilidade, etc.) que são detectáveis e mensuráveis;
- o objetivo principal da avaliação é descobrir e medir esses atributos;
- o processo de avaliação e seus resultados são considerados fundamentalmente benignos ou mesmo benéficos (embora efeitos colaterais infelizes possam ser reconhecidos e tentativas feitas para melhorá-los).

Este trabalho relata uma investigação sobre a imparcialidade e subjetividade da correção de questões discursivas em provas de Matemática. Nas três fases da pesquisa são investigados dois aspectos da multcorreção. Por um lado, é questionada a existência de um efeito “*halo*” (Kahneman, 2012), em que o desempenho nas primeiras questões influencia o professor na atribuição de notas às demais questões. Por outro lado, é observada a ampla variação das notas atribuídas por grupos de licenciandos e de professores ao corrigir uma mesma prova. Os resultados apontam na direção de que a aplicação de testes não pode ser encarada como um instrumento justo e objetivo de avaliação.

Uma breve síntese da história das avaliações escolares

Guba e Lincoln (2011) apontam quatro gerações de avaliações escolares ao longo da história: geração da mensuração; geração da descrição e do juízo de valor, amparadas pelo paradigma científico e a avaliação de quarta geração, denominada por Fernandes (2000) como geração da negociação e construção.

Na primeira dessas concepções, ocorrida prioritariamente no início do século XX, a avaliação era entendida como uma medida e que, através de testes bem construídos, era possível medir com rigor e isenção as aprendizagens escolares dos alunos. Para Fernandes (2000), foi influenciada por alguns fatores históricos como a evolução da gestão científica na economia que procurava tornar o trabalho mais eficiente e os marcantes sucessos da Matemática e das Ciências Experimentais que ocorreram no final do século XVIII e no início do século XIX. Algumas das perspectivas mais marcantes dessa fase perduram até hoje, tais como as concepções de que: a avaliação é um instrumento para a classificação, seleção e certificação; que os conhecimentos são o único objeto da avaliação; os alunos não participam do processo avaliativo; a avaliação é

descontextualizada; a avaliação busca a objetividade e neutralidade do professor e é referida a uma norma ou padrão, por isso os resultados dos alunos são comparados.

A geração da descrição, nos anos 30 e 40, tentou superar algumas limitações da primeira. Enquanto na geração anterior o sistema educacional era avaliado a partir do desempenho dos estudantes, nesta perspectiva outros fatores foram, como por exemplo, os curriculares. A avaliação não se limitava a medir, ela passou a descrever até que ponto os alunos atingiam determinados objetivos. “A mensuração então deixou de ser tratada como equivalente à avaliação, sendo, porém, definida como um dos vários instrumentos que poderiam ser empregados a seu serviço” (Guba & Lincoln, 2011, p.35). Essa geração foi influenciada pelos estudos de Ralph Tyler, um pesquisador norte americano considerado o pai da avaliação educacional, que inovou ao propor objetivos comportamentais associados à avaliação e apresentava uma concepção diferenciada de currículo, caracterizado por “um conjunto ampliado e formulado de experiências formativas que ocorriam na escola” (Fernandes, 2000, p.47).

Nas décadas de 60 e 70 surge a terceira geração que, de modo análogo à segunda, tentou superar os pontos fracos das gerações anteriores. Os avaliadores acrescentaram à função descritiva das avaliações o papel de juízo de valor. É nessa geração, também denominada de Idade do Desenvolvimento, que os conceitos de avaliação escolar se tornam mais sofisticados. Surge em 1967, por intermédio de Michael Scriven, o conceito de “avaliação somativa, mais associada à prestação de contas, à certificação e à seleção, e o conceito de avaliação formativa, mais associada ao desenvolvimento, à melhoria das aprendizagens e à regulação dos processos de ensino e de aprendizagem” (Fernandes, 2000, p. 49). A geração da avaliação como juízo de valor é caracterizada por novas ideias, tais como as recomendações de que a avaliação deve induzir e/ou facilitar a aprendizagem; a coleta de informações deve ir além dos resultados dos testes; a avaliação deve envolver alunos, professores, pais, ... e os contextos de ensino devem ser considerados.

