


## Os efeitos da aprendizagem individualizada de línguas guiada por IA: uma metanálise

Hansol Lee<sup>1</sup>

Hansol Lee é Professor de Inglês da Academia Militar da Coreia, na Coreia do Sul. Ele obteve seu doutorado em Educação pela Universidade da Califórnia, Irvine. Suas áreas de interesse são métodos de pesquisa e estatísticas, tecnologia educacional, alfabetização e avaliação de programas.

 <https://orcid.org/0000-0002-6912-7128>

Jang Ho Lee<sup>2</sup>

Jang Ho Lee obteve seu DPhil em educação pela Universidade de Oxford e atualmente é Professor na Universidade Chung-Ang, na Coreia do Sul. Suas áreas de interesse são ensino de línguas assistido por computador (CALL, em inglês), uso da L1 no ensino de L2 e aquisição de vocabulário. Toda correspondência relacionada a esta publicação deve ser dirigida a ele.

 <https://orcid.org/0000-0003-2767-3881>

Tradutora: Lara Pereira de Sousa<sup>3</sup>

Graduanda do curso de Línguas Estrangeiras Aplicadas às Negociações Internacionais (LEA-NI), da Universidade Federal da Paraíba (UFPB).

Tradutora: Isa De Oliveira Gonçalves Gomes<sup>4</sup>

Graduanda do curso de Línguas Estrangeiras Aplicadas às Negociações Internacionais (LEA-NI), da Universidade Federal da Paraíba (UFPB).

 <http://lattes.cnpq.br/0184711662341479>

Data de recebimento: 20 de fevereiro de 2025.

Data de aceite: 15 de abril de 2025.

**Como citar este artigo:** LEE, H. LEE, H. J. Os efeitos da aprendizagem individualizada de línguas guiada por IA: Uma metanálise. *Revista InterCulturas*, João Pessoa, v. 1, n. 2, p. e74198, mai. 2025. Tradução de: Lara Pereira de Sousa e Isa De Oliveira Gonçalves Gomes. Texto fonte publicado em: LEE, H. LEE, H. J. Effects of content and language integrated learning at the primary school level: A multi-level meta-analysis. *Language Learning & Technology*, Honolulu, v. 28, n. 2. pp. 134–162, 2024.

### Resumo

A inteligência artificial (IA) avançou consideravelmente os métodos de individualização das oportunidades de aprendizagem de línguas, como a avaliação do progresso da aprendizagem e a recomendação de instrução individual eficaz. No presente estudo, conduzimos uma metanálise para sintetizar descobertas empíricas recentes sobre o uso da aprendizagem de línguas guiadas por IA e coletamos 61 amostras (N=8.282) de 17 projetos de pesquisas (ex. *Assessment to Instruction [A2I]*, *Duolingo* e *Project LISTEN*). Os resultados da nossa metanálise confirmaram que a aprendizagem individualizada de línguas guiada por IA foi eficaz para o desenvolvimento

<sup>1</sup> Endereço eletrônico: [hansol@kma.ac.kr](mailto:hansol@kma.ac.kr)

<sup>2</sup> Endereço eletrônico: [jangholee@cau.ac.kr](mailto:jangholee@cau.ac.kr)

<sup>3</sup> Endereço eletrônico: [lara.pereira0348@gmail.com](mailto:lara.pereira0348@gmail.com)

<sup>4</sup> Endereço eletrônico: [isa.goncalves@academico.ufpb.br](mailto:isa.goncalves@academico.ufpb.br)

linguístico dos estudantes ( $d = 1,18$ , com base em 26 amostras intragrupo,  $N = 2.262$ ) e apresentou um efeito positivo em comparação com as condições habituais de ensino ( $d = 0,39$ , com base em 35 amostras entre grupos,  $N = 6.020$ ). Além disso, os resultados das nossas análises de moderação do efeito do tratamento revelaram que a aprendizagem de línguas guiada por IA com aprendizagem de máquinas e sistemas híbridos foi mais impactante do que aquelas dos sistemas baseados em regras, os quais, por sua vez, podem ser mais úteis (em comparação com os primeiros) para compreender como as previsões são feitas sob uma perspectiva pedagógica. Implicações baseadas em evidências são apresentadas como base nos resultados desta meta-análise.

**Palavras-chave:** Inteligência Artificial (IA), Instruções Individualizadas, Aprendizagem de Línguas, Metanálise.

## Introdução

A inteligência artificial (IA) tem se tornado comum no nosso dia a dia; mesmo os celulares são equipados com assistentes de IA (ex. Siri e Assistente Google). Inclusive, o crescimento e a proliferação do Chat Generative Pretrained Transformer (ChatGPT), desenvolvido pela OpenAI (Ghumra 2022), em um ritmo extremamente rápido, aceleraram a introdução da IA ao público (Lund; Wang, 2023). Conseqüentemente, definir a IA não é mais tão complexo quanto era antes. Entre as várias perspectivas em como definir IA no contexto da educação (ex., Chen et al., 2020; Pokrivcakova, 2019; Sharma et al., 2019), no presente estudo, a definimos IA como “inteligência ou capacidade da máquina de demonstrar algum nível de inteligência e desempenhar uma gama de funções que requerem capacidades semelhantes às humanas” (p. 75267), adotando a definição de Chen et al. (2020) no campo da educação.

A IA tem sido integrada a inúmeros domínios (Xu et al., 2021), e o campo da educação não é uma exceção (ex. Du Boulay, 2016; Florea & Radu, 2019; Kandlhofer et al., 2019). A comunidade educacional está agora levantando preocupações sobre o potencial da IA de substituir os professores humanos, embora estudos existentes tenham revelado algum ceticismo em relação a essa possibilidade entre os professores em serviço (ex. Chan; Tsi, 2023; Tao et al., 2019). Apesar desse ceticismo, estatísticas sugerem que a IA desempenhará um grande papel na educação no futuro (GMI, 2020). No campo do ensino e aprendizagem de línguas, o potencial da IA tem sido amplamente explorado na área de *Computer Assisted Language Learning* (CALL) (Kessler, 2023).

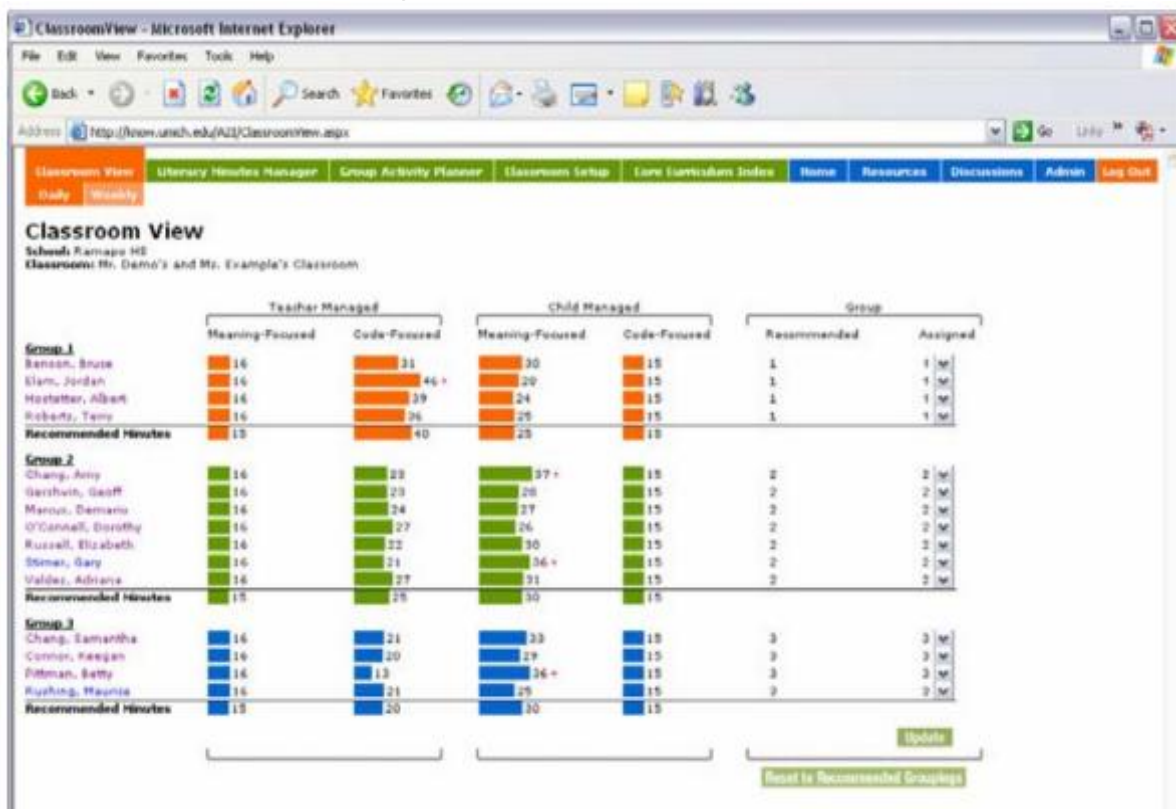
No presente estudo, sintetizamos as contribuições pedagógicas da IA no auxílio à aprendizagem individualizada de línguas. A ideia de prover uma aprendizagem individualizada para diferentes grupos de estudantes na mesma sala de aula (ou seja, instrução diferenciada) remete ao início do século XX (Washburne, 1925). A maioria dos professores concordam que identificar as diferenças dos seus estudantes é uma questão importante na aprendizagem de línguas e que ser capaz de responder a essa diversidade é um aspecto integral de uma aprendizagem de línguas de alta qualidade (Puzio et al., 2020). De qualquer forma, isso não é uma tarefa fácil. Por exemplo, de acordo com uma pesquisa nacional nos EUA, 84% dos professores acharam difícil implementar tais métodos de ensino (Loveless et al., 2008). A metanálise de Puzio et al. (2020) revelou que o efeito geral da instrução diferenciada em alfabetização foi marginal ( $d = 0,13$ ), o que deixa o potencial desta instrução em questão.

O crescimento recente da IA tornou essa abordagem pedagógica muito mais efetiva e eficiente (Chen et al., 2020), e essa contribuição foi proposta como “a maior oportunidade para a IA na educação” (Lee; Qiufan, 2021, p.118). Assim, acreditamos fortemente que a aprendizagem individualizada de línguas guiada por IA tem um efeito pedagógico mais promissor na melhoria do desempenho dos alunos do que os resultados de Puzio et al. (2020). Portanto, o objetivo do presente estudo foi sintetizar as descobertas empíricas sobre as oportunidades de aprendizagem individualizada de línguas guiadas por IA por meio de uma abordagem metanalítica.

