

Novas possibilidades de avaliação em larga escala na educação básica através do uso de EDM e *Learning Analytics*

**Raphael Dourado¹, Rodrigo Lins Rodrigues², Jorge Cavalcanti³,
Alex Sandro Gomes¹**

¹Centro de Informática, Universidade Federal de Pernambuco (UFPE)
Av. Jornalista Aníbal Fernandes, s/n - Cidade Universitária - 50.740-560 - Recife – PE

²Departamento de Educação, Universidade Federal Rural de Pernambuco (UFRPE)
Rua Dom Manoel de Medeiros, s/n, Dois Irmãos - 52171-900 - Recife – PE

³CECOMP, Universidade Federal do Vale do São Francisco (UNIVASF)
Av. José de Sá Maniçoba, S/N - Centro CEP: 56304-917 – Petrolina – PE
rasd2@cin.ufpe.br, rlr@ded.ufrpe.br, jorge.cavalcanti@univasf.edu.br,
asg@cin.ufpe.br

Abstract. *In Brazil, the main strategy for basic education assessment consists in standardized tests, especially those applied by the Basic Education Assessment System (SAEB). However, many researchers criticize the use of this strategy as the only means to diagnose the quality of Brazilian public schools, pointing out, among other limitations, its sparse periodicity. In this context, this paper proposes as a challenge the conception and validation of new evaluation mechanisms, able to provide information about students' performance in real time, through a computational approach based on Learning Analytics and Educational Data Mining (EDM) techniques.*

Resumo. *A estratégia de avaliação da educação básica no Brasil é hoje fortemente fundamentada em exames padronizados, especialmente os aplicados no âmbito do Sistema de Avaliação da Educação Básica (SAEB). No entanto, muitos pesquisadores criticam o uso desta estratégia como único meio para diagnóstico da qualidade das escolas públicas brasileiras, apontando, dentre outras limitações, a sua pontualidade. Nesse sentido, este artigo propõe como desafio a concepção e validação de novos mecanismos de avaliação, capazes de oferecer informações sobre o desempenho escolar dos alunos de forma contínua e em tempo real, por meio de uma abordagem computacional baseada em técnicas advindas das áreas de Learning Analytics e Mineração de Dados Educacionais (EDM).*

1. Introdução

Nos últimos anos, a estratégia de avaliação e acompanhamento da qualidade da educação básica no Brasil tem consistido majoritariamente na aplicação de exames padronizados em larga escala, com periodicidade anual ou bianual (GATTI, 2009). Este tipo de avaliação dos sistemas educacionais surgiu nos EUA a partir da década de 50,

com a aplicação de testes em larga escala, e tem como marco a publicação, em 1966, do Relatório Coleman, que pretendia estudar a segregação racial no sistema educacional norte-americano (FERNANDES e GREMAUD, 2009). A partir de então, os resultados deste tipo de avaliação tem sido utilizados em muitos países como um dos principais instrumentos de fomento à cultura de *accountability* nas escolas, isto é, “um tipo de política de responsabilização dos professores e gestores de escolas pelos resultados nas avaliações externas” (ALAVARSE, 2015).

Embora esta estratégia de avaliação tenha contribuído para um melhor entendimento das deficiências dos sistemas de ensino brasileiros (ALAVARSE et. al, 2013), diversos pesquisadores acreditam que mecanismos de avaliação complementares são necessários, principalmente voltados para o acompanhamento diagnóstico e longitudinal do desempenho dos discentes (Silva et. al., 2016; Esteban e Fetzner, 2015; Travitzki, 2013; Fernandes e Gremaud, 2009; Franco, 2001).

Nesse sentido, a crescente adoção de Ambientes Virtuais de Aprendizagem (AVAs) no ensino presencial – arranjo conhecido como *ensino híbrido*, ou *b-learning* – pode viabilizar novas possibilidades para a avaliação da aprendizagem (WELLS et. al., 2016). Estes ambientes são capazes de coletar e armazenar uma grande quantidade de dados relacionados ao processo de ensino-aprendizagem, os quais podem ser usados como insumos para a criação de mecanismos inovadores de avaliação e acompanhamento dos discentes.

