O USO DO NOTEBOOKLM E DE AGENTES AVALIADORES DE INTELIGÊNCIA ARTIFICIAL NA PRODUÇÃO TEXTUAL PARA O ENEM: UM ESTUDO COMPARATIVO NO ENSINO MÉDIO

Igor Alexandre Batista de Sá[1]
Deivison Gomes de Amorim[2]

INTRODUÇÃO

A expansão das tecnologias digitais alterou as formas de acesso, circulação e produção do conhecimento e passou a interpelar diretamente as práticas escolares de leitura e escrita. No campo da educação linguística, esse cenário não se reduz à incorporação de dispositivos: exige discutir como os estudantes localizam fontes, avaliam sua confiabilidade, relacionam perspectivas e transformam informação em conhecimento autoral. A emergência de sistemas de inteligência artificial generativa (IAG) intensifica essa discussão, pois oferece respostas fluentes e rápidas, mas nem sempre verificáveis, completas ou adequadas ao contexto de uso.

No Ensino Médio, a produção de textos dissertativo-argumentativos envolve operações complexas: compreender uma questão social, formular uma tese, selecionar repertórios pertinentes, organizar argumentos, construir relações coesivas e propor encaminhamentos compatíveis com os direitos humanos. No Exame Nacional do Ensino Médio (ENEM), essas dimensões são avaliadas por cinco competências, que abrangem desde o domínio da modalidade escrita formal até a elaboração de uma proposta de intervenção. A Cartilha do Participante explicita que não basta inserir referências ou conectivos: é necessário empregá-los de modo produtivo na construção de um projeto argumentativo coerente (BRASIL, 2024).

É nesse contexto que se insere o presente estudo, cujo foco é o uso articulado do NotebookLM, como ambiente de consulta e organização de materiais selecionados, e de agentes avaliadores configurados para oferecer devolutivas conforme a matriz do ENEM. A relevância pedagógica dessa combinação está na possibilidade de aproximar pesquisa, planejamento, produção e revisão textual em um ciclo mais frequente de feedback. Tal possibilidade, entretanto, deve ser examinada sem atribuir à tecnologia uma capacidade autônoma de ensinar ou de assegurar aprendizagem: a qualidade da experiência depende da curadoria das fontes, dos comandos utilizados, da participação ativa do estudante e da mediação do professor.

[1] Mestre em Letras pelo PROFLETRAS – Universidade de Pernambuco (UPE). E-mail: igoralexandrebatista@gmail.com;
[2] Mestre em Informática Aplicada – Universidade Federal Rural de Pernambuco (UFRPE). E-mail: deivison.gomes.amorim@gmail.com.

O objetivo geral é analisar as diferenças descritivas de desempenho entre dois grupos de estudantes do Ensino Médio submetidos a formas distintas de acompanhamento da produção textual, com atenção às médias obtidas nas cinco competências do ENEM. Busca-se, em particular, compreender em quais dimensões aparecem as maiores variações, discutir como a consulta orientada e a devolutiva automatizada podem apoiar o processo de escrita e identificar limites da intervenção observada. A pesquisa original informa a participação de 20 estudantes, sendo 11 no grupo experimental e nove no grupo de controle, com médias finais registradas em relatório de avaliação de redações.

O trabalho é apresentado como estudo aplicado de abordagem quali-quantitativa e delineamento comparativo com intervenção pedagógica. Como a documentação disponível não especifica elementos como duração exata, tema de cada redação, existência de avaliação diagnóstica inicial, critérios de alocação dos estudantes e procedimento de dupla correção, os resultados são tratados como descritivos e exploratórios, e não como prova de causalidade ou de significância estatística. A discussão articula os dados registrados com estudos sobre multiletramentos, avaliação formativa e diretrizes para o uso responsável de IA na educação. Ao final, sustenta-se que os achados são promissores, sobretudo para repertório e organização argumentativa, mas apontam a necessidade de novas investigações, especialmente quanto à proposta de intervenção e à autonomia crítica dos estudantes.