A quarta geração proposta por Guba e Lincoln, a avaliação como uma negociação e construção, constitui uma ruptura epistemológica. Essa nova concepção avaliativa, de raízes construtivistas, baseia-se em uma ideia de diálogo entre todos os pares do processo avaliativo. (como citado em Fernandes, 2000). Nesta concepção,

- os professores devem partilhar o poder de avaliar com os alunos e outros atores e devem utilizar uma variedade de estratégias e instrumentos avaliativos;
- a avaliação deve ser integrada no processo de ensino-aprendizagem;
- a avaliação formativa deve ser privilegiada em detrimento da avaliação somativa;
- o feedback se constitui como um indispensável elemento no processo;
- a avaliação assume uma função mais de suporte à aprendizagem do que de julgamento e classificação;
- a avaliação passa a ser interpretada como uma construção social do conhecimento;
- deve-se privilegiar métodos qualitativos aos métodos quantitativos. (Fernandes, 2000).

Por considerar que este método é, de fato, uma ruptura epistemológica, Fernandes (2000) alerta para as dificuldades e barreiras em uma possível implementação de tais concepções: “poderemos nos interrogar se os sistemas educativos e, mais concretamente, as pessoas que nele trabalham, aguentam certo tipo de modificações” (p.53). Talvez por esse motivo, as avaliações atuais ainda apresentam características predominantes das três primeiras gerações.

O paradigma psicométrico da avaliação (escolar)

Na predominância desse modelo positivista de avaliação escolar configura-se o que Fernandes (2009) denomina de *paradigma psicométrico de avaliação*. Segundo este autor, há uma tendência de a avaliação centrar-se mais nos resultados ou nos produtos do que no processo de aprendizagem. As três características principais do paradigma são:

é possível determinar exatamente o que os alunos sabem e são capazes de fazer;
as aprendizagens dos alunos constituem uma realidade que pode ser avaliada de forma objetiva, neutra e sem quaisquer inferências valorativas;
testes de naturezas diversas – cientificamente construídos e, como tal, objetivos e neutros – permitem a quantificação das aprendizagens dos alunos. (Fernandes, 2000, p.81-82)

Há alguns anos, diversos estudos realizados por pesquisadores, em lugares e em contextos distintos contestam a concepção positivista presente na avaliação escolar (Pacheco, 1998; Morgan, 2000; Hadji, 2001; Fernandes, 2009; Buriasco et al., 2009; Santos; Pinto, 2018; Vaz; Nasser, 2018).

via de regra, negligencia-se que o quantitativo advém do qualitativo, e, no caso da avaliação, a nota atribuída não emerge de maneira pura e unívoca dos instrumentos utilizados, mas é produzida pelo avaliador, que, para fazê-lo, pode se valer de instrumentos. Por fim, o rito de avaliar se constitui numa prática que confere uma validade ilusória ao mito da possibilidade do exercício da precisão e da justiça. (Buriasco et al, 2009, p.72)

Segundo Hadji (2001), o julgamento do avaliador é “sempre infiltrado por elementos provenientes do contexto escolar e social, desde a carga afetiva e a dimensão emocional devido à presença efetiva dos alunos” e, geralmente, “ignora que se baseia em parte em uma representação construída do aluno e em convicções íntimas que nada tem de científicas” (p.32). Para esse autor, a avaliação não é uma medida, porque “o avaliador não é um instrumento e porque o que é avaliado não é um objeto no sentido imediato” (p.34).

Fischer (2008) realizou uma pesquisa com professores universitários que atuam em cursos de licenciatura, a fim de investigar suas concepções em relação à avaliação. Neste trabalho, foi constatado que os professores formadores associam a objetividade na avaliação escolar à clareza, à uniformidade nos critérios de avaliação e à neutralidade no campo da matemática. A autora cita a fala ambígua de um professor que diz privilegiar “a exatidão do conteúdo, mas valoriza, na correção, o desenvolvimento apresentado pelo aluno” (p.85) e conclui que a pesquisa realizada

mostrou que os professores de matemática ainda refletem, em suas práticas avaliativas, muitas das concepções acerca do conhecimento matemático e de seu ensino construídas com base no discurso de modernidade. As características apontadas como constituintes do *habitus* desse professor, como a busca pela objetividade, a concepção positivista de rigor no trato dessa ciência e de seu ensino, um certo descrédito do fazer pedagógico e a adoção de uma postura pouco flexível, têm fortes marcas desse paradigma de ciência. (Fischer, 2008, p.96)

Mesmo com tantos estudos que contestam o paradigma psicométrico da avaliação escolar, as mudanças de tais concepções no meio escolar e no meio acadêmico parecem não ter se fortalecido a ponto de romper o forte bloqueio na crença nesse paradigma. Este trabalho foi inspirado no desafio de romper esse bloqueio.