A partir deste contexto, este artigo propõe como desafio a concepção e validação de novos mecanismos de avaliação em larga escala, capazes de oferecer a gestores e professores informações sobre o desempenho escolar dos alunos de forma contínua e em tempo real. Estes novos mecanismos fomentariam a adoção de uma cultura de avaliação formativa em larga escala nas escolas brasileiras, sanando um dos principais problemas do sistema de avaliação em larga escala atual: a falta de velocidade na detecção das dificuldades de aprendizagem dos discentes.

O restante deste artigo está organizado da seguinte forma: a Seção 2 descreve a história, características e limitações dos sistemas de avaliação em larga escala no Brasil, a Seção 3 faz um apanhado dos trabalhos recentes que utilizam uma abordagem computacional para auxiliar na avaliação da aprendizagem, a Seção 4 articula o desafio proposto e sugere uma metodologia para ataca-lo e na Seção 5 são traçadas algumas considerações finais.

2. História e cenário atual das avaliações em larga escala no Brasil

No Brasil, as primeiras iniciativas de avaliação dos sistemas de ensino iniciaram-se na década de 60, tendo como marcos o desenvolvimento de testes educacionais pelo Centro de Estudos de Testes e Pesquisas Psicológicas (CETPP, FGV-RJ) e também os estudos desenvolvidos pelo Programa de Estudos Conjuntos de Integração Econômica Latinoamericana (ECIEL) em meados dos anos 1970 (GATTI, 2009). Porém, é apenas no final da década de 80 que, motivado pelas altas taxas de repetência e evasão nas escolas públicas, o governo brasileiro iniciou os esforços, através do MEC, para institucionalizar um sistema de avaliação do rendimento escolar na educação básica.

Os baixos resultados obtidos nas primeiras experiências de avaliação levaram a um aumento no interesse pelos processos avaliativos, e culminou na criação do Sistema de Avaliação da Educação Básica (SAEB)¹. Atualmente, o SAEB é composto por três exames: a Avaliação Nacional da Educação Básica (Aneb), de caráter amostral, aplicada bianualmente a alunos do 5º e 9º ano do Ensino Fundamental e do 3º ano do Ensino Médio; a Avaliação Nacional do Rendimento Escolar (Anresc), também conhecida como “Prova Brasil”, de caráter censitário e realizada bianualmente envolvendo os alunos do 5º e 9º ano do Ensino Fundamental das escolas públicas; e a Avaliação Nacional da Alfabetização (ANA), também de caráter censitário e aplicada aos alunos do 3º ano do Ensino Fundamental, com o objetivo de avaliar os níveis de alfabetização e letramento em Língua Portuguesa e Matemática na rede pública de ensino. As provas do SAEB são construídas com base na Teoria de Resposta ao Item (TRI) e seus resultados são utilizados para compor o Índice de Desenvolvimento da Educação Básica (IDEB), principal indicador utilizado pelo governo brasileiro para avaliar a qualidade dos sistemas de ensino.

Além dos dados coletados pelo SAEB, os resultados do Exame Nacional do Ensino Médio (ENEM) — instituído em 1998 e adotado como critério de seleção para o ensino superior em 2009 — também têm sido utilizados para geração de indicadores da qualidade do ensino básico. Os exames do ENEM, assim como os do SAEB, são construídos atualmente de acordo com a Teoria da Resposta ao Item, porém avaliam os estudantes em cinco áreas: Ciências da Natureza, Ciências Humanas, Linguagem, Matemática e Redação.

No entanto, muitos pesquisadores têm apontado as limitações do uso de exames padronizados como única estratégia de diagnóstico da qualidade das escolas públicas brasileiras. Silva et. al. (2016) argumenta que “[as avaliações em larga escala] têm contribuído para essa perda de autonomia escolar, na medida em que essas avaliações não priorizam os processos, mas somente o produto final.” Esteban e Fetzner (2015) criticam o estreitamento curricular e redução do papel da escola provocado pelos exames padronizados, argumentando que “A escola é ocupada por treinamentos para a prova, com exercícios que não apresentam a discussão de questões que possam ser consideradas interessantes pelos alunos ou relevantes para a comunidade” e concluem que “O bom desempenho nas provas transforma-se em objetivo e fim da escola”, visão compartilhada também por Fernandes e Gremaud (2009), que apontam ainda o risco de *gaming*, dado que “as escolas podem adotar estratégias para alterar os resultados, mas que não mudam a qualidade do ensino ministrado como, por exemplo, treinar e motivar os estudantes para os testes ou excluir dos exames alunos de baixa proficiência”.