REFERENCIAL TEÓRICO

Escrita argumentativa, multiletramentos e curadoria de fontes

A escrita escolar pode ser compreendida como prática social que articula propósitos, interlocutores, gêneros e condições concretas de circulação, e não como simples exercício de aplicação de regras gramaticais. Essa perspectiva desloca a atenção do texto entendido como produto acabado para as decisões que o autor toma durante o planejamento, a textualização e a revisão. No gênero dissertativo-argumentativo exigido pelo ENEM, tais decisões tornam-se visíveis na relação entre recorte temático, tese, seleção de evidências, progressão dos argumentos e encaminhamento conclusivo.

A pedagogia dos multiletramentos amplia a noção de letramento ao considerar a diversidade cultural e a multiplicidade de modos, mídias e linguagens presentes na vida contemporânea. O New London Group (1996) propõe uma abordagem que associa participação em práticas sociais, construção de sentidos e análise crítica das condições de produção. Rojo e Moura (2012), ao discutirem a escola brasileira, destacam que o trabalho com textos e mídias

diversas pode ampliar repertórios, desde que não se limite ao uso instrumental das tecnologias. Para a produção de redações, isso implica ensinar o estudante a converter fontes diversas em argumentos pertinentes, em vez de apenas acumular citações ou informações.

O letramento digital, portanto, envolve habilidades de busca e operação, mas também julgamento: reconhecer autoria e propósito, confrontar dados, distinguir evidência de opinião e identificar lacunas ou vieses. Uma ferramenta de consulta apoiada em fontes pode favorecer o retorno a documentos e a organização de ideias, mas não elimina a necessidade de leitura direta. Resumos automáticos podem comprimir nuances, e a seleção inicial dos materiais delimita o horizonte de respostas. A mediação pedagógica deve tornar explícitos esses limites e convidar os estudantes a verificar afirmações relevantes nas fontes originais.

No estudo em questão, o NotebookLM foi empregado como ambiente para organizar artigos, relatórios e materiais de apoio e apoiar a síntese de informações temáticas. Essa função pode ser interpretada como suporte à curadoria e ao planejamento, não como substituição da leitura, da autoria ou do conhecimento disciplinar. Para que o repertório seja produtivo, é necessário que o estudante explique a relação entre a referência mobilizada e o problema debatido, avalie sua pertinência e a integre a uma linha de raciocínio própria. A distinção entre repertório legítimo e repertório meramente decorativo é decisiva para o desempenho argumentativo.

Avaliação formativa, revisão e feedback mediado por IA

A avaliação formativa assume função pedagógica quando as informações sobre o desempenho são convertidas em decisões de ensino e aprendizagem. Em vez de encerrar a atividade com uma nota, a devolutiva deve tornar compreensível o que já atende aos critérios, qual aspecto demanda revisão e que ação pode aproximar o estudante do objetivo. Hattie e Timperley (2007) observam que o feedback pode favorecer ou prejudicar a aprendizagem conforme o foco e a forma de apresentação; sua eficácia depende de oferecer informação utilizável sobre a tarefa e sobre os próximos passos, e não apenas julgamentos genéricos.

Na escrita, esse princípio se traduz em ciclos de planejamento, produção, leitura crítica, revisão e reescrita. Uma devolutiva organizada pelas competências do ENEM pode tornar os critérios mais visíveis e ajudar o estudante a localizar fragilidades específicas, como tese pouco delimitada, progressão insuficiente ou proposta de intervenção incompleta. Ainda assim, a classificação automática de um trecho não equivale à compreensão do texto em sua totalidade. É necessário que o estudante interprete o comentário, decida se ele procede e realize as alterações sob sua responsabilidade autoral.