A pesquisa de multitorreção

Esta investigação faz parte de um estudo de multitorreção em provas de matemática, que se iniciou em 2017 e está em sua terceira fase. Nas fases iniciais foi investigada a subjetividade presente na correção realizada por futuros professores (alunos de licenciatura em matemática). O objetivo específico na fase inicial da pesquisa, mais especificamente, foi investigar se, e como, o ordenamento das questões influenciaria a correção dos avaliadores. Na terceira fase o objeto de estudo passou a ser o professor de matemática formado e já atuante, com o objetivo de verificar a possível amplitude de notas atribuídas por professores em exercício na correção de uma mesma prova eram compatíveis com a grande amplitude observada na Fase 2. As três fases estão relacionadas ao objetivo geral de investigar o papel da subjetividade na correção de provas e exames.

Fase 1

O efeito *halo* consiste “no julgamento do todo a partir de características obtidas inicialmente e, se apresenta quando uma impressão é formada a partir de uma característica inicial influenciando múltiplos julgamentos ou classificações de fatores não relacionados” (Vaz & Nasser, 2018). Na correção de provas, o efeito halo induziria o professor a corrigir de modo distinto as provas de seus alunos a partir das respostas às primeiras questões, ou seja, caso o efeito halo aconteça, as resoluções das questões iniciais dos estudantes influenciarão o corretor durante a correção das questões seguintes.

A primeira investigação foi realizada com 40 alunos da Licenciatura em Matemática do IFRJ, uma instituição pública, federal, localizada no estado do Rio de Janeiro, Brasil sobre a ocorrência do efeito halo na correção de provas. Os estudantes foram divididos em dois grupos: A e B. Cada aluno corrigiu uma prova contendo 4 questões discursivas sobre a resolução de equações do 2º grau e atribuiu uma nota de 0 a 10. O valor de cada questão foi estipulado em 2,5, ficando a cargo de cada corretor atribuir outros valores a respostas parcialmente incorretas. Havia uma diferença entre o ordenamento das questões contidas nas duas provas: no teste A, a primeira questão continha uma solução correta enquanto que a última apresentava uma solução incorreta. No teste B, questões e soluções eram reposicionadas na ordem inversa, para verificar se os licenciandos atribuiriam notas semelhantes às provas dos tipos A e B. Em ambos os testes, as questões intermediárias apresentavam resoluções parcialmente corretas.

Os resultados desta fase sugeriram a existência do efeito halo. No teste A, a média, a moda e a mediana das notas foram menores do que no teste B. A conclusão desta fase é que, possivelmente, os futuros professores tentaram realizar uma compensação ao corrigir o teste B, em que a solução da primeira questão estava incorreta, sendo mais benevolente ao atribuir notas às questões seguintes.

Tabela 1. Medidas de tendência central - fase 1

	Teste A	Teste B
Média	6,43	7,17
Moda	6,25	7,25

Mediana	6,5	7,25
---------	-----	------

Fonte: Vaz & Nasser (2018).

Fase 2

A segunda fase da pesquisa utilizou dois campi do IFRJ, diversificando a amostra e ampliando-a ligeiramente (45 estudantes). O objetivo era retificar ou ratificar os resultados da fase 1 a partir de um novo teste com 5 questões, versando sobre áreas de figuras planas. Todos os avaliadores (licenciandos em Matemática) receberam um teste para ser corrigido a partir de um gabarito. De modo análogo, seria atribuída uma nota de zero a dez, cada questão “valendo” 2 pontos. Novamente a pontuação parcial ficou a cargo dos corretores.

Nesta fase não ficou configurada a existência do efeito *halo*. o entanto, a objetividade na correção pode ser questionada a partir da imensa amplitude entre as notas atribuídas. A mesma avaliação obteve desde a nota 2,5, por arbitrada por dois avaliadores, até a nota 7,0, estipulada por um terceiro avaliador. A figura 1 mostra a frequência das anotas atribuídas pelos licenciandos na fase 2 da pesquisa.

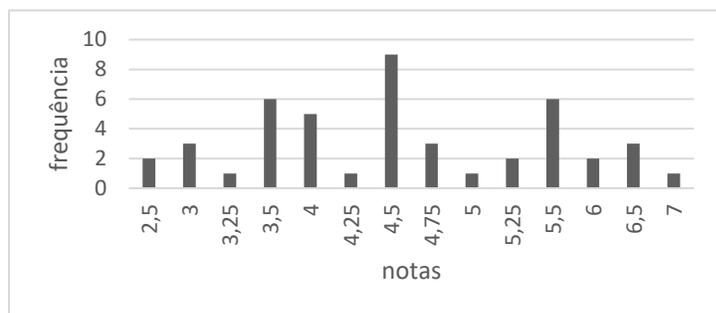


Figura 1. Notas atribuídas na fase 2.