Franco (2001) aponta ainda outras duas limitações do SAEB: a insuficiência de suas medidas para auxiliar na compreensão de quais fatores influenciam no desempenho dos alunos e a ausência de dados longitudinais, os quais poderiam possibilitar uma avaliação mais abrangente e o acompanhamento do desenvolvimento dos alunos. Lee (2010) também enfatiza a importância dos dados longitudinais para o diagnóstico da qualidade do ensino, e cita o projeto GERES (FRANCO et. al., 2015) como exemplo de iniciativa bem sucedida nesse sentido. Em relação ao ENEM, Travitzki (2013) faz um

¹ <http://portal.inep.gov.br/web/guest/educacao-basica/saeb>

estudo abrangente sobre este exame e alerta que “a supervalorização desses testes cria o risco de colonização do cotidiano escolar pela razão instrumental, empobrecendo as relações intersubjetivas e as práticas pedagógicas”.

Com base nas limitações expostas nesta seção, é possível perceber que há uma carência por mecanismos de avaliação em larga escala que sejam contínuos, abrangentes e diversificados. A próxima seção descreve trabalhos recentes da comunidade de pesquisa em Ciência da Computação que se propõem a resolver problemas semelhantes.

3. Contribuições da computação para avaliação da aprendizagem

Nos últimos anos, os Ambientes Virtuais de Aprendizagem (AVAs) têm sido cada vez mais usados no ensino presencial, dando origem à modalidade conhecida como *ensino híbrido* ou, em inglês, *Blended Learning* ou *b-learning* (BONK e GRANHAM, 2012). No *b-learning*, as melhores características do ensino presencial e à distância são combinadas de forma a utilizar a abordagem mais eficaz para cada momento do processo de ensino-aprendizagem.

Dada a capacidade dos AVAs de coletar, armazenar e disponibilizar grandes quantidades de dados estruturados relativos à interação do aluno com o ambiente, Carvalho et. al. (2014) enxerga na utilização destes ambientes, em contextos de ensino híbrido, um grande potencial para o desenvolvimento de novos mecanismos de avaliação e acompanhamento contínuo dos discentes. Nessa mesma direção, uma nova área de pesquisa da Ciência da Computação tem se consolidado nos últimos anos: a Mineração de Dados Educacionais (EDM, do inglês *Educational Data Mining*). A EDM surge ao mesmo tempo como uma ciência do aprendizado e também como uma rica área de aplicação para Mineração de Dados, dado a crescente disponibilidade de dados relativos a contextos educacionais, possibilitando decisões baseadas em evidência para aperfeiçoar as práticas e materiais de ensino (CALDERS e PECHENIZKIY, 2011).

Tendo em vista as possibilidades da EDM, Mislevy et. al. (2012) enxergam na interseção desta área com o Design Centrado em Evidência (ECD) e a Psicometria um caminho para novas formas de avaliação da aprendizagem, especialmente em larga escala. Os autores propõem uma estratégia para a concepção, desenvolvimento, implementação e validação de mecanismos de avaliação, citando como exemplos de aplicação bem sucedida o programa de treinamento *Cisco Networking Academy*. Já Sandeen (2013) acredita que os MOOCs (*Massive Online Open Courses*) estão promovendo avanços importantes nas estratégias de avaliação, e as novas técnicas que vem sendo desenvolvidas para estes ambientes serão naturalmente adotadas por outras modalidades de ensino.

Outra iniciativa relevante é apresentada por Pimentel e Omar (2008): um modelo para possibilitar a avaliação formativa de alunos em AVAs. Por meio do uso de EDM e índices de mensuração de capacidades e habilidades cognitivas como o KML, KMA (TOBIAS e EVERSON, 2002) e KMB (GAMA, 2004), os autores provaram ser possível o monitoramento cognitivo e metacognitivo dos discentes em um AVA.