Os agentes avaliadores utilizados no grupo experimental foram configurados para analisar os textos segundo as cinco competências. Em tese, a rapidez do retorno pode ampliar as oportunidades de revisão e liberar tempo docente para acompanhamento mais focalizado. Contudo, a automação pode reproduzir inconsistências, apresentar avaliações excessivamente seguras ou não captar ambiguidades, ironia, repertórios locais e a relação entre partes do texto. Por isso, a devolutiva deve ser considerada hipótese diagnóstica, sujeita à validação do professor, e não decisão final sobre a qualidade do texto ou sobre a nota do estudante.

Essa cautela é compatível com as orientações da UNESCO para IA generativa em educação e pesquisa, que propõem uma abordagem centrada no ser humano, com atenção à privacidade, à validação pedagógica, à inclusão e à agência dos aprendizes (MIAO; HOLMES, 2023). Em contexto escolar, a responsabilidade não se restringe a evitar erros factuais: envolve também discutir o que se compartilha com plataformas, quais fontes sustentam as respostas e como preservar a autoria. O professor permanece responsável por contextualizar critérios, identificar necessidades e garantir que a tecnologia esteja subordinada a objetivos formativos.

A matriz do ENEM como referência de ensino

A matriz de redação do ENEM oferece uma linguagem comum para organizar o ensino e o feedback. A Competência 1 trata da modalidade escrita formal; a Competência 2 articula compreensão do tema, atendimento ao gênero e mobilização de conhecimentos; a Competência 3 observa a seleção e organização de informações e argumentos; a Competência 4 incide sobre mecanismos de coesão; e a Competência 5 considera a elaboração de proposta de intervenção relacionada ao problema, com respeito aos direitos humanos (BRASIL, 2024). Essas dimensões se inter-relacionam: um repertório só contribui para a argumentação quando integrado ao projeto de texto, e a proposta final depende da leitura adequada do problema desenvolvido.

Usar a matriz como ferramenta didática exige evitar que a produção se transforme em preenchimento mecânico de uma lista de requisitos. O ensino precisa mostrar como cada competência participa da construção de sentido e como escolhas textuais podem atender a mais de uma dimensão simultaneamente. O feedback automatizado pode funcionar como uma oportunidade de tornar os critérios explícitos, desde que não induza à produção padronizada ou à busca de pontuação desvinculada da reflexão crítica.

METODOLOGIA

A investigação caracteriza-se como pesquisa aplicada, de abordagem quali-quantitativa, desenvolvida no contexto do Ensino Médio. Considerando que o material de origem descreve

dois grupos submetidos a modalidades diferentes de acompanhamento, adota-se neste texto a denominação estudo comparativo com intervenção pedagógica. A designação “experimental” é mantida apenas para identificar o grupo que utilizou as ferramentas digitais, sem pressupor randomização ou controle integral de variáveis, pois essas informações não foram fornecidas.

A amostra foi composta por 20 estudantes, distribuídos em grupo experimental (G1; n = 11) e grupo de controle (G2; n = 9). No G1, os estudantes utilizaram o NotebookLM para consultar e sintetizar materiais de apoio e, após a produção inicial, submeteram os textos a um agente avaliador configurado com base nas cinco competências do ENEM. A devolutiva foi posteriormente mediada e validada pelo professor regente. No G2, a pesquisa temática foi realizada com coletâneas impressas e consultas genéricas, e a correção ocorreu pelo método docente tradicional, sem suporte do agente de IA.

A coleta quantitativa, conforme descrita no estudo-base, consistiu nas notas de 0 a 200 atribuídas a cada uma das cinco competências, totalizando até 1.000 pontos por redação. A dimensão qualitativa foi associada à observação do engajamento dos estudantes e à análise dos relatórios diagnósticos gerados pela ferramenta e pelas correções docentes. O material consultado não informa instrumentos padronizados para registrar engajamento, protocolo de codificação qualitativa, quantidade de redações por estudante, tema(s) trabalhado(s), duração da intervenção ou características dos avaliadores. Assim, não se atribuem categorias qualitativas ou análises estatísticas além daquelas efetivamente reportadas.