É fundamental destacar que “a maioria das ferramentas de AIEd [IA na Educação] tende a impulsionar a homogeneização dos alunos” (Holmes; Tuomi, 2022, p.559), embora muitos aplicativos guiados por IA afirmam oferecer oportunidades de aprendizagem de línguas individualizadas. Portanto, neste estudo, buscamos restringir nossa interpretação da aprendizagem de línguas individualizadas, considerando as duas ramificações distintas da tecnologia de IA no domínio da aprendizagem de línguas. Primeiramente, uma das ramificações desenvolveu um programa de aprendizagem baseado em uma IA, que oferece aos alunos trajetórias individualizadas, ajustando as abordagens instrucionais-alvo (ou seja, instrução individualizada). A estrutura A2i de Connor et al. (2007, 2009, 2011a, 2011b, 2013) é um exemplo relevante, no qual os resultados do desempenho individual dos alunos, obtidos

por meio de avaliações formativas, são consideradas no algoritmo da IA. Essa, por sua vez, recomenda diferentes combinações de instruções para cada aluno (ver figura 1, onde o A2i sugere a quantidade de instruções recomendada para alunos de forma individualizada com base em suas pontuações de vocabulário e leitura de palavras).

Figura 1: Visualização da sala de aula A2i



Fonte: Adaptado de Connor et al. (2007).

A segunda ramificação oferece plataformas instrutivas em que os alunos recebem trajetórias individualizadas por meio de diferentes itens do idioma-alvo, com base em seu progresso (ou seja, em suas respostas a itens anteriores). Um exemplo bem conhecido nessa categoria é o Project LISTEN de Mostow et al. (2003a, 2003b, 2013), que oferece tutores de leitura, um Sistema de Tutor Inteligente (ITS, em inglês), com ferramentas de avaliação, planejamento e suporte ao desempenho, permitindo que os alunos aprimorem sua leitura oral dentro do programa de leitura. Além disso, o Duolingo (ex. Al Otaiba et al., 2011; Purwanto et

al., 2022), uma plataforma de aprendizagem de línguas fundada em 2011, é outro exemplo relevante.

Vale ressaltar que não consideramos tecnologias baseadas em IA, como chatbots, robôs sociais, e experiência de escrita automática (AWE, em inglês), que já foram amplamente estudadas, como evidenciado por revisões existentes (ex. Huang et al., 2022; Lee; Lee, 2022a; Nunes et al., 2022).

## Contexto

### Individualização para a aprendizagem de línguas

A noção de oferecer oportunidades de aprendizagem individualizadas deu origem a diversos tipos de programas em contextos acadêmicos (Wang; Walberg, 1985). O campo de aprendizagem de línguas também reconheceu a importância da instrução diferenciada (International Reading Association, 2000), com diversos programas desenvolvidos para esse fim (ex. School Wide Enrichment Model in Reading, Reis et al., 2011; Individualized Student Instruction, Connor et al., 2007; Integrated Curriculum Model, Van Tassel-Baska; Wood, 2010). Teoricamente fundamentada na Zona de Desenvolvimento Proximal de Vygotsky (1978), a instrução individualizada busca identificar com precisão o estágio atual do estudante e, em seguida, proporcionar uma oportunidade de aprendizagem mais ajustada do que o modelo padronizado para todos. Dessa forma, a aprendizagem pode ser “personalizada, ajustada ou adaptada às experiências, habilidades e interesses de cada aluno” (Waxman et al., 2013, p.405).

Propõe-se que a aprendizagem individualizada será significativamente aprimorada com a integração de tecnologias guiadas por IA (Lee; Qiufan, 2021), que fornecem um caminho de aprendizagem personalizado para cada aluno ao analisar dados relacionado ao seu comportamento, desempenho e padrões de aprendizagem. Sob essa perspectiva, de acordo com a revisão de literatura de Chassignol et al. (2018) sobre AIED, a IA tem sido incorporada em três áreas: (a) administração, (b) instrução e (c) aprendizagem. As comunidades de ensino e de estudantes de línguas também adotaram recentemente a IA em seus respectivos

domínios. Por exemplo, ao adotar o escopo de Chassignol et al., o Alpie's Alley (ex. Chambers et al., 2008a, 2008b), um ITS para aprendizagem de línguas, abrange três áreas: (a) avaliação e fornecimento de feedback (administração), (b) determinação de planos individualizados (instrução) e (c) detecção do status de aprendizagem dos alunos e abordagem de suas deficiências (aprendizagem).

Apesar do entusiasmo em torno das tecnologias guiadas por IA para a instrução individualizada, Holmes e Tuomi (2022) levantaram algumas preocupações, duas das quais são discutidas a seguir. Em primeiro lugar, ao contrário do que afirmam os defensores entusiastas das tecnologias guiadas por IA no domínio da educação, a instrução individualizada pode ser, e tem sido, oferecida por professores humanos sem o auxílio dessas tecnologias. Além disso, não há evidências sólidas de que a primeira seja mais eficaz do que a segunda. Em segundo lugar, os caminhos de aprendizagem individualizados sugeridos por tecnologias guiadas por IA podem não ser úteis para os alunos atuais, pois são desenvolvidos com base nos alunos anteriores. Diante dessas questões, é importante examinar até que ponto a aprendizagem individualizada guiada por IA pode beneficiar o desenvolvimento linguístico dos estudantes de línguas.

Apenas uma pequena quantidade de metanálises sintetizou descobertas empíricas relacionadas à aprendizagem guiada por IA (ex. Lin et al., 2022; Martínez et al., 2023; Zheng et al., 2023). Dentre elas, vale a pena revisar a metanálise de Lin et al., (2022), que ofereceu uma percepção sobre o efeito da IA na aprendizagem de línguas. Embora não tenham explorado o conceito de individualização, nove dos 21 estudos empíricos focam na aprendizagem de línguas. Seus resultados revelaram que o uso da IA na aprendizagem de línguas apresentou um efeito de tamanho pequeno ( $d = 0,26$ ). Como o principal objetivo dessa metanálise foi fornecer perspectivas abrangentes sobre a AIEd, ela apresenta uma limitação que buscamos abordar no presente estudo. Mais especificamente, não foram apresentadas descobertas mais detalhadas e aprofundadas sobre nosso interesse, como os efeitos de moderação (ou seja, como diferentes características dos estudos ou dos alunos influenciam a eficácia da IA na aprendizagem de línguas), não podendo ser encontradas em estudos anteriores nenhuma das metanálises, incluindo Lin et al. (2022). Nosso objetivo foi investigar o papel das seguintes

variáveis moderadoras, que são particularmente significativas no uso de tecnologias de IA em contextos de aprendizagem de línguas.

### **Moderadores potenciais**

Essa seção apresenta os cinco moderadores identificados, tendo como hipótese suas influências na eficácia da aprendizagem individualizada de línguas guiada por IA.

### **Idade**

Pesquisas sobre a aprendizagem de línguas sugerem que diferentes faixas etárias utilizam diferentes mecanismos de aprendizagem (Bley-Vroman, 1990, 2009) e diferem em termos de experiência ou expertise na aprendizagem de línguas (Johnson, 2005). Essas diferenças entre faixas etárias podem moderar os efeitos de um determinado tipo de instrução ou programa de aprendizagem (ex. Johnson; Swain, 1994; Lee; Macaro, 2013). Da mesma forma, pode-se esperar que os efeitos da aprendizagem individualizada de línguas guiada por IA variem entre faixas etárias. Temos como hipótese que faixas etárias mais jovens, com experiência de aprendizagem limitada e menor maturidade cognitiva, se beneficiam mais dos caminhos de aprendizagem personalizados oferecidos pelas ferramentas de IA, devido aos seus níveis mais baixos de experiência em aprendizagem de línguas e maturidade cognitiva em comparação com seus colegas mais velhos.

### **Contexto de aprendizagem**

A aprendizagem individualizada de línguas guiada por IA tem sido analisada em diferentes contextos de aprendizagem, incluindo línguas maternas, segundas línguas e línguas estrangeiras. No caso da aprendizagem da língua materna, na qual a instrução diferenciada está amplamente associada ao desenvolvimento da alfabetização (International Reading Association, 2000), a aprendizagem individualizada de línguas guiada por IA também tem sido direcionada para a melhoria da alfabetização (ou seja, habilidades de leitura) (ex. Chambers et

al., 2008a, 2008b, 2011; Connor et al., 2007, 2009, 2011a, 2011b; Mills-Tetty et al., 2009; Mostow et al., 2013). Outro grupo de estudos (ex. Ali, 2021; Aulia et al., 2020; Kusumadewi; Widyastuti, 2018; Nagata, 1993; Purwanto et al., 2022) foi conduzido em contextos de línguas estrangeiras, nos quais o idioma-alvo não é falado fora da sala de aula e a motivação para aprendê-lo geralmente está ligada a objetivos profissionais (Dewaele, 2009). Por fim, alguns estudos (ex., Heilman et al., 2010; Loewen et al., 2020; Poulsen, 2004; Vesselinov; Grego, 2012) tiveram como alvo alunos de um segundo idioma, para os quais o idioma-alvo é o idioma falado na comunidade em que residem atualmente. Dada as diversas diferenças entre os contextos (ex., a motivação para aprender o idioma-alvo, a quantidade de exposição ao idioma fora da sala de aula e o status da língua-alvo), temos a hipótese de que o contexto pode ser um moderador importante que influencia os efeitos da aprendizagem individualizada de línguas guiada por IA.

### **Domínio da língua-alvo**

Como mencionado anteriormente, a instrução individualizada tem sido amplamente associada ao desenvolvimento da alfabetização. Assim, espera-se que uma grande parte das pesquisas sobre instrução individualizada guiada por IA tenha estabelecido a leitura (ou seja, a alfabetização) como o domínio da língua-alvo e tenha examinado os efeitos dos programas dos pesquisadores no desenvolvimento da alfabetização. Esses programas incluem o Alphonie's Alley de Chambers et al. (2008a, 2008b, 2011), o ISI com o software A2i de Connor et al. (2007, 2009, 2011a, 2011b, 2013) e o Project LISTEN Reading Tutor de Mostow et al. (2003a, 2003b, 2013). Além da leitura, o vocabulário também tem sido amplamente abordado nas pesquisas sobre instrução individualizada guiada por IA, presumivelmente devido às “características das aulas de vocabulário, nas quais os itens lexicais-alvo podem ser apresentados individualmente e de forma independente” (Lee; Lee, 2022a, p. 121). Exemplos incluem o Duolingo (ex. Ali, 2021; Aulia et al., 2020; Kusumadewi; Widyastuti, 2018; Zamzami, 2019) e o Memrise (ex. Esmaili; Shahrokhi, 2020; Nuralisah; Kareviati, 2020; Rohim et al., 2022). Por fim, outras habilidades, como a compreensão oral, a produção oral e a gramática, foram abordadas em um número

limitado de estudos (ex. Hassani et al., 2016; Loewen et al., 2020; Maesaroh, 2021; Purwanto et al., 2022; Tsiriga; Virvou, 2004).