Além da mineração de dados, outras técnicas têm sido usadas para apoiar as tarefas de avaliação e acompanhamento dos discentes. Otsuka et. al. (2002), por exemplo, mostram como a utilização de agentes de *software* pode auxiliar na promoção

de modelos de avaliação formativa em AVAs. Os autores descrevem a utilização desta técnica para incorporar mecanismos de avaliação formativa no ambiente TelEduc.

Como pode ser visto nesta seção, já existem iniciativas na comunidade de pesquisa em Computação que propõem novas estratégias de avaliação da aprendizagem amparadas nos dados coletados por AVAs. Na próxima seção, é proposto como desafio alavancar as iniciativas existentes no sentido de criar mecanismos de avaliação contínua e em larga escala para a educação básica.

4. Avaliação contínua em larga escala: um desafio para a computação

Visto a necessidade de mecanismos eficazes e diversificados para avaliação contínua em larga escala dos processos educacionais na educação básica, este artigo propõe como desafio a concepção e validação de tais mecanismos, por meio de uma abordagem computacional amparada no uso de técnicas advindas das áreas de *Learning Analytics* e Mineração de Dados Educacionais (EDM).

De forma a enfrentar o desafio proposto num período de dez anos, descreve-se a seguir três ações e como cada uma delas pode ser avaliada:

1. Ampliar a utilização de AVAs na educação básica, bem como adequar as práticas de ensino para utilizar estes ambientes de forma efetiva, no sentido de promover a modalidade de ensino híbrido. Além de ter potencial para trazer ganhos ao ensino-aprendizagem, o efetivo uso de estratégias de *b-learning* irá garantir a existência de uma massa de dados para suportar a criação de novos mecanismos de avaliação baseados em EDM e *Learning Analytics*. O sucesso desta ação pode ser mensurado a partir da quantidade de escolas utilizando adequadamente estratégias híbridas de ensino, assim como a percepção da comunidade escolar sobre os ganhos com esta mudança.
2. Utilizando-se dos dados disponíveis nos AVAs, desenvolver modelagens comportamentais capazes de medir as mesmas capacidades e habilidades avaliadas por exames como o SAEB e o ENEM, porém de forma longitudinal. Para avaliar o sucesso desta ação, propõe-se incorporar as modelagens desenvolvidas a um ou mais AVAs em uso na educação básica — na forma de um painel de acompanhamento (*dashboard*) — e verificar tanto sua precisão diagnóstica quanto sua contribuição para a melhoria do processo de ensino-aprendizagem. Uma das plataformas que podem ser usadas para este experimento é o Openredu², um AVA de código aberto atualmente em uso por 96 escolas públicas de educação básica em Recife/PE.
3. Investigar que competências, habilidades e fenômenos cognitivos e metacognitivos ainda não avaliados por exames como SAEB e ENEM que podem ser medidos explorando a riqueza de dados disponíveis nos AVAs — como interação entre pares em fóruns de discussão e *chats*; tempo gasto, número de tentativas e estratégias utilizadas na resolução de questões —

² OpenRedu – Rede Social Educacional. <http://www.openredu.org>

por meio de técnicas de EDM e *Learning Analytics*. A efetividade desta ação poderá ser avaliada da mesma forma que a anterior, ou seja, através da realização de experimentos com usuários reais utilizando AVAs em contextos de ensino híbrido.

5. Considerações Finais

Neste artigo, foi apresentada a estratégia atual de avaliação em larga escala da educação básica brasileira, suas características e limitações. Foram mostradas também algumas iniciativas na área da Ciência da Computação que propõem novos mecanismos de avaliação da aprendizagem, amparados principalmente em dados provenientes de AVAs. A partir deste contexto, foi proposto como desafio o desenvolvimento de novos mecanismos de avaliação capazes de oferecer a gestores e professores informações sobre o desempenho escolar dos alunos de forma contínua e em tempo real — e não apenas anualmente/bianualmente — possibilitando assim a tomada de ações pedagógicas imediatas para melhorar a qualidade do ensino e fomentar a cultura de avaliação formativa nas escolas.