Os dados foram organizados por médias de grupo em cada competência, e a diferença foi calculada subtraindo-se a média do G2 da média do G1. A interpretação é descritiva: não foram disponibilizados dados individuais, medidas de dispersão, notas anteriores à intervenção ou resultados de testes inferenciais. Consequentemente, a diferença entre médias não permite afirmar que o uso das ferramentas causou o desempenho observado nem que as diferenças sejam estatisticamente significativas. O estudo informa que os procedimentos respeitaram consentimento e anonimato; para reprodução ou publicação ampliada, recomenda-se explicitar o protocolo de consentimento, a faixa etária, as autorizações institucionais e, quando aplicável, os procedimentos de ética em pesquisa e proteção de dados.

RESULTADOS E DISCUSSÃO

A média geral reportada para o G1 foi de 643,64 pontos, enquanto a do G2 foi de 568,89, correspondendo a uma diferença descritiva de 74,75 pontos. A Tabela 1 apresenta os resultados por competência, conforme os valores do relatório original. A soma das médias das cinco

competências coincide com as médias gerais informadas, consideradas as aproximações decorrentes do arredondamento.

Tabela 1 – Médias por competência na produção textual dos grupos

Competência avaliada	G1 – com IA	G2 – controle	Diferença
C1 – Modalidade escrita formal	134,55	120,00	+14,55
C2 – Tema, gênero e repertório	141,82	111,11	+30,71
C3 – Projeto de texto e argumentação	123,64	97,78	+25,86
C4 – Coesão e mecanismos linguísticos	127,27	120,00	+7,27
C5 – Proposta de intervenção	116,36	120,00	−3,64
Média Geral	**643,64**	**568,89**	**+74,75**

Fonte: Dados da pesquisa (2026), conforme relatório de avaliação de redações informado pelos autores.

Desempenho global e interpretação dos dados

A diferença de 74,75 pontos favorece o grupo que participou da intervenção com ferramentas de IA. Esse dado oferece um indício descritivo relevante para reflexão pedagógica, mas não deve ser apresentado como "avanço significativo" em sentido estatístico. Sem notas de entrada, distribuição das pontuações, informação sobre equivalência inicial entre os grupos e teste inferencial, não é possível separar o possível efeito da intervenção de diferenças prévias, da experiência docente, do tema da redação ou de outras condições de aprendizagem. A formulação mais rigorosa é que, ao final do período observado, o G1 apresentou média superior à do G2.

Essa distinção não diminui o valor do achado; ao contrário, estabelece com precisão o alcance da evidência. Em turmas reais, estudos exploratórios podem indicar dimensões promissoras e orientar aperfeiçoamentos, desde que se evite transformar associação em causalidade. Investigações subsequentes poderão incluir avaliação diagnóstica e final com temas equivalentes, rubrica comum, correção independente por avaliadores humanos, registro das versões de texto e análise de medidas de dispersão.

Competência 2: compreensão temática e repertório sociocultural

A maior diferença entre os grupos ocorreu na Competência 2: 30,71 pontos a favor do G1. O resultado é compatível com a hipótese de que a consulta organizada a materiais de apoio tenha facilitado o acesso a conceitos e referências relacionados ao tema. O NotebookLM pode ter contribuído para aproximar a pesquisa da etapa de planejamento, permitindo que estudantes retomassem conteúdos e identificassem conexões com a questão proposta. Todavia, os dados não permitem afirmar qual fonte foi consultada, se as informações foram confirmadas nos textos originais ou se houve mudança qualitativa na pertinência do repertório mobilizado.

A interpretação pedagógica mais produtiva é tratar a ferramenta como apoio à leitura e à curadoria, não como mecanismo de fornecimento de repertório pronto. Uma sequência didática consistente pode solicitar que o estudante registre a fonte, explique com suas palavras a ideia selecionada e mostre como ela sustenta uma parte da tese. Essa prática reduz o risco de citações ornamentais e favorece a transferência do conhecimento para a argumentação, em consonância com a concepção de multiletramentos como prática crítica e situada (NEW LONDON GROUP, 1996; ROJO; MOURA, 2012).