## O papel de assistência da IA

Quando as tecnologias da IA são adotadas em estudos empíricos, seus papéis variam de acordo com os objetivos do estudo. De acordo com Lin et al. (2022), por exemplo, aproximadamente 78% dos estudos coletados para sua metanálise utilizaram ITS para fornecer aos seus alunos tutores virtuais para aprendizagem individualizada de línguas (os 22% restantes não eram relevantes para nossa definição de aprendizagem de língua individualizada). No caso de Zheng et al. (2003), onde um total de sete papéis diferentes foram identificados para suas análises dos moderadores, os (a) ITS, os (b) sistemas de teste e diagnóstico/sistemas de avaliação ou monitoramento e os (c) sistemas de aprendizado adaptativo (ALS, em inglês) / ferramentas de suporte ao aprendizado foram relevantes para a aprendizagem individualizada de línguas guiada por IA. De maneira similar, embora a metanálise de Hwang (2022) tenha se concentrado em matemática, este estudo encontrou que a maioria (aproximadamente 63%) das instruções anteriores guiadas por IA utilizou (a) ITS ou (b) ALS. Outros utilizaram ferramentas de aprendizado de línguas para fornecer instruções a alunos individuais de acordo com seus níveis e ritmos de aprendizado, usando sistemas de IA (veja Godwin-Jones, 2021, para uma tendência recente do uso de big data no aprendizado de línguas).

## Tipo de Sistema de IA

De modo geral, a IA aceita dados de entrada (ex. texto, fala, imagens), processa esses dados de acordo com diversos métodos (ex, sistemas de algoritmos baseados em regras, que aplicam regras especificadas manualmente para resolver problemas, ou sistemas de algoritmos de aprendizagem de máquina, que fazem previsões com base em regras aprendidas automaticamente) e fornece resultados, como previsões ou classificações (veja Lee et al.,

2019a, para um estudo exemplar sobre aprendizado de línguas). Dependendo de como a IA é projetada, especialmente dos tipos de algoritmos utilizados (ou seja, sistemas de algoritmos baseados em regras, sistemas de algoritmos de aprendizagem de máquina ou ambos), os resultados podem ser analisados e usados para ajustar o sistema e, em seguida, repetir todo o procedimento (isto é, um loop) até que o estado final desejado seja alcançado (Witten et al., 2016). Um exemplo de um sistema de IA baseado em regras é o framework A2i de Connor et al. (ex., 2007, 2009, 2011a, 2011b), cujo algoritmo de IA desenvolvido foi modelado com base na gestão de sala de aula dos instrutores. O sistema de IA analisava os dados de avaliação para fornecer recomendações ideais com base em regras predefinidas. Considerando os sistemas de IA baseados em aprendizagem de máquina, como o Duolingo (ex. Ali, 2021; Aulia et al., 2020; Esmaeili; Shahrokhi, 2020), o algoritmo de IA determina as lições mais adequadas com base nas necessidades específicas dos alunos. O algoritmo então aprende a melhorar sua precisão em resposta às interações dos alunos. Dado que o número de usuários no terceiro trimestre de 2022 era aproximadamente 56,5 milhões (dados da empresa), melhorar a precisão do algoritmo de IA do Duolingo teria sido irrealista. Portanto, diferentes tipos de sistemas de IA podem influenciar a eficácia geral de seu uso para alunos individuais. Embora metanálises anteriores (ex. Hwang, 2022; Lin et al., 2022; Zheng et al., 2023) tenham identificado os diferentes papéis da IA empregados em estudos empíricos anteriores, nenhuma delas abordou os tipos de sistemas ou algoritmos de IA.

## **O presente estudo**

O objetivo do presente estudo foi sintetizar os resultados empíricos das instruções individualizadas orientadas por IA para a aprendizagem de línguas, por meio de uma perspectiva metanalítica. Para obter uma visão mais equilibrada dos efeitos das instruções individualizadas orientadas por IA, identificamos tanto os estudos dentro do grupo (ou seja, análises pré-pós) quanto os estudos entre grupos (ou seja, aqueles que comparam os grupos experimental e de rotina). As três questões de pesquisa seguintes guiaram a metanálise atual para alcançar este objetivo de pesquisa.

1. Em que medida a aprendizagem individualizada guiada por IA é eficaz para a aprendizagem de línguas?
2. Em que medida a aprendizagem individualizada de línguas guiada por IA tem um efeito de tratamento significativo em comparação à condição de controle?
3. Quais são os fatores importantes e contextuais dos alunos que influenciam os efeitos do tratamento na aprendizagem individualizada de línguas guiada por IA? Em que medida os diversos papéis e algoritmos da IA estão relacionados aos seus efeitos de tratamento?

## MÉTODOS

### Busca literária

Para iniciar a metanálise, realizamos uma busca da literatura após definir os quatro critérios de inclusão seguintes. Para ser incluído na metanálise atual, um estudo deve (a) ser projetado para melhorar os resultados de aprendizado de línguas dos alunos, (b) adotar IA para apoiar a aprendizagem de línguas individualizada, (c) relatar estatísticas descritivas suficientes para calcular o  $d$  de Cohen para tamanhos de efeito (ex. médias e desvios padrão dos resultados dos testes e tamanhos das amostras), e (d) ser escrito em inglês. Para identificar o maior número possível de estudos, não restringimos os anos de publicação ou os tipos de artigos. Os autores do presente estudo realizaram uma busca na literatura, que consistiu em uma busca em banco de dados, revistas e indentificação de referências retrospectivas/prospectivas.

Primeiro, realizamos uma busca em bancos de dados. Para isso, visitamos Web of Science (WoS), ProQuest e SCOPUS e usamos a seguinte combinação de palavras-chave nos campos de título e resumo, de acordo com quatro parâmetros principais: (a) tipos de instrução: sistema de tutoria inteligente ('intellige\* tutor\*' OU 'intellig\* agent' OU 'cognit\* tutor\*' OU 'adapt\* tutor\*' OU 'cognit\* virtual companion\*' OU 'intellige\* coaching system\*' OU 'adapt\*

instruct\*’ OU ‘adapt\* teach\*’), instruções diferenciadas (‘differentiat\* instruct\*’ OU ‘differentiat\* teach\*’), ou instruções individualizadas (‘individuali\* instruct\*’ OU ‘individuali\* teach\*’); (b) tópicos de aprendizagem de línguas: aprendizagem de primeira língua (‘language learning’ OU ‘language teaching’ OU ‘language acquisition’), aprendizagem de segunda/lingua estrangeira (‘second language’ OU ‘foreign language’ OU L2), ou aprendizagem de inglês (‘English learning’ OU ‘English teaching’ OU ‘English acquisition’ OU EFL OU ESL); (c) adoção de sistemas de IA (CALL OU agent OU interface OU ‘artificial intelligence’ OU AI OU ‘computer-assisted’ OU ‘computer\*’); e (d) quantificação (‘impact\*’ OU ‘effect\*’). Assim, identificamos os estudos primários relevantes na metanálise atual.

Para as buscas em periódicos, visitamos revistas acadêmicas relevantes, incluindo a *Annual Review of Applied Linguistics*, o *CALICO Journal*, a *Computer-Assisted Language Learning*, a *Educational Research Review*, a *Language Learning*, a *Language Learning & Technology*, a *Language Teaching Research*, a *ReCALL*, a *Review of Educational Research*, *Studies in Second Language Acquisition* e a *TESOL Quarterly*. As palavras-chaves mencionadas anteriormente foram utilizadas novamente para as buscas em periódicos.

Após identificar a literatura relevante sobre o tema, realizamos buscas retroativas e prospectivas de referências. Para a busca retroativa de referências, verificamos as listas de referências dos estudos identificados; para a busca prospectiva, utilizamos a função “Citado por” do Google Acadêmico para localizar estudos subsequentes que citaram os estudos identificados. Um total de 186 casos foram identificados após a remoção de duplicatas.

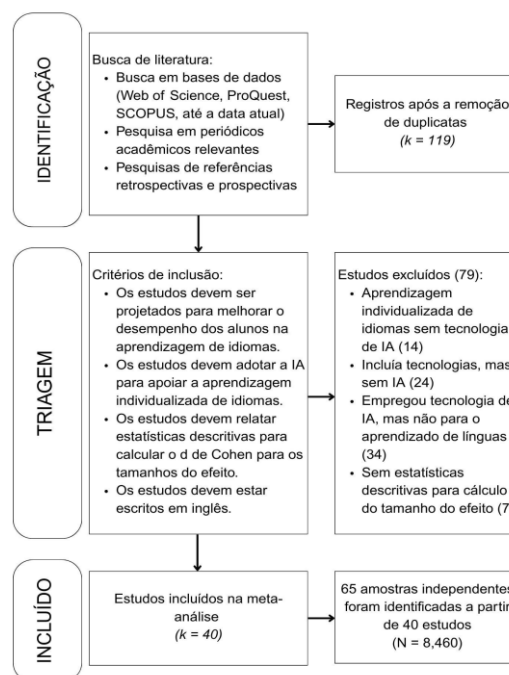
Os títulos e resumos desses 186 casos foram então revisados quanto à sua relevância para o tema, e 119 casos permaneceram elegíveis para uma revisão completa, a fim de verificar se atendiam aos quatro critérios de inclusão. Como mostrado na Figura 2, um total de 79 estudos foram excluídos pelos seguintes motivos: (a) os estudos incluíam instruções individualizadas (ex. instruções diferenciadas, tutoria individual), mas não utilizavam tecnologias de IA (k = 14); (b) os estudos usavam diversas tecnologias computacionais (ex. sistemas de webconferência, coaching online, avaliação formativa computadorizada), mas não incluíam tecnologias de IA (k = 24); (c) os estudos empregam tecnologias de IA, mas não mediam os resultados do aprendizado de línguas (k = 34); e (d) os estudos não apresentavam

dados descritivos para o cálculo do tamanho do efeito ( $k = 7$ ). Como resultado, 40 estudos foram selecionados para a metanálise atual (veja o Material Suplementar para a tabela que apresenta as informações de cada estudo).

## Construção do conjunto de dados

Esta seção descreve todo o procedimento para a construção dos conjuntos de dados: (a) identificação de amostras independentes e suas estruturas de dados multiníveis, (b) estimativa dos tamanhos do efeito, (c) detecção de valores atípicos e casos influentes, além da verificação de qualquer indício de efeitos de estudos pequenos e (d) codificação dos moderadores. Como incluímos tanto estudos de um único grupo (ou seja, com design pré-pós) quanto estudos experimentais com dois grupos (ou seja, grupo experimental vs. grupo de controle tradicional), havia dois tipos de conjuntos de dados: um para tamanhos do efeito do grupo e outro para tamanhos de efeito entre grupos. Todos os procedimentos foram realizados pelos autores do presente estudo. Quaisquer divergências foram resolvidas por meio de discussões produtivas.

Figura 2: Processo de busca e exclusão (*diagrama de fluxo PRISMA*)



Fonte: Elaboração própria (2024).