6. Agradecimentos

Raphael Dourado é bolsista de doutorado do CNPq, processo nº 140973/2017-6. Alex Sandro Gomes é Pesquisador DTI Nível 2 CNPq nos processos 310.466/2012-1, 475634/2013-6 e 307202/2015-1. Esta pesquisa também é apoiada pela FACEPE APQ 0525-1.03/14.

Referências

- ALAVARSE, O. M., BRAVO, M. H., & MACHADO, C.. Avaliações externas e qualidade na educação básica: articulações e tendências. *Estudos Em Avaliação Educacional*, 24(54), 12–31, 2013.
- ALAVARSE, O. M. Avaliações externas e seus efeitos. *Anais CAEM, USP*, 2015.
- BONK, Curtis J.; GRAHAM, Charles R. *The handbook of blended learning: Global perspectives, local designs*. John Wiley & Sons, 2012.
- BOUSBIA, Nabila; BELAMRI, Idriss. Which Contribution Does EDM Provide to Computer-Based Learning Environments?. In: *Educational data mining*. Springer International Publishing, 2014. p. 3-28.
- CALDERS, Toon; PECHENIZKIY, Mykola. Introduction to the special section on educational data mining. *ACM SIGKDD Explorations Newsletter*, v. 13, n. 2, p. 3-6, 2012.
- CARVALHO, R. S et al. Discussão sobre as Tecnologias de Acompanhamento e Avaliação da Aprendizagem no *Blended Learning*. In: *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*. 2014. p. 223.
- ESTEBAN, Maria Teresa; FETZNER, Andréa Rosana. A redução da escola: a avaliação externa e o aprisionamento curricular. *Educar em Revista*, p. 75-92, 2015.

- FERNANDES, Reynaldo; GREMAUD, Amaury Patrick. Qualidade da educação: avaliação, indicadores e metas. Educação básica no Brasil: construindo o país do futuro. Rio de Janeiro: Elsevier, v. 1, p. 213-238, 2009.
- FRANCO, Creso. The Brazilian System of Assessment for Basic Education (SAEB): potential, problems and challenges. Revista Brasileira de Educação, n. 17, p. 127-133, 2001.
- FRANCO, Creso; BROOKE, Nigel; ALVES, Fátima. Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. Revista Ensaio: Avaliação e Políticas Públicas em Educação, v. 16, n. 61, p. 625-638, 2015.
- GATTI, B. Avaliação de sistemas educacionais no Brasil. Sísifo: Revista de Ciências Da Educação, (9), 7–18, 2009.
- GAMA, C. Towards a model of meta-cognition instruction in interactive learning environments [doctoral thesis]. Sussex, England: University of Sussex, 2004.
- LEE, V. E. Dados longitudinais em educação : um componente essencial da abordagem de valor agregado no que se refere à avaliação de desempenho escolar. Estudos Em Avaliação Educacional, 21(47), 531–542, 2010.
- MISLEVY, R. J., BEHRENS, J. T., DICERBO, K. E., & LEVY, R. Design and discovery in educational assessment: Evidence-centered design, psychometrics, and Educational Data Mining. Journal of Educational Data Mining, 4(1), 11–48 (2012).
- OTSUKA, J. L., VIEIRA, H., & ROCHA, D. (2002). Avaliação formativa em ambientes de EaD. XIII Simpósio Brasileiro de Informática Na Educação – SBIE – UNISINOS.
- SANDEEN, C. Assessment's Place in the New MOOC World. Research and Practice in Assessment, 8, p. 5–12, 2013.
- SILVA, Angela Maria Ferreira; LOPES, Pedro Isaac Ximenes; CASTRO, Alda Maria Duarte Araújo. Avaliação da educação no Brasil: a centralidade dos testes em larga escala. HOLOS, v. 7, p. 388-401, 2016.
- TOBIAS, S.; EVERSON, H.T. Knowing what you don't know: Further research on metacognitive knowledge monitoring [College Board Research Report 2002-3]. New York: College Entrance Examination Board (2002).
- TRAVITZKI, Rodrigo. ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar. 2013. Tese (Doutorado em Educação) - Faculdade de Educação, Universidade de São Paulo, São Paulo, 2013.
- WELLS, Craig S.; FAULKNER-BOND, Molly; HAMBLETON, Else. Educational Measurement: From Foundations to Future. Guilford Publications, 2016.