

Sistema Inteligente para Gestão e Correção de Avaliações Acadêmicas

**Fabiane Sorbar, Rafael H. Bordini, Flávio Rech Wagner, Jomi F. Hübner,
Daniel J. Souza, Matheus E. Giombelli, Victor B. Nischak**

¹Faculdade Donaduzzi, Avenida Max Planck, 3797 – Ed. Charles Darwin
85919-899 – Biopark – Toledo, PR – Brasil

fabiane.sorbar@bpkedu.com.br, R.Bordini@durham.ac.uk
jomi@inf.furb.br
daniel.souza@alunos.bpkedu.com.br
matheus.giombelli@alunos.bpkedu.com.br
victor.nischak@alunos.bpkedu.com.br

Abstract. *Manual grading of academic assessments creates teacher overload, rework, and delays in feedback to students, while the absence of centralized platforms limits question reuse and instrument standardization. This paper presents an integrated digital platform, under active development, for the creation, application, and correction of academic assessments, covering both multiple-choice and open-ended questions. The architecture is modular, comprising a question bank, test composition, correction, and reporting modules. Multiple-choice items are automatically graded through classical computer vision techniques applied to standardized answer sheets, whereas open-ended items are graded by the instructor through a dedicated interface within the system. This work focuses on the automated correction module for multiple-choice questions, situated at the intersection of Software Engineering and applications of Artificial Intelligence to education.*

Resumo. *A correção manual de avaliações acadêmicas gera sobrecarga docente, retrabalho e atraso na devolutiva aos estudantes, e a ausência de plataformas centralizadas limita o reaproveitamento de questões e a padronização dos instrumentos. Este trabalho apresenta uma plataforma digital integrada, em desenvolvimento ativo, para criação, aplicação e correção de avaliações acadêmicas, contemplando questões objetivas e dissertativas. A arquitetura é modular, organizada em torno de banco de questões, montagem de provas, correção e relatórios. As objetivas são corrigidas automaticamente por técnicas clássicas de visão computacional sobre cartões-resposta padronizados; as dissertativas são corrigidas pelo docente em interface dedicada do próprio sistema. O foco deste trabalho é o módulo de correção automatizada de questões objetivas, situado na interseção entre Engenharia de Software e aplicações de Inteligência Artificial à educação.*

1. Introdução

A avaliação da aprendizagem é componente central do ensino superior, exigindo instrumentos confiáveis e devolutiva ágil para orientar intervenções pedagógicas. Na Faculdade Donaduzzi, a correção manual de provas objetivas gera sobrecarga docente, retrabalho e atraso no retorno aos estudantes, ao passo que a ausência de uma plataforma centralizada

limita o reaproveitamento de questões e a padronização dos instrumentos avaliativos. Esses fatores motivam a investigação de soluções que aliem Engenharia de Software à correção automatizada por visão computacional, com vistas a reduzir o tempo gasto pelo docente sem comprometer a confiabilidade do processo.

2. Objetivo

O objetivo deste trabalho é desenvolver uma plataforma digital integrada para criação, aplicação e correção de avaliações acadêmicas no contexto da Faculdade Donaduzzi, contemplando questões objetivas e dissertativas. Como objetivos específicos: (i) construir um banco compartilhado de questões dos dois tipos; (ii) permitir a montagem de provas com cartões-resposta padronizados; (iii) implementar um módulo de correção automatizada para itens objetivos, complementado por interface dedicada para correção das dissertativas pelo docente; e (iv) avaliar a confiabilidade do módulo de correção automatizada em condições controladas e adversas.

3. Fundamentação Teórica

3.1. Avaliação da Aprendizagem e Sistemas de Avaliação Digital

O feedback imediato é fator essencial para a eficácia do processo avaliativo, orientando intervenções pedagógicas [Jurāne-Brėmane 2023]. Sistemas de avaliação digital têm se consolidado, com AVAs como o Moodle oferecendo recursos avaliativos e gestão de bancos de questões. Para a correção de itens objetivos, técnicas de reconhecimento óptico de marcas (OMR) são amplamente utilizadas, variando de processamento clássico de imagem a abordagens baseadas em aprendizado profundo [Mondal et al. 2024].