Fonte: Vaz & Nasser. 2018

De modo geral, o que se constatou nas duas primeiras fases, em que os licenciandos corrigiram as provas, é que há um teor de subjetividade inerente ao corretor. Tal resultado, redirecionou a pesquisa ao seguinte questionamento: essa amplitude seria observada caso essa avaliação fosse corrigida por professores formados? Para responder a essa pergunta, iniciou-se a terceira fase: a investigação da multicorreção com professores de Matemática em exercício.

Fase 3

Com o objetivo de alcançar uma amostra de professores experientes, atuando em diversos tipos de escolas, nos diferentes níveis de ensino, a fase 3 foi aplicada a professores participantes de um projeto de extensão. Por facilidade de acesso, foi escolhido o Projeto Fundão, um grupo colaborativo de pesquisa e extensão da Universidade Federal do Rio de Janeiro.

Os professores participantes desse projeto foram convidados a corrigir o mesmo teste aplicado na fase 2. Cada um dos 14 professores presentes na data da aplicação recebeu uma cópia do teste respondido. A correção deveria ser realizada com amparo de um gabarito fornecido, com a informação de que cada questão valia 2,0 pontos, mas sem critérios para a atribuição de notas parciais.

Foi constatada uma redução da amplitude das notas do teste corrigido por professores (fase 3) se comparado ao mesmo teste corrigido por licenciandos (fase 2). Anteriormente, as notas

oscilaram entre 2,5 e 7,0, com uma variação de “4,5 pontos”. Já com os professores, a oscilação foi de “3,5 pontos”, de 4,0 a 7,5. Essa diferença é significativa. Supondo, por hipótese, que em uma escola a nota mínima exigida para aprovação é 6,0 (um padrão nas escolas brasileiras), esse estudante fictício seria habilitado para aprovação por 6 professores corretores, enquanto para os outros 8 professores, seu desempenho estaria aquém do desejado.

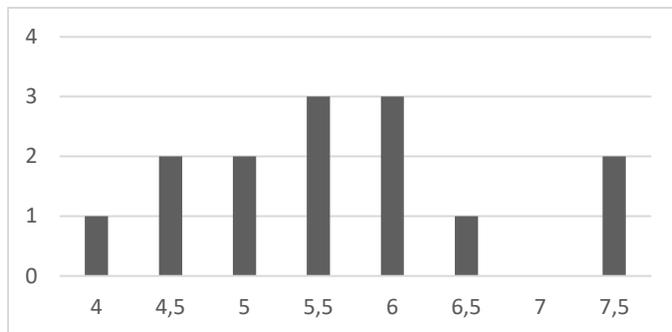


Figura 2. Notas atribuídas na fase 3

Fonte: os autores

A amplitude das notas atribuídas a uma mesma prova é um indício relevante do forte grau de subjetividade na correção de testes escolares. Mesmo considerando a hipótese que o conhecimento de alguém possa ser mensurado, que os professores sejam absolutamente neutros em sua atuação profissional e que todos os testes fossem construídos com embasamento científico de neutralidade e objetividade, a crença na possibilidade de usar o teste para “medir” de alguma forma o conhecimento também é questionada a partir desses resultados.

Uma das questões propostas para a correção apresentava uma solução de um estudante fictício, do nono ano do ensino fundamental, que comete um erro ao realizar uma divisão na segunda parte da solução (em três partes). Alguns professores atribuíram a esta questão pontuação de 1,8 ou valores próximos, enquanto outros professores atribuíram 1,0 ponto.