## Compreendendo a estrutura dos dados

Os 40 estudos identificados foram então revisados para compreender a estrutura do conjunto de dados. Primeiramente, alguns dos estudos utilizaram várias amostras independentes. Por exemplo, Chambers et al. (2011) incluíram alunos do primeiro e do segundo ano e relataram os resultados separadamente para cada grupo, representando assim, duas amostras independentes (como se fossem dois estudos distintos; portanto, referimo-nos a eles como Chambers et al., 2011[1] e Chambers et al., 2011[2]). Da mesma forma, Mills-Tetty et al. (2009) recrutaram três grupos diferentes de (a) uma escola privada, (b) uma escola pública e (c) um programa educacional informal, e relataram os resultados separadamente para cada grupo. Portanto, identificamos três amostras neste estudo, referidas como Mills-Tetty et al. (2009) [1], Mills-Tetty et al. (2009) [2] e Mills-Tetty et al. (2009) [3]. Como resultado, identificamos um total de 65 amostras ( $N = 8.460$ ), consistindo em 26 amostras de pré-pós-teste (isto é, dentro do grupo; diferenças dentro dos grupos de tratamento [entre o pré-teste e o pós-teste] sem grupos de controle) ( $N = 2.262$ ) e 39 amostras entre grupos (isto é, diferenças entre os grupos de tratamento e de controle [com base nos resultados do pós-teste]) ( $N = 6.198$ ).

Em segundo lugar, alguns estudos foram conduzidos pelos mesmos ou quase os mesmos pesquisadores como parte de um projeto de pesquisa maior. Chambers et al. (2008a), Chambers et al. (2008b), Chambers et al. (2011) [1] e Chambers et al. (2011) [2] utilizaram o mesmo ITS, denominado Alphie's Alley. Da mesma forma, Ali (2021), Aulia et al. (2020), Kusumadewi e Widyastuti (2018), Purwanto et al. (2022) [1], Purwanto et al. (2022) [2] e Zamzami (2019) utilizaram o Duolingo. Se o mesmo número de participantes foi medido em vários estudos, os estudos mais recentes foram incluídos na metanálise. No entanto, nenhum mediu os mesmos participantes várias vezes (ou seja, cada estudo tinha amostras independentes). Assim, concluímos que nossas 65 amostras estavam inseridas em 17 projetos de pesquisa.

## Estimando os tamanhos de efeito

A estimativa do tamanho do efeito é essencial para metanálises, pois permite comparar matematicamente os achados empíricos de diferentes estudos com distintas medidas de resultado. Uma das abordagens mais comuns é calcular a diferença média padronizada (DMP), também conhecida como  $d$  de Cohen, dividindo a diferença média de cada resultado entre dois grupos pelo desvio padrão agrupado entre esses grupos. Mais precisamente, os resultados do pré-teste e pós-teste são considerados para estudos com um único grupo, enquanto, no caso de estudos experimentais com dois grupos, são utilizados os resultados do pós-teste dos grupos de tratamento e de controle. Assim, os tamanhos de efeito padronizados têm a vantagem de serem escalas numéricas para medir a força dos efeitos do tratamento. Portanto, podemos sintetizar objetivamente as magnitudes dos efeitos da aprendizagem individualizada de idiomas guiada por IA entre diferentes estudos empíricos. Durante a estimativa, considerando que alguns estudos tinham tamanhos de amostra pequenos ( $<50$ ), adicionamos o fator de correção  $J$  ao cálculo do tamanho do efeito para evitar possíveis vieses (Hedges, 1981); a unidade dos nossos tamanhos de efeito foi o  $d$  Cohen não tendencioso (também conhecido como  $g$  de Hedges). Após estimar os tamanhos do efeito para cada resultado, calculamos um total de 119 tamanhos de efeito.

## Detectando Valores atípicos/ casos influentes e viés de publicação

Após o cálculo do tamanho do efeito, realizamos diagnósticos de valores atípicos/casos influentes utilizando o pacote *metafor* (Viechtbauer, 2010) no software R. Essa abordagem quantifica como certos casos são considerados valores atípicos e casos influentes com base em vários índices matemáticos, como esíduos estudentizados excluídos, DFFITS, distância de Cook e COVRATIO (ver Viechtbauer; Cheung, 2010, para suas definições). Com base em diagnósticos sistemáticos quantificados de forma sistemática, essa abordagem detecta matematicamente valores atípicos e casos influentes cuja exclusão pode impactar consideravelmente a

metanálise geral. Assim, valores atípicos devem ser excluídos, enquanto uma análise de sensibilidade de acompanhamento é necessária para os casos influentes, a fim de verificar se a inclusão ou a exclusão desses casos pode ter afetado os resultados da metanálise (ver Viechtbauer; Cheung, 2010, para a justificativa e Lee; Lee, 2022b, para um guia prático). Os resultados indicaram que nenhuma das amostras dentro do grupo foi detectada como valor atípico; no entanto, quatro tamanhos de efeito foram identificados no conjunto de amostras entre grupos (ou seja, nenhum caso influente foi detectado em ambos os conjuntos de dados). Portanto, excluímos os quatro possíveis valores atípicos do conjunto de dados correspondente. Como resultado, permaneceram 115 tamanhos de efeito de 61 amostras ( $N = 8.282$ ) baseados em 17 projetos, sendo 39 tamanhos de efeito de 26 amostras dentro do grupo ( $N = 2.262$ ) baseados em 14 projetos e 76 tamanhos de efeito de 35 amostras entre grupos ( $N = 6.020$ ) baseados em 13 projetos. As Tabelas 1 e 2 apresentam as estruturas dos conjuntos de dados.

Após detectar os valores atípicos, verificamos sinais de viés de publicação. Considerando que estudos com significância estatística tendem a ser publicados com mais frequência, a síntese de estudos primários publicados sobre a metanálise pode negligenciar estudos não publicados com achados nulos ou resultados negativos (Song et al., 2010), que provavelmente estão sub-representados na literatura. Mais especificamente, se os tamanhos do efeito calculados a partir dos estudos coletados estiverem predominantemente em uma única direção, isso pode indicar viés de publicação, sugerindo que o conjunto de dados pode estar canonizando fatos falsos. Um gráfico de funil é uma abordagem amplamente utilizada para visualizar a distribuição simétrica em torno dos tamanhos de efeito e dos erros-padrão (Debray et al., 2018). O teste de regressão de Egger (Egger et al., 1997) é outro método comum para verificar estatisticamente o viés de publicação.

**Tabela 1:** Estrutura completa do conjunto de dados da amostra dentro do grupo

| <b>Nível 3: Projeto (14)</b> | <b>Nível 2: Amostra (26)</b>        | <b>Nível 1: Resultado (39)</b> |
|------------------------------|-------------------------------------|--------------------------------|
| Alphie's Alley               | Chambers et al. (2008a)             | 1                              |
| Autotutor Software           | Mohammadzadeh & Sarkhosh (2018) [1] | 1                              |
|                              | Mohammadzadeh & Sarkhosh (2018) [2] | 1                              |
| Babbel                       | Loewen et al. (2020)                | 3                              |
| Bussu                        | Maesaroh (2021)                     | 1                              |
|                              | Ali (2021)                          | 1                              |
| Duolingo                     | Vesselinov & Grego (2012)           | 1                              |
|                              | Purwanto et al. (2022) [1]          | 1                              |
|                              | Purwanto et al. (2022) [2]          | 1                              |
|                              | Al Otaiba et al. (2011)             | 3                              |
|                              | Connor et al. (2007)                | 3                              |
| ISI with A2i                 | Connor et al. (2009)                | 3                              |
|                              | Connor et al. (2011a)               | 2                              |
|                              | Connor et al. (2011b)               | 2                              |
| iSTART-E                     | McCarthy et al. (2020)              | 1                              |
| ITSS                         | Wijekumar et al. (2012)             | 1                              |
| IVELL                        | Hassani et al. (2016)               | 4                              |
|                              | Esmaili & Shahrokhi (2020)          | 1                              |
|                              | Nuralisah & Kareviati (2020)        | 1                              |
| Memrise                      | Rohim et al. (2022)                 | 1                              |
|                              | Taebenu & Katemba (2021) [1]        | 1                              |
|                              | Taebenu & Katemba (2021) [2]        | 1                              |
| Moodle                       | Jia et al. (2012)                   | 1                              |
| Negative Knowledge           | Demir (2020)                        | 1                              |
| Nihongo-CALI                 | Nagata (1995)                       | 1                              |
| Web-based ICALL              | Tsiriga & Virvou (2004)             | 1                              |

Fonte: Elaboração própria (2024).

**Tabela 2:** Estrutura completa do conjunto de dados da amostra entre grupos

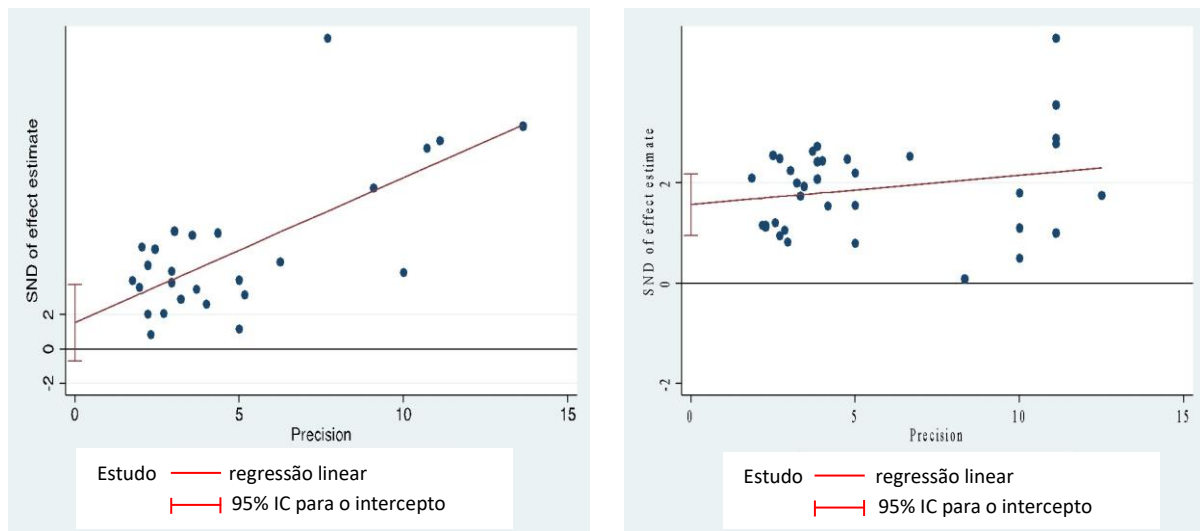
| Nível 3: Projeto (13) | Nível 2: Amostra (35)         | Nível 1: Resultado (76) |
|-----------------------|-------------------------------|-------------------------|
| Alphie's Alley        | Chambers et al. (2008a)       | 4                       |
|                       | Chambers et al. (2008b)       | 3                       |
|                       | Chambers et al. (2011) [1]    | 3                       |
|                       | Chambers et al. (2011) [2]    | 2                       |
|                       | Ali (2021)                    | 1                       |
| Duilingo              | Aulia et al. (2020)           | 1                       |
|                       | Purwanto et al. (2022) [1]    | 1                       |
|                       | Purwanto et al. (2022) [2]    | 1                       |
|                       | Al Otaiba et al. (2011)       | 5                       |
|                       | Connor et al. (2007)          | 3                       |
| ISI with A2i software | Connor et al. (2011a)         | 2                       |
|                       | Connor et al. (2011b)         | 2                       |
|                       | Connor et al. (2013) [1]      | 1                       |
|                       | Connor et al. (2013) [2]      | 1                       |
|                       | Connor et al. (2013) [3]      | 1                       |
| iSTART                | McNamara et al. (2006)        | 2                       |
| ITSS                  | Wijekumar et al. (2012)       | 1                       |
| Memrise               | Esmaeili & Shahrokhi (2020)   | 1                       |
|                       | Rohim et al. (2022)           | 1                       |
| Moodle                | Jia et al. (2012)             | 1                       |
| Negative Knowledge    | Demir (2020)                  | 1                       |
|                       | Nagata (1993)                 | 1                       |
| Nihongo-CALI          | Nagata (1995)                 | 1                       |
|                       | Nagata (1996)                 | 1                       |
|                       | Mostow et al. (2003a)         | 2                       |
| Project LISTEN        | Mostow et al. (2003b)         | 1                       |
|                       | Mostow et al. (2013)          | 12                      |
|                       | Poulsen (2004)                | 5                       |
|                       | Aist et al. (2001) [1]        | 1                       |
|                       | Mills-Tetty et al. (2009) [1] | 2                       |
| REAP                  | Mills-Tetty et al. (2009) [2] | 2                       |
|                       | Mills-Tetty et al. (2009) [3] | 2                       |
|                       | Heilman et al. (2010)         | 3                       |
| The First 400 Words   | Fehr et al. (2012)            | 3                       |
| Web-based ICALL       | Tsiriga & Virvou (2004)       | 1                       |