3.2. Engenharia de Software e Usabilidade em Sistemas Educacionais

Sistemas educacionais demandam arquitetura modular e atributos como usabilidade, escalabilidade e segurança. A usabilidade é especialmente crítica para adoção docente, sendo afetada por resistência à mudança e complexidade das interfaces [Jurāne-Brėmane 2023, Villamañe and Álvarez 2024].

3.3. Inteligência Artificial na Educação

A aplicação de IA na educação tem demonstrado ganhos em eficiência, personalização e análise de desempenho [Zawacki-Richter et al. 2019], incluindo geração automática de questões [Das et al. 2022]. Este trabalho foca na correção automatizada de questões objetivas por meio de visão computacional. Embora métodos baseados em redes neurais apresentem alto desempenho [Mondal et al. 2024], técnicas clássicas oferecem maior previsibilidade e menor custo computacional, sendo adequadas quando se prioriza robustez operacional.

4. Metodologia

Os requisitos foram levantados a partir da observação de problemas recorrentes pela equipe e de entrevistas com a docente responsável pelo projeto. Foram identificadas como principais dificuldades o tempo elevado de correção manual, a ausência de repositório centralizado, a limitada padronização dos instrumentos e a demora na devolutiva.

Esses pontos orientaram a definição dos requisitos funcionais — banco de questões, montagem de provas, correção automatizada, compartilhamento e geração de relatórios — e dos requisitos não funcionais, com ênfase em usabilidade, desempenho e segurança.

A arquitetura segue o padrão cliente-servidor, organizada em três camadas: apresentação, responsável pelas interfaces docentes; aplicação, composta por uma API REST que implementa a lógica de negócio; e persistência, destinada ao armazenamento de questões, avaliações, cartões corrigidos e dados de usuários. O sistema agrupa-se em quatro módulos: banco de questões, que abriga itens objetivos e dissertativos com criação, categorização e reutilização; avaliações, responsável pela composição das provas e pela geração de cartões-resposta padronizados; correção, que combina o componente automatizado para itens objetivos com interface dedicada para a correção das dissertativas pelo docente; e relatórios, que consolida os resultados em diferentes níveis de análise. Dos quatro módulos, o componente de correção automatizada de questões objetivas — foco deste trabalho — encontra-se em estado consolidado e foi submetido a avaliação experimental; os demais permanecem em desenvolvimento ativo, com partes funcionais coexistindo com partes ainda em implementação.

A camada de apresentação foi implementada em Flutter, para construção multiplataforma. A camada de aplicação utiliza o framework FastAPI, em Python, com API REST e validação de tipos; a persistência é realizada em PostgreSQL. O módulo de correção foi desenvolvido em Python com a biblioteca OpenCV e segue a tradição clássica de OMR [Nguyen et al. 2011], com três decisões de projeto que o diferenciam de implementações genéricas. Primeiro, como o próprio sistema gera o cartão, a posição de cada bolha é derivada de constantes geométricas do *template*, e não detectada via segmentação de contornos — estratégia robusta a bolhas saturadas pela marcação e a rasuras adjacentes. Segundo, a retificação da imagem é feita por homografia (RANSAC) sobre os dezesseis cantos dos quatro marcadores ArUco impressos nos vértices do cartão, o que torna a normalização geométrica invariante à orientação da foto e tolerante a inclinações expressivas. Terceiro, o limiar de classificação “bolha marcada” versus “bolha vazia” é calibrado por folha, a partir da distribuição de intensidades das bolhas amostradas, utilizando um *Gaussian Mixture Model* raso de dois componentes, com *fallback* determinístico via Otsu quando a separação bimodal não é suficiente; a saída por questão é classificada em estados distintos (ok, vazia, múltipla, ambigua) com índice de confiança associado.