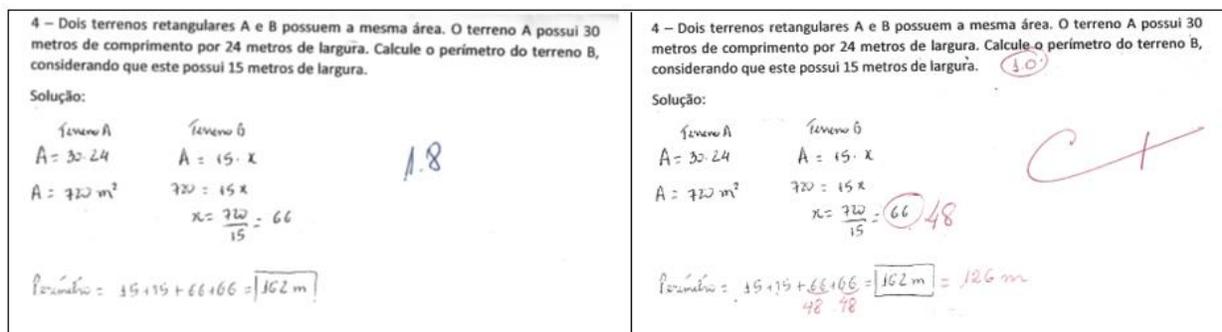


Figura 3. Comparativo de correção

Fonte: os autores

Observe que na solução da direita mesmo identificando que o estudante cometeu um único erro no procedimento ($720/15$), o professor atribui a metade da pontuação ao estudante (1,0), enquanto na solução da esquerda o professor desconta apenas 10% da pontuação na questão (1,8). Se a nota da prova se configura em uma medida, como justificar que o mesmo objeto pode ser associado a medidas distintas?

Considerações Finais

Os resultados descritos neste estudo evidenciam que há uma problemática na utilização de testes escolares: a correção. Corrigir o teste e atribuir a nota definitivamente não pode ser encarado como sinônimo de medir o conhecimento de alguém sobre algo, como se achava durante o início do século passado na geração da medida. (Guba & Lincoln, 2011). A ideia de associar a nota de um teste a uma medição é, definitivamente, um mito. O teste está muito mais próximo de uma leitura do que de uma medida. Uma leitura não somente da aprendizagem, mas também do ensino. A variação de notas atribuídas por licenciandos e professores ao corrigir uma mesma prova indica diferentes leituras, estruturadas na subjetividade do corretor.

A partir daí, é necessário que os testes devem ser ressignificados no meio escolar e acadêmico, abandonando o papel de único instrumento avaliativo (comumente utilizado na disciplina de Matemática). As avaliações escolares devem ser mais diversificadas, pois mais instrumentos podem oferecer uma melhor leitura do ensino-aprendizagem, respeitando mais as peculiaridades de cada estudante. A avaliação como uma negociação e construção, descrita por citada por Fernandes (2000), se constitui uma ruptura epistemológica difícil, mas necessária para modificar a filosofia positivista que acompanha as concepções sobre a avaliação escolar.

Referências y bibliografia

- Buriasco, R. L. C., Ferreira, P. E. A., & Ciani, A. B. (2009). Avaliação como prática de investigação (alguns apontamentos). *Boletim de Educação Matemática*, 22 (33), 69-96.
- Fernandes, D. (2009). *Avaliar para aprender: fundamentos, práticas e políticas*. São Paulo, BR: Unesp.
- Fischer, M. C. B. (2008) Os formadores de professores de matemática e suas práticas avaliativas. In: Valente, W. R. (Org.). *Avaliação em matemática: história e perspectivas atuais*. Campinas: Papirus. p. 75 -100.
- Guba, E. G., & Lincoln, Y. S. (2011) *Avaliação de quarta geração*. Campinas, BR: Editora Unicamp.
- Hadji, C. (2001). *Avaliação desmistificada*. Porto Alegre, BR: Artmed Editora.
- Kahneman, D. *Rápido e devagar: duas formas de pensar*. 1. ed. Rio de Janeiro: Objetiva, 2012. 608 p.
- Morgan, C. (2000). Better assessment in mathematics education? A social perspective. In: Boaler, J. (Org.). *Multiple Perspectives on Mathematics Teaching and Learning*. Westport, AblexPublishing, 225-242
- Pacheco, J. A. (1998). A avaliação da aprendizagem. In: Almeida, L. S.; Tavares, J. (Org.). *Conhecer, aprender, avaliar*. Porto: Porto Editora, 111-132.
- Santos, L., & Pinto, J. (2018). Ensino de conteúdos escolares: A avaliação como Fator estruturante. In: Veiga, F. H. (Org.). *O Ensino como fator de envolvimento numa escola para todos*. Lisboa: Climepsi Editores, 503-539.
- Vaz, R. F. N., Nasser, L. (2018). Avaliação em matemática: um estudo sobre multitorreção. In: Seminário Internacional de Pesquisa em Educação Matemática, 7., *Anais...Foz do Iguaçu*: SBEM.