Fonte: Elaboração própria (2024).

Os resultados do teste de Egger revelaram que os tamanhos do efeito calculados para as amostras dentro do grupo não apresentaram sinais de viés de publicação ( $\beta = 1,53$ ,  $EP = 1,08$ ,  $p = 0,17$ ), enquanto aqueles para as amostras entre grupos mostraram um sinal estatisticamente significativo de viés ( $\beta = 1,56$ ,  $EP = 0,30$ ,  $p < 0,001$ ) (ver Figura 3). Durante a fase de busca na literatura, localizamos literatura cinzenta (ex. artigos de conferências e

relatórios de pesquisa) da forma mais abrangente possível, mas nossos esforços não resolveram esse problema. No entanto, identificamos quatro variáveis de controle: (a) duração da instrução (em semanas), (b) ano de publicação (média centralizada), (c) desenho do estudo (ou seja, ensaio clínico randomizado vs. estudo quase-experimental) e (d) confiabilidade do teste (ou seja, teste padronizado com confiabilidade relatada vs. teste desenvolvido pelo pesquisador sem confiabilidade relatada), as quais dizem respeito aos tipos de publicação dos estudos primários. Nós as incluímos em nossos modelos de meta-regressão para verificar os efeitos do viés de publicação (Lee et al., 2022; Peng et al., 2021). Os resultados não mostraram significância estatística, indicando que os tipos de publicação no conjunto de dados não influenciaram os resultados gerais. No entanto, considerando que um dos dois conjuntos de dados (ou seja, as amostras entre grupos) possivelmente apresentava um viés de publicação, evitamos fazer uma argumentação definitiva ao interpretar os resultados relacionados.

**Figura 3:** Resultados do teste de Egger para amostras dentro do grupo (esquerda) e entre grupos (direita)



### Codificando variáveis de moderação

Ao verificar o viés de publicação, identificamos características que podem demonstrar como os estudos coletados diferem entre si. Controlar estatisticamente essas variáveis (características do moderador), forneceria um ângulo mais confiável para os seus efeitos em

diferentes contextos (Woltman et al., 2012). Como discutido anteriormente, identificamos cinco características de moderadores: (a) idade dos alunos, (b) contextos de aprendizagem de línguas, (c) domínios da língua-alvo, (d) funções da assistência por IA e (e) tipos de sistemas de IA, todos os quais podem influenciar o impacto da aprendizagem de línguas individualizada com orientação de IA. A Tabela 3 apresenta as variáveis moderadoras codificadas.

Primeiro, os estudos incluíram participantes de diversas faixas etárias. Com base nas classificações de idade da Associação Médica Americana (NIH, 2022), categorizamos os participantes em três grupos: crianças (de 1 a 12 anos), adolescentes (de 13 a 17 anos) e adultos (18 anos ou mais).

Segundo, houve três tipos de contextos de aprendizagem de línguas: aprendizagem de língua materna (por exemplo, aprender inglês nos EUA), aprendizagem de segunda língua (por exemplo, aprender espanhol nos EUA) e aprendizagem de língua estrangeira (por exemplo, aprender inglês na Indonésia).

Terceiro, os focos do domínio da língua-alvo foram diversos, embora uma grande proporção dos tamanhos de efeito estivesse associada a habilidades de leitura e conhecimento de vocabulário, e um número menor de tamanhos de efeito estivesse associado a habilidades de produção oral, habilidades de compreensão oral, conhecimento gramatical e habilidades linguísticas gerais. Durante a fase de análise da meta-regressão, descobrimos que não poderíamos manter todas as habilidades individualmente para este moderador, pois isso tornava inestimáveis os modelos de regressão devido à falta de propriedade estatística (ou seja, muitas categorias, mas amostras muito pequenas). Assim, tivemos que agrupar alguns dos valores de acordo com seus tamanhos amostrais. Além disso, descobrimos que o conjunto de tamanhos de efeito entre grupos não incluía amostras relevantes para “habilidades de produção oral” e “habilidades linguísticas gerais”. Portanto, designamos três valores categóricos: (a) habilidades de leitura, (b) conhecimento de vocabulário e (c) outros. Este último valor categórico incluiu os casos de “conhecimento gramatical”, “habilidades de produção oral”, “habilidades de compreensão oral” e “habilidades linguísticas gerais”.

Quarto, o papel dos sistemas de IA teve três vertentes: alguns estudos empregaram ITS baseados em IA para fornecer instruções e feedback individualizados (ver Ma et al., 2014, para

uma revisão). Outro conjunto de estudos utilizou sistemas de IA para auxiliar professores a analisar dados de avaliação de seus alunos e recomendar a quantidade e o tipo adequados de instrução de letramento para cada estudante (isto é, gestão da aprendizagem; ver Deunk et al., 2018, para uma visão geral da abordagem de gestão individualizada da aprendizagem). O último tipo incluiu ferramentas individuais de aprendizagem de línguas equipadas com sistemas de aprendizagem adaptativa, como o Duolingo.

Por fim, os tipos de sistemas de IA utilizados em cada estudo primário foram diversos. Após revisar os estudos primários coletados para a metanálise, descobrimos que o fato de o sistema de IA ser projetado para ajustar suas regras e algoritmos manual ou automaticamente constituía uma das características mais distintivas entre os sistemas adotados. Quanto a outras características dos sistemas de IA, como os formatos de dados de entrada, os métodos de processamento ou os tipos de resultados, o escopo das variações não permitiu uma categorização que possibilitasse uma comparação objetiva das perspectivas teórica e prática (Witten et al., 2016). Por isso, classificamos os sistemas de IA em três categorias: (a) sistemas baseados em regras (sem ajustes após cada iteração), (b) sistemas de *machine learning* (aprendizagem de máquina) (onde a inteligência artificial ajusta o sistema automaticamente a cada iteração até atingir os objetivos), e (c) sistemas híbridos (com ajustes manuais feitos por programadores humanos com base nos resultados).

**Tabela 3:** Descrição das variáveis de moderação codificadas

| Variável moderadora                | # de valores entre grupos da amostra (k = 35) |
|------------------------------------|---|
| <i>Idade</i>                       |   |
| Crianças                           | 22  |
| Adolescentes                       | 4   |
| Adultos                            | 9   |
| <i>Contexto de aprendizagem</i>    |   |
| Aprendizagem da primeira língua    | 21  |
| Aprendizagem da segunda língua     | 2   |
| Aprendizagem de língua estrangeira | 12  |
| <i>Domínio da língua-alvo</i>      |   |
| Leitura                            | 20  |
| Vocabulário                        | 8   |
| Outros (Gramática)                 | 5   |
| (Produção oral)                    | 0   |
| (Compreensão oral)                 | 2   |

|                                       |    |
|---------------------------------------|----|
| (Habilidades gerais de linguagem)     | 0  |
| <i>Papel da assistência de IA</i>     |    |
| Sistema de tutoria inteligente        | 20 |
| Gestão de aprendizagem                | 7  |
| Ferramenta de aprendizagem individual | 8  |
| <i>Tipo de sistema de IA</i>          |    |
| Sistema baseado em regras             | 21 |
| Sistema de aprendizado de máquina     | 6  |
| Sistema híbrido                       | 8  |

Fonte: Elaboração própria (2024).

## Análise de Dados

Após calcular o tamanho dos efeitos e concluir o procedimento de codificação, as análises subsequentes dos dados consistiram em duas etapas: (a) sintetizar o tamanho dos efeitos calculados para computar o tamanho médio dos efeitos globais e (b) avaliar como os moderadores identificados influenciaram os efeitos médios globais. Para ambas as fases, utilizamos o STATA versão 16 (StataCorp, 2019) para empregar uma abordagem de metanálise multinível usando um modelo de efeitos aleatórios. Considerando que os conjuntos de dados tinham uma estrutura de três níveis (nível 1: resultado; nível 2: amostra; nível 3: projeto) e que os tamanhos do efeito (no nível 1) foram estimados com base em resultados derivados de diferentes populações (ou seja, sujeitos a erros-padrão), nossa abordagem foi estimar uma regressão de três níveis com um modelo de variância conhecida (ver Hox et al., 2018 para o embasamento teórico; ver Lee et al., 2019b, Lee et al., 2020, para diretrizes práticas e exemplos). Assim, na primeira etapa, estimamos um modelo de regressão de linha de base para identificar os tamanhos médios do efeito dos dois conjuntos de dados. Dado que o intercepto de um modelo de linha de base (ou seja, sem variáveis independentes) é um parâmetro para o tamanho do efeito total, o valor do intercepto representa o tamanho médio do efeito global (Raudenbush; Bryk, 2002).

Na segunda etapa, investigamos os efeitos da moderação para o segundo conjunto de dados (ou seja, 35 amostras entre grupos). Quando moderadores são incluídos como variáveis independentes no modelo de linha de base, todos os coeficientes do modelo são interpretados como diferenças entre os valores modelados e os de referência (Raudenbush; Bryk, 2002;

Woltman et al., 2012). Assim, incluímos cada variável moderadora em modelos de regressão separados (ou seja, modelos de regressão simples) para calcular os tamanhos médios do efeito para cada valor categórico dos moderadores (isto é, análise de subgrupos). Em seguida, cada variável moderadora foi incluída no modelo de linha de base para investigar os efeitos de moderação globais (ou seja, meta-regressão; Borenstein et al., 2021), após controlar por cada moderador, juntamente com as quatro variáveis de controle relacionadas aos tipos de publicação mencionadas anteriormente. Quando identificamos variáveis moderadoras com significância estatística nessa etapa, conduzimos outra análise de meta-regressão utilizando essas variáveis significativas para obter estimativas mais precisas e acuradas (Hox et al., 2018). Como os resultados da meta-regressão são geralmente considerados mais precisos do que os da análise de subgrupos, incluímos estes últimos no Apêndice devido às limitações de espaço.

## Resultados

### **Pergunta de Pesquisa #1. Até que ponto a aprendizagem individualizada guiada por IA é eficaz para o aprendizado de idiomas?**

A primeira pergunta de pesquisa foi respondida com base no conjunto de dados de amostras intragrupo, composto por 39 tamanhos de efeito provenientes de 26 amostras intragrupo ( $N = 2.262$ ), com base em 14 projetos. Esses efeitos foram calculados dividindo as diferenças médias entre os resultados dos testes pré e pós-intervenção pelos seus desvios padrão combinados. Além disso, eles representam o grau de desenvolvimento que os estudantes de idiomas alcançam por meio da aprendizagem individualizada guiada por IA. Assim como detalhado no gráfico de floresta (Figura 4), os efeitos variaram de 0,23 a 2,90 nas 26 amostras intragrupo. Embora dois deles tenham incluído zero em seus intervalos de confiança – indicando que esses estudos relataram efeitos nulos da aprendizagem individualizada guiada por IA –, a maioria das amostras mostrou contribuições positivas ( $dp > 0$ ) da tecnologia de IA no desenvolvimento linguístico.

Como mostra a Tabela 4, após sintetizar os efeitos dessas 26 amostras, constatamos que o tamanho médio global do efeito foi de 1,18 ( $EP = 0,17$ ,  $p < 0,001$ ). Isso indica que, em

média, a aprendizagem individualizada de idiomas guiada por IA contribuiu de forma estatisticamente significativa para o desenvolvimento linguístico, com um impacto considerado grande ( $d > 0,80$ ) de acordo com os parâmetros de interpretação do tamanho do efeito de Cohen.

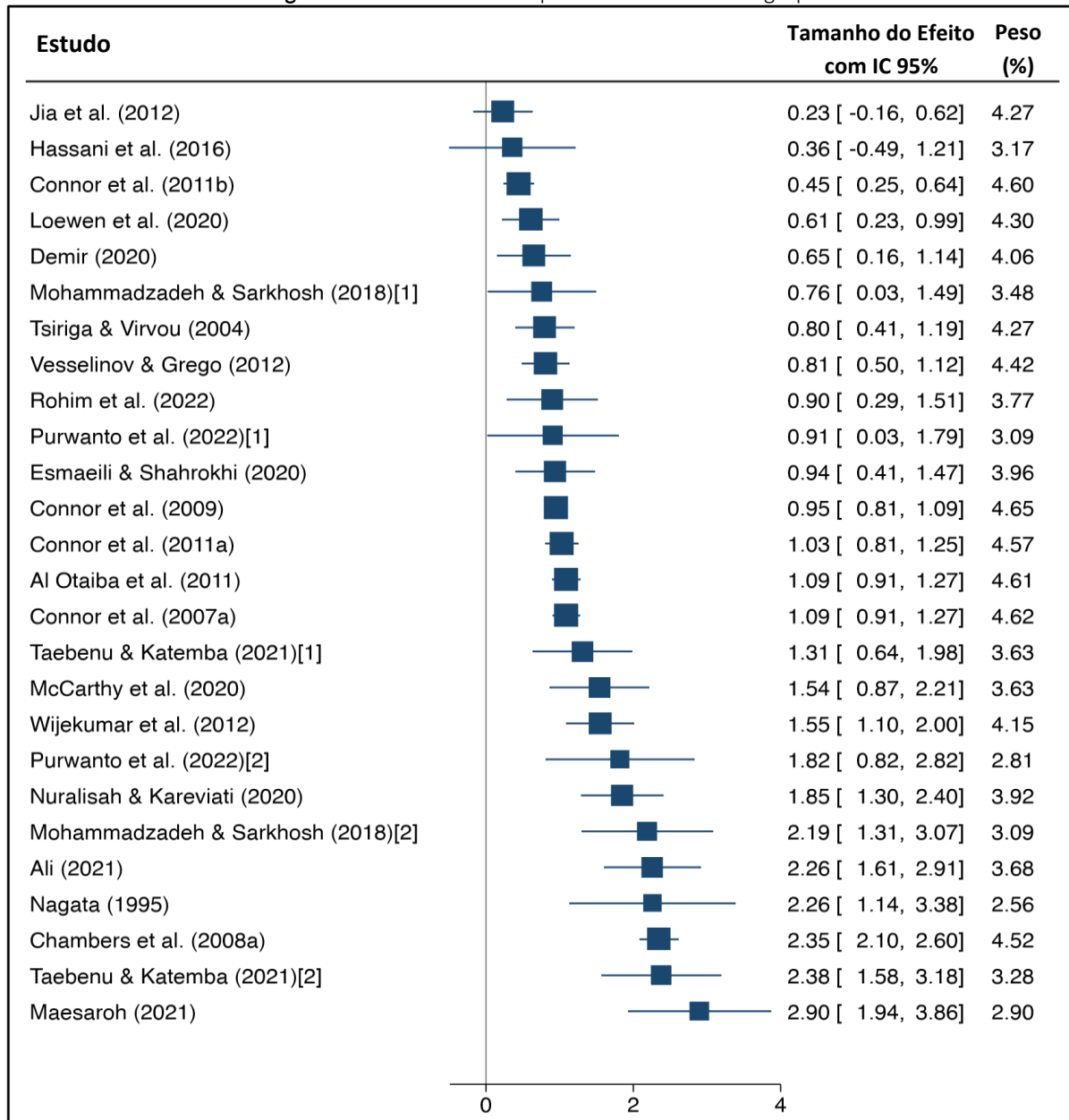
**Tabela 4:** Resultados da meta-análise para o tamanho médio do efeito intra-grupo (N = 2.262)

| # do Tamanho do Efeito | # da Amostra | # do Projeto | Média do Tamanho do Efeito | EP     |
|------------------------|--------------|--------------|----------------------------|--------|
| 39                     | 26           | 14           | 1,18***                    | (0,17) |

Nota: \*\*\*  $p < ,001$

Fonte: Elaboração própria (2024).

Figura 4: Gráfico de floresta para 26 amostras intra-grupo



Fonte: Elaboração própria (2024).

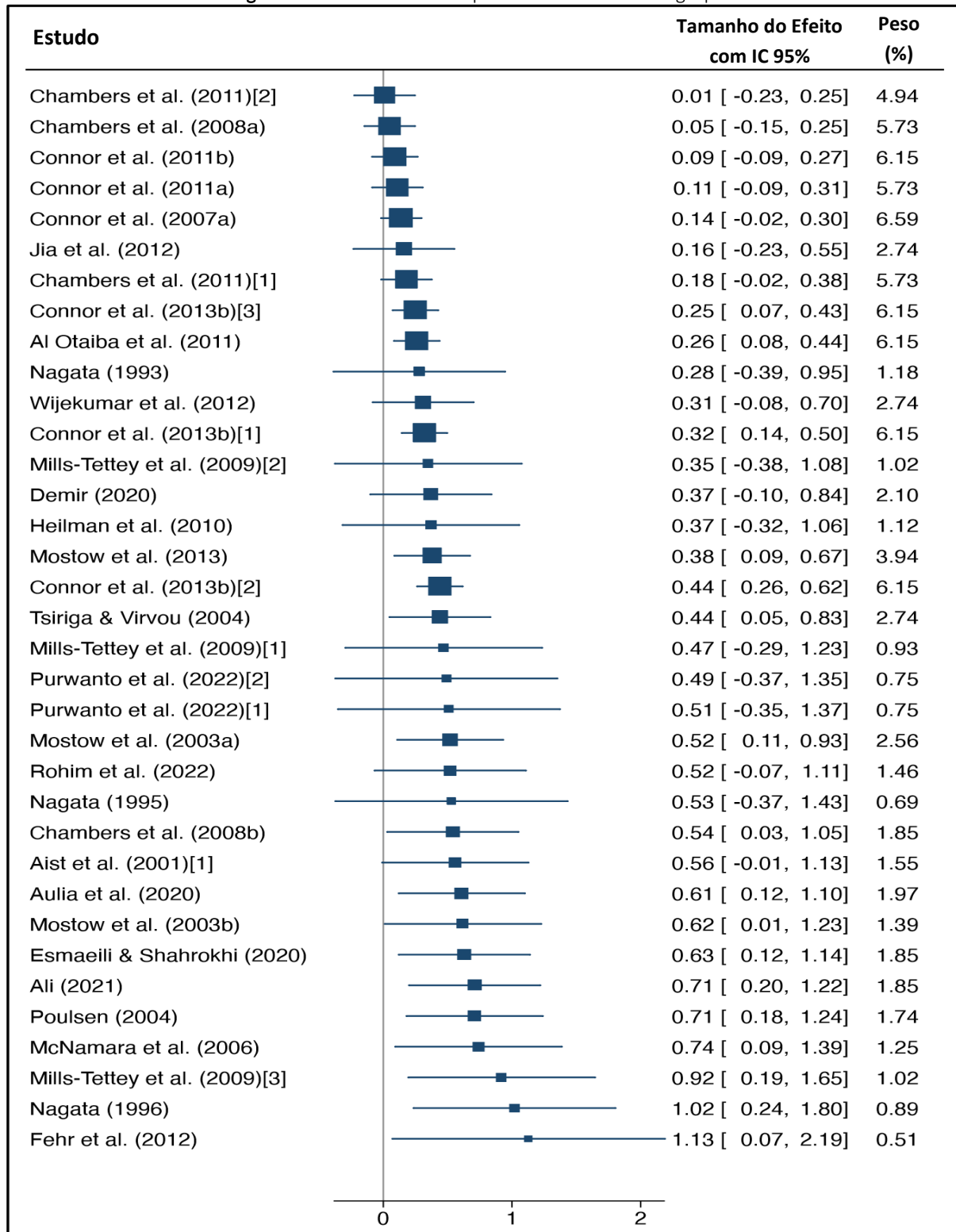
Pergunta de Pesquisa #2. Até que ponto a aprendizagem individualizada de idiomas guiado por IA apresenta um efeito de tratamento significativo em comparação com condições tradicionais de ensino?

Para responder à segunda pergunta de pesquisa, analisamos o conjunto de dados de amostras entre grupos, composto por 76 tamanhos de efeito de 35 amostras (N = 6.020), com

base em 13 projetos. Os tamanhos de efeito calculados a partir das amostras entre grupos representam os efeitos do tratamento da aprendizagem individualizada guiada por IA em comparação com condições tradicionais de ensino, conforme as diferenças padronizadas de médias entre os resultados pós-teste dos grupos de tratamento e dos grupos de controle.

Portanto, qualquer valor positivo indica a superioridade do aprendizado guiado por IA em relação às condições convencionais de aprendizagem de idiomas. Conforme detalhado no gráfico de floresta (Figura 5), os tamanhos de efeito calculados variaram de 0,01 a 1,13, e nenhuma das 35 amostras apresentou um tamanho de efeito negativo estatisticamente significativo. Esses resultados indicam que o aprendizado individualizado guiado por IA foi, no mínimo, tão eficaz quanto o aprendizado em condições convencionais. Aproximadamente metade (17 amostras) das amostras apresentou tamanhos de efeito positivos com significância estatística ( $dp > 0$ ), sugerindo que a aprendizagem individualizada guiada por IA pode ter sido mais eficaz no aprimoramento do desenvolvimento linguístico dos alunos do que as condições tradicionais de sala de aula.

Figura 5: Gráfico de floresta para 35 amostras entre grupos



Fonte: Elaboração própria (2024).

Como mostra a Tabela 5, a média geral dos tamanhos de efeito entre grupos foi de 0,39 (EP = 0,07,  $p < 0,001$ ), indicando que, em média, aprendizagem individualizada de idiomas guiada por IA teve um impacto educacional significativamente maior do que os métodos convencionais (ou seja, um efeito de tratamento positivo). Além disso, o tamanho médio do efeito do tratamento foi considerado pequeno ( $0,20 \leq d < 0,50$ ), de acordo com os parâmetros de interpretação do  $d$  de Cohen.

**Tabela 5:** Resultados da metanálise para o tamanho médio do efeito entre grupos (N = 6.020)

| # do Tamanho do Efeito | # da Amostra | # do Projeto | Média do Tamanho do Efeito | EP     |
|------------------------|--------------|--------------|----------------------------|--------|
| 76                     | 35           | 13           | 0.39***                    | (0,07) |

Nota: \*\*\*  $p < ,001$

Fonte: Elaboração própria (2024).

**Pergunta de Pesquisa #3.** Quais são os fatores importantes dos estudantes e os contextuais que influenciam os efeitos do tratamento da aprendizagem de idiomas individualizada guiada por IA? Em que medida os vários papéis e algoritmos da IA se relacionam com seus efeitos de tratamento?

Para responder à terceira questão de pesquisa, realizamos uma análise de regressão múltipla (também conhecida como meta-regressão), incluindo cinco variáveis moderadoras alinhadas com nossos interesses e quatro variáveis de controle relacionadas aos tipos de publicação (ver Tabela 6). Esse procedimento teve como objetivo investigar se existiam diferenças estatisticamente significativas entre os valores categóricos das variáveis moderadoras. Conforme mencionado na seção de Análise de Dados, os resultados da nossa análise de subgrupos, como parte dessa análise de moderadores, estão incluídos no Apêndice.

**Tabela 6:** Resultados da análise de moderador (meta-regressão) para amostras entre grupos (N = 6.020)

| Variável independente                 | Análise de regressão múltipla (meta-regressão) |       |                                     |        |
|---------------------------------------|--|-------|-------------------------------------|--------|
|                                       | Coefficiente                                   | EP    | Média do Tamanho do efeito ajustado | EP     |
| <i>Idade</i>                          |  |       |                                     |        |
| Crianças                              | 0,00   | (,)   | 0,45***                             | (0,07) |
| Adolescentes                          | 0,01   | -0,26 | 0,46*                               | (0,22) |
| Adultos                               | -0,4   | -0,32 | 0,05                                | (0,27) |
| <i>Contexto de aprendizagem</i>       |  |       |                                     |        |
| Aprendizagem da primeira língua       | 0,00   | (,)   | 0,31***                             | (0,09) |
| Aprendizagem da segunda língua        | 0,01   | -0,21 | 0,32                                | (0,19) |
| Aprendizagem de língua estrangeira    | 0,45   | -0,4  | 0,76*                               | (0,33) |
| <i>Domínio da língua-alvo</i>         |  |       |                                     |        |
| Compreensão escrita                   | 0,00   | (,)   | 0,35***                             | (0,10) |
| Vocabulário                           | 0,29   | -0,34 | 0,63*                               | (0,26) |
| Outros (Gramática)                    | 0,03   | -0,49 | 0,37                                | (0,41) |
| <i>Papel da assistência de IA</i>     |  |       |                                     |        |
| Sistema de tutoria inteligente        | 0,00   | (,)   | 0,44***                             | (0,09) |
| Gestão de aprendizagem                | 0,06   | -0,09 | 0,50***                             | (0,11) |
| Ferramenta de aprendizagem individual | -0,46  | -0,57 | -0,02                               | (0,50) |
| <i>Tipo de sistema de IA</i>          |  |       |                                     |        |
| Sistema baseado em regras             | 0,00   | (,)   | 0,28***                             | (0,06) |
| Sistema de aprendizado de máquina     | 0,42   | -0,37 | 0,70*                               | (0,34) |
| Sistema híbrido                       | 0,23*  | -0,1  | 0,51***                             | (0,08) |

Nota: \*\*\* p < ,001, \* p < ,05

Fonte: Elaboração própria (2024).

Os resultados indicaram que apenas uma variável moderadora apresentou significância estatística: o tipo de sistema de IA. Entre os três valores categóricos, a aprendizagem individualizada de idiomas guiada por IA com sistemas híbridos apresentou um coeficiente estatisticamente significativo de 0,23 (EP = 0,10, p = 0,017) quando comparado com sistemas baseados em regras (nível de referência).

**Tabela 7:** Resultados da meta-regressão de acompanhamento com moderador significativo para amostras entre grupos (N = 6.020)

| Variável independente             | Coefficiente | EP     | Média do Tamanho do efeito ajustado | EP     |
|-----------------------------------|--------------|--------|-------------------------------------|--------|
| <i>Tipo de sistema de IA</i>      |              |        |                                     |        |
| Sistema baseado em regras         | 0,00         | (,)    | 0,29***                             | (0,05) |
| Sistema de aprendizado de máquina | 0,30         | (0,24) | 0,59**                              | (0,22) |
| Sistema híbrido                   | 0,22**       | (0,08) | 0,51***                             | (0,07) |

Nota: \*\*\*  $p < ,001$ , \*\*  $p < ,01$

Fonte: Elaboração própria (2024).

Para obter estimativas mais precisas, conduzimos uma análise de meta-regressão de acompanhamento. Os resultados são mostrados na Tabela 7. Semelhante ao modelo completo de meta-regressão, a aprendizagem individualizada de idiomas guiada por IA usando sistemas híbridos teve um coeficiente estatisticamente significativo de 0,22 (EP = 0,08,  $p = 0,007$ ). Adicionalmente, o tipo de sistema de aprendizado de máquina teve um coeficiente maior ( $\beta = 0,30$ ) sem significância estatística; entretanto, não alcançou significância estatística devido ao seu grande erro-padrão (EP = 0,24), que presumivelmente ocorreu devido ao seu tamanho amostral relativamente pequeno e estrutura de dados complexa (isto é, seis tamanhos de efeito aninhados em seis amostras de dois projetos). De acordo com as estimativas ajustadas do tamanho médio do efeito, identificamos que ambos os sistemas híbridos ( $d = 0,51$ , EP = 0,07) e sistemas de aprendizado de máquina ( $d = 0,59$ , EP = 0,22) tiveram tamanhos médios de efeito moderados, enquanto sistemas baseados em regras tiveram tamanho médio de efeito pequeno ( $d = 0,29$ , EP = 0,05).

## DISCUSSÃO

Esta metanálise teve como objetivo sintetizar os resultados quantitativos de estudos empíricos primários que adotaram a aprendizagem individualizada de idiomas guiada por IA para facilitar a aprendizagem de línguas pelos alunos. Contribuímos para este ramo de pesquisa nos seguintes dois aspectos: (a) a eficácia geral e os efeitos positivos do tratamento ao usar a aprendizagem individualizada de idiomas guiada por IA e (b) as diretrizes sobre como os sistemas de aprendizagem de idiomas guiados por IA podem ser usados com maior eficácia. Em conjunto, o ensino de idiomas individualizado guiado por IA, se equipado com currículos e

materiais de aprendizagem bem desenvolvidos (elaborados por especialistas humanos) e modelos preditivos altamente precisos (seja por regras pré-definidas, big data coletados pela própria máquina ou ambos), pode ser utilizado com maior eficácia para determinados tipos de alunos em ambientes de aprendizagem específicos. Por exemplo, considerando alunos que têm dificuldade em acessar um ambiente educacional adequado, como quando as escolas são tão remotas que há poucos professores do idioma-alvo, esse método seria o mais eficaz. As seções a seguir delineiam os dois principais achados deste estudo.

Primeiro, este estudo metanalítico fornece evidências robustas e sumárias da eficácia geral da individualização guiada por IA no domínio da aprendizagem de línguas. Nossos resultados em relação às amostras intragrupo revelaram que a aprendizagem individualizada de idiomas guiada por IA pode ser considerada como tendo um tamanho médio de efeito grande (ou seja,  $d \geq 0,80$ ) ( $d = 1,18$ ). Essa descoberta está alinhada com a premissa geral da teoria sociocultural (Bruner, 1983; Vygotsky, 1978), que ressalta a importância de avaliar o estágio atual de cada aluno e oferecer instrução e suporte individualizados, corroborando a percepção positiva geral quanto à integração de tecnologias de IA na educação (Lee; Qiufan, 2021). Quando comparado com ambientes convencionais de aprendizagem de línguas, um tamanho médio de efeito pequeno (ou seja,  $0,20 \leq d < 0,50$ ) ( $d = 0,39$ ) foi observado para as amostras entre grupos, confirmando que o aprendizado de idiomas individualizado guiado por IA foi pelo menos tão eficaz quanto a aprendizagem em condições convencionais.

Portanto, os achados do presente estudo contribuem para dissipar o ceticismo existente sobre a individualização por meio de tecnologias de IA na educação (Holmes; Tuomi, 2022). Consequentemente, formuladores de políticas e desenvolvedores de materiais no âmbito do ensino de línguas devem considerar o desenvolvimento e a adoção da aprendizagem individualizada guiada por IA em seus contextos. Por exemplo, ao elaborar um currículo para desenvolvimento de leitura, os desenvolvedores de materiais podem consultar programas existentes de individualização guiada por IA (como a estrutura A2i de Connor et al., 2007; 2009; 2011a, 2011b) e incorporar componentes relevantes em seus projetos. Além disso, devem explorar continuamente como os sistemas e aplicações de aprendizagem de línguas guiada por

IA podem complementar o aprendizado dos alunos quando há escassez de professores e recursos educacionais em sala de aula.

Segundo, outra contribuição deste estudo vem dos resultados das análises de moderadores, que nos permitiram obter uma compreensão mais profunda dos tipos de tecnologia de IA que poderiam ajudar a aprendizagem individualizada de idiomas com mais eficácia. Nossa metanálise demonstrou que, entre os diferentes tipos de sistemas (ou seja, sistemas baseados em regras, sistemas de aprendizado de máquina ou híbridos), a aprendizagem individualizada de idiomas guiada por IA poderia produzir o maior efeito de tratamento com um sistema de algoritmo híbrido que incorpora o envolvimento humano em sistemas de algoritmo de aprendizado de máquina. O Projeto LISTEN (por exemplo, Beck et al., 2003; Mostow, 2016; Mostow et al., 2013) é um sistema de IA exemplar que combina sistemas baseados em regras e de aprendizado de máquina. Seus dados coletados, incluindo uma infinidade de palavras ouvidas pelo reconhecedor de fala e os comportamentos de clique dos alunos, foram usados para adquirir modelos preditivos como parte de uma abordagem de aprendizado de máquina (por exemplo, Beck; Woolf, 2000). Enquanto isso, os sistemas de IA foram atualizados por pesquisadores humanos com base nos resultados dos testes de campo (por exemplo, González-Brenes; Mostow, 2011). Com base nessa descoberta, propomos que a experiência e os conhecimentos teóricos dos professores possam ser usados como lógica orientadora para tomar decisões com base em inferências feitas por algoritmos de aprendizado de máquina a partir de grandes quantidades de dados do aluno, porque o componente humano desempenha um papel crucial no sistema híbrido, que tira proveito de ambos os sistemas (Lessware, 2021). Dessa forma, quando o sistema híbrido fornece recomendações individualizadas de instruções exclusivas para atender às necessidades específicas de cada aluno, os professores podem não apenas entender como essa recomendação é feita e como ela ajudaria seus alunos, mas os algoritmos de aprendizado de máquina também podem garantir que a recomendação seja prevista com a mais alta precisão.

Além da superioridade do sistema híbrido, demonstramos que a aprendizagem individualizada de idiomas guiada por IA apresentou resultados positivos gerais de aprendizagem, independentemente do tipo de sistema (ou seja, sistemas baseados em regras,

sistemas de aprendizagem de máquina ou híbrido). Portanto, sugerimos que os professores e alunos usem cada tipo de sistema de algoritmo de IA de acordo com seus objetivos de ensino e aprendizagem. Por exemplo, os sistemas de IA de aprendizado de máquina geralmente tendem a ter um poder preditivo mais preciso do que os sistemas baseados em regras, enquanto os sistemas baseados em regras exigem que os seres humanos sugiram regras interpretáveis para a tomada de decisões em sistemas de algoritmos. Mais especificamente, embora os sistemas de IA de aprendizagem de máquina tenham, em média, efeitos de tratamento mais significativos no aprimoramento da aprendizagem de idiomas do que os sistemas de IA baseados em regras, os últimos têm benefícios específicos, como o fato de podermos entender e interpretar seu mecanismo, que consiste em regras predeterminadas por humanos. Por exemplo, para aplicativos como o Duolingo, que se baseia em um sistema de algoritmo de aprendizagem de máquina, os usuários querem receber as lições mais adequadas para seu progresso de aprendizagem em vez de entender por que determinadas lições são dadas a eles. Portanto, o uso de sistemas de IA de aprendizado de máquina nesse caso seria apropriado do ponto de vista dos alunos. Por outro lado, em um ambiente de sala de aula, o software A2i de Connor, que se utiliza de sistemas baseados em regras, recomenda que os professores forneçam lições específicas para determinados grupos de alunos e, naturalmente, os professores podem entender a lógica por detrás de tais recomendações e interpretá-las para suas futuras instruções; portanto, os sistemas baseados em regras podem ser mais úteis do ponto de vista dos professores.

### **Direcionamento e limitações para pesquisas futuras**

Além das principais contribuições deste estudo, nossa metanálise apresenta algumas limitações. Em primeiro lugar, embora tenhamos tentado localizar e revisar o maior número possível de estudos relevantes por meio de diversos canais, existe a possibilidade de estudos importantes não identificados (daí o indicativo de viés de publicação em nosso conjunto de amostras entre grupos, conforme discutido anteriormente); alguns exemplos são trabalhos não publicados com resultados não significativos, nulos ou até negativos. Futuros esforços

metanalíticos devem incluir um número maior de estudos empíricos, o que permitiria examinar uma gama mais ampla de moderadores (por exemplo, ambientes de aprendizagem de línguas [como laboratório, em sala de aula, fora da sala] e tipos de hardware de IA [como smartphones, computadores, robôs]). Em segundo lugar, considerando que a operacionalização da IA é diversa entre pesquisadores, técnicos de informática, engenheiros e leigos (por exemplo, Hechler et al., 2020), reconhecemos que nossa operacionalização da tecnologia guiada por IA pode não abranger todas as perspectivas existentes sobre o que constitui verdadeiramente “IA” no contexto educacional. Essa limitação conceitual pode afetar a generalização de nossos resultados.

Em terceiro lugar, como mencionado na Introdução, não consideramos estudos sobre aprendizagem de idiomas com chatbots, robôs sociais ou AWE (Avaliação Automatizada de Escrita, sigla em inglês). Entretanto, isso não significa necessariamente que essas tecnologias não possam ser usadas para aprendizagem individualizada de idiomas (por exemplo, chatbots ou robótica desenvolvidos para fins gerais podem ser aplicados a contextos de aprendizagem de idiomas de acordo com o desenho de estudo exclusivo dos instrutores). Futuras metanálises podem estender nosso esforço considerando estudos sobre os efeitos de tais tecnologias voltadas para instrução individualizada. Finalmente, houve algum desequilíbrio no número de estudos primários associados aos 17 projetos identificados, com alguns projetos (por exemplo, Project LISTEN, A2i), tendo sido mais extensivamente pesquisados (isto é, usando mais amostras) do que outros (por exemplo, ITS iSTART), embora tenhamos tentado resolver esse desequilíbrio conduzindo uma metanálise de três níveis.

Com base nos resultados do presente estudo, propomos as seguintes três direções para pesquisas futuras sobre aprendizagem individualizada de idiomas guiada por IA: Primeiro, com os ITS sendo os mais amplamente categorizados (por exemplo, Heilman et al., 2010; McNamara et al., 2006; Mills-Tetty et al., 2009; Poulsen, 2004), parece necessário examinar com mais detalhe os aspectos dos ITS que facilitam a aprendizagem de idiomas, idealmente em conjunto com dados qualitativos (por exemplo, examinando padrões de usuários). Segundo, o papel dos instrutores humanos merece avaliação mais próxima dentro da ampla gama de emprego de tecnologia de IA, particularmente em vista do ceticismo existente sobre

o papel da IA na educação (por exemplo, Chan; Tsi, 2023; Tao et al., 2019), incluindo desenvolver e projetar currículos e gerenciar as regras de atribuição de material de aprendizagem. Uma questão empírica exemplar seria se o envolvimento de instrutores humanos (ou especialistas) seria mais eficaz do que sua ausência em termos de utilizar os dados obtidos da avaliação formativa para instrução posterior. Terceiro, pesquisas futuras sobre essa questão devem refinar a taxonomia da IA, com base na qual implicações concretas podem ser fornecidas para programadores e pesquisadores de tecnologia baseada em IA voltada para instrução individualizada.

### Agradecimentos

Este trabalho foi apoiado pelo Instituto de Pesquisa Hwarangdae da Academia Militar da Coreia (2023B1001). Gostaríamos de agradecer aos revisores por seus valiosos comentários.

### Referências

AIST, G. et al. Computer-assisted oral reading helps third graders learn vocabulary better than a classroom control—About as well as one-on-one human-assisted oral reading. In: MOORE, J. D.; REDFIELD, C. L.; JOHNSON, W. L. (Eds.). *Artificial intelligence in education: AI-ED in the wired and wireless future*. IOS Press, 2001. p. 267–277.

AL OTAIBA, S. et al. Assessment data–informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal*, v. 111, n. 4, p. 535–560, 2011. <https://doi.org/10.1086/659031>

ALI, N. The effect of Duolingo usage on students' vocabulary mastery of seventh-grade students at MTs PAB 1 Helvetia. *Journal Serunai Bahasa Inggris*, v. 13, n. 1, p. 23–30, 2021. <https://doi.org/10.37755/jsbi.v13i1.347>

AULIA, H. R.; WAHJUNINGSIH, E.; ANDAYANI, R. The effect of Duolingo application on students' English vocabulary mastery. *English Language Teaching Research Journal*, v. 4, n. 2, p. 29–37, 2020. <https://doi.org/10.37147/eltr.v4i2.71>

BECK, J. E. et al. Can automated questioning help children's reading comprehension? [Apresentação de trabalho] 10th International *Conference on Artificial Intelligence in Education*, 2003, julho 20-24.

BECK, J. E.; WOOLF, B. P. High-level student modeling with machine learning. In: GAUTHIER, G.; FRASSON, C.; VANLEHN, K. (Eds.). *Intelligent tutoring system: Lecture notes in Computer Science*. Springer, 2000. p. 584–593.

BLEY-VROMAN, R. The logical problem of foreign language learning. *Linguistic Analysis*, v. 20, n. 1, p. 3–49, 1990.

BLEY-VROMAN, R. The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, v. 31, n. 2, p. 175–198, 2009. <https://doi.org/10.1017/S0272263109090275>

BORENSTEIN, M. et al. *Introduction to meta-analysis*. 2.<sup>a</sup> ed. John Wiley & Sons, 2021.

BRUNER, J. *Child's talk: Learning to use language*. Oxford University Press, 1983.

CHAMBERS, B. et al. Computer-assisted tutoring in success for all: Reading outcomes for first graders. *Journal of Research on Educational Effectiveness*, v. 1, n. 2, p. 120–137, 2008a. <https://doi.org/10.1080/19345740801941357>

CHAMBERS, B. et al. Technology infusion in Success for All: Reading outcomes for first graders. *The Elementary School Journal*, v. 109, n. 1, p. 1–15, 2008b. <https://doi.org/10.1086/592364>

CHAMBERS, B. et al. Small-group, computer-assisted tutoring to improve reading outcomes for struggling first and second graders. *The Elementary School Journal*, v. 111, n. 4, p. 625–640, 2011. <https://doi.org/10.1086/659035>

CHAN, C. K. Y.; TSI, L. H. Y. The AI revolution in education: Will AI replace or assist teachers in higher education? arXiv, 2023. <https://doi.org/10.48550/arXiv.2305.01185>

CHASSIGNOL, M. et al. Artificial Intelligence trends in education: A narrative overview. *Procedia Computer Science*, v. 136, p. 16–24, 2018. <https://doi.org/10.1016/j.procs.2018.08.233>

CHEN, L. et al. Artificial intelligence in Education: A review. *IEEE Access*, v. 8, p. 75264–75278, 2020. <https://doi.org/10.1109/ACCESS.2020.2988510>

CONNOR, C. M. et al. Algorithm-guided individualized reading instruction