5. Procedimento Experimental

A avaliação do módulo de correção foi conduzida com a pipeline em operação *standalone*, isolada do restante do sistema. Foram impressos cartões com até 20 questões objetivas, preenchidos com configurações variadas — bolhas únicas, vazias e múltiplas marcações intencionais — e fotografados com câmera de dispositivo móvel, totalizando 20 imagens. As condições de captura variaram em três eixos: iluminação ambiente normal e reduzida; incidência de sombra parcial, incluindo casos com sombra da mão do operador sobre a folha; e inclinação da câmera entre 35° e 40° em parte das amostras, aferida pelo giroscópio do dispositivo. As métricas registradas foram a classificação produzida para cada questão, comparada ao gabarito; o tempo total de execução por cartão, medido fim a fim; e a ocorrência de abortos da pipeline, acionados quando o erro de reprojeção da homografia excede o limite operacional ou quando algum dos quatro ArUcos não é detectado.

O hardware foi um *notebook*, com execução exclusivamente em CPU.

6. Resultados e Discussão

Em condições normais de iluminação e enquadramento, as 20 amostras foram corretamente classificadas em todas as questões, com tempo de execução por cartão tipicamente abaixo de 300 ms. A normalização via homografia sobre os dezesseis cantos dos marcadores ArUco mostrou-se eficaz em preservar a precisão da classificação nas amostras com inclinação entre 35° e 40° , intervalo só acessível porque a retificação corrige perspectiva projetiva — em contraste com *pipelines* OMR que aplicam apenas correção de *skew* planar via transformada de Hough [Nguyen et al. 2011] e, portanto, pressupõem capturas próximas ao alinhamento frontal. Sombras parciais leves não produziram degradação mensurável das classificações, comportamento atribuído à equalização adaptativa de histograma (CLAHE) no pré-processamento. Em três imagens com sombra mais pronunciada, uma linha por foto recaiu nos estados *múltipla* ou *ambigua*, sem que nenhuma resposta incorreta fosse classificada como *ok* — a ambiguidade foi sinalizada explicitamente em vez de propagar como acerto silencioso, consistente com o princípio de defesa em profundidade da pipeline.

Observou-se ainda um modo de falha característico em baixa iluminação: a detecção dos próprios marcadores ArUco compromete-se antes que erros de classificação de bolha apareçam, fazendo a pipeline abortar cedo, com diagnóstico, em vez de retornar resultado retificado de forma imprecisa.

Duas limitações da correção automatizada permanecem. A primeira é a restrição a itens objetivos: as dissertativas são hoje corrigidas pelo docente em interface dedicada do próprio sistema, com correção assistida por processamento de linguagem natural [Das et al. 2022] prevista como evolução. A segunda é a dependência da detectabilidade dos marcadores fiduciais, que motiva, como próxima frente, a orientação ativa do usuário durante a captura, com verificação prévia da legibilidade dos quatro ArUcos.

7. Conclusão

Este trabalho apresentou uma plataforma em desenvolvimento ativo para o ciclo completo de avaliações acadêmicas, com correção automatizada para itens objetivos e interface dedicada de correção pelo docente para os dissertativos. Os experimentos com 20 cartões em condições variadas confirmaram a viabilidade do módulo de correção automatizada, com classificação correta sob iluminação e enquadramento normais, robustez a inclinações de até 40° pela normalização via ArUco e tempo abaixo de 300 ms por cartão em CPU. Como evolução, prevê-se correção assistida por processamento de linguagem natural [Das et al. 2022] para dissertativas, geração assistida de questões e integração a AVAs institucionais.

Referências

- Das, B., Majumder, M., Phadikar, S., and Sekh, A. A. (2022). Automatic question generation and answer assessment for subjective examination. *Cognitive Systems Research*, 72:14–22.
- Jurâne-Brêmane, A. (2023). Digital assessment in technology-enriched education: Thematic review. *Education Sciences*, 13(5):522.

- Mondal, S., De, P., Malakar, S., and Sarkar, R. (2024). Omrnet: A lightweight deep learning model for optical mark recognition. *Multimedia Tools and Applications*, 83:14011–14045.
- Nguyen, T. D., Hoang Manh, Q., Bui Minh, P., Nguyen Thanh, L., and Hoang, T. M. (2011). Efficient and reliable camera based multiple-choice test grading system. In *International Conference on Advanced Technologies for Communications (ATC)*, pages 268–271.
- Villamañe, M. and Álvarez, A. (2024). Facilitating and automating usability testing of educational technologies. *Computer Applications in Engineering Education*, 32(3):e22725.
- Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16:39.