Competência 3: projeto de texto e organização argumentativa

Na Competência 3, o G1 superou o G2 em 25,86 pontos. A diferença sugere que o planejamento apoiado por consulta temática e feedback estruturado pode ter favorecido a explicitação da tese e a ordenação dos argumentos. Uma ferramenta avaliadora que aponta problemas de progressão ou de relação entre ideias pode levar o estudante a reler o texto com maior atenção ao conjunto, especialmente quando a devolutiva é acompanhada por orientação para reescrita.

Não se conhece, contudo, a natureza dos comentários gerados nem a forma como os estudantes os utilizaram. Portanto, não se pode atribuir diretamente o resultado a um mecanismo específico da ferramenta. A mediação docente permanece central para distinguir uma sugestão pertinente de uma recomendação genérica e para solicitar que o estudante justifique as escolhas feitas na revisão. O ganho observado deve ser entendido como diferença de média, cuja explicação requer evidências processuais adicionais, como versões sucessivas, planos de texto e registros de decisão.

Competências 1 e 4: convenções da escrita e coesão

As diferenças foram positivas, mas menores, na Competência 1 (+14,55) e na Competência 4 (+7,27). Esses resultados são coerentes com a possibilidade de um agente avaliador sinalizar desvios gramaticais, repetições ou conexões pouco claras com rapidez. A utilidade didática do apontamento depende, porém, de que ele seja compreendido e convertido em revisão consciente. Se o estudante apenas aceita correções automáticas sem compreender a regra ou a relação semântica envolvida, a alteração local pode não produzir aprendizagem transferível para novas produções.

Também é importante que o ensino da modalidade formal não reduza a avaliação a uma contagem de erros. A adequação linguística precisa ser examinada em relação à clareza, à organização sintática e à construção de sentidos. De modo semelhante, o uso de conectivos não

assegura coesão quando as relações entre as ideias não estão semanticamente estabelecidas. Um feedback formativo deve explicar a função do problema e priorizar poucos aspectos revisáveis, evitando listas extensas de correções que possam sobrecarregar o estudante (HATTIE; TIMPERLEY, 2007).

Competência 5: proposta de intervenção

A Competência 5 foi a única em que a média do G2 superou a do G1, por 3,64 pontos. A diferença é pequena e, sem dispersão ou teste estatístico, não deve ser interpretada como evidência de superioridade robusta do grupo de controle. Ainda assim, o resultado aponta uma questão didática importante: a proposta de intervenção requer articulação entre o problema discutido e uma ação viável, com identificação de agentes, meios, finalidade e respeito aos direitos humanos. Uma devolutiva genérica pode não ajudar o estudante a especificar esses componentes nem a verificar sua coerência com os argumentos do desenvolvimento.

O dado sugere que, em futuras intervenções, vale dedicar instrução explícita à elaboração e à revisão da proposta, com análise de exemplos, comparação entre soluções vagas e detalhadas e verificação de consistência entre diagnóstico e ação. O agente pode funcionar como lista de verificação para localizar elementos ausentes, mas o professor precisa conduzir a avaliação de viabilidade e pertinência social. Assim, o resultado não indica que a tecnologia tenha prejudicado a Competência 5; indica que, na configuração descrita, não se observou vantagem média nessa dimensão.

Implicações pedagógicas e limites

Em conjunto, os resultados oferecem um quadro heterogêneo, e não uma validação uniforme da ferramenta: as diferenças são mais expressivas nas competências relacionadas a repertório e projeto argumentativo, menores em aspectos linguísticos e coesivos, e ligeiramente negativas na proposta de intervenção. A heterogeneidade recomenda alinhar cada recurso a objetivos de aprendizagem definidos. A tecnologia pode apoiar pesquisa, geração de perguntas para revisão e feedback rápido, mas não substitui a leitura crítica, a instrução explícita nem o acompanhamento da autoria.

Entre as limitações estão o tamanho reduzido da amostra, a ausência de informação sobre alocação dos participantes, a falta de dados de desempenho inicial e de medidas de variabilidade, bem como a descrição insuficiente da duração e dos instrumentos qualitativos. Também não foram apresentados os comandos de configuração dos agentes, o conjunto de fontes usado no NotebookLM, critérios para validar a correção automatizada ou exemplos

comparáveis das redações. Esses elementos restringem a replicabilidade e impedem inferências causais. A generalização para outras escolas, séries e contextos socioeconômicos deve, portanto, ser feita com prudência.

Em investigações futuras, recomenda-se documentar a sequência didática e os materiais, preservar versões sucessivas dos textos, utilizar temas equivalentes e rubrica compartilhada, incluir correção humana independente e registrar como os estudantes aceitam, adaptam ou rejeitam sugestões da IA. Também é importante acompanhar questões de acesso tecnológico, privacidade e consentimento, sobretudo por se tratar de estudantes do Ensino Médio. Esses procedimentos permitiriam avaliar não apenas a nota final, mas também os processos de aprendizagem, a autonomia crítica e a capacidade de transferir conhecimentos para novas tarefas.

CONSIDERAÇÕES FINAIS

Este estudo analisou o uso combinado do NotebookLM e de agentes avaliadores configurados para as competências do ENEM em uma intervenção de produção textual no Ensino Médio. Nas médias reportadas, o grupo que utilizou as ferramentas obteve pontuação geral superior à do grupo de controle, com maiores diferenças nas Competências 2 e 3. Esse padrão é compatível com a hipótese de que a organização de fontes e a devolutiva estruturada possam apoiar a mobilização de repertório e o planejamento argumentativo, mas os dados disponíveis não permitem estabelecer causalidade nem significância estatística.

A Competência 5 merece atenção particular: o grupo de controle apresentou média ligeiramente superior, o que reforça a necessidade de ensino explícito sobre a construção de propostas de intervenção detalhadas e coerentes com o problema. O conjunto dos resultados mostra que a contribuição da IA não é automática nem homogênea. Seu valor pedagógico depende de objetivos claros, fontes confiáveis, transparência sobre as limitações dos sistemas, validação docente e participação autoral do estudante.

Conclui-se que NotebookLM e agentes avaliadores podem integrar um ambiente de apoio à pesquisa e à revisão, desde que permaneçam subordinados a uma concepção formativa da escrita. A tecnologia não deve produzir a redação pelo estudante nem substituir o professor; pode ampliar oportunidades de consulta e feedback quando utilizada criticamente. Como agenda de pesquisa, propõe-se ampliar a amostra, descrever com precisão os procedimentos e observar as transformações entre versões, com especial atenção à autonomia, à qualidade do repertório e à elaboração da proposta de intervenção.

REFERÊNCIAS

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. A redação do Enem 2024: cartilha do(a) participante. Brasília, DF: Inep, 2024. Disponível em: https://www.gov.br/inep/pt-br/centrais-de-conteudo/acervo-linha-editorial/publicacoes-institucionais/avaliacoes-e-exames-da-educacao-basica/a-redacao-do-enem-cartilha-do-a-participante. Acesso em: 29 out. 2026.

HATTIE, John; TIMPERLEY, Helen. The power of feedback. Review of Educational Research, v. 77, n. 1, p. 81–112, 2007. DOI: https://doi.org/10.3102/003465430298487.

MIAO, Fengchun; HOLMES, Wayne. Guidance for generative AI in education and research. Paris: UNESCO, 2023. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000386693. Acesso em: 29 out. 2026.

NEW LONDON GROUP. A pedagogy of multiliteracies: designing social futures. Harvard Educational Review, v. 66, n. 1, p. 60–93, 1996. DOI: https://doi.org/10.17763/haer.66.1.17370n67v22j160u.

ROJO, Roxane; MOURA, Eduardo (org.). Multiletramentos na escola. São Paulo: Parábola Editorial, 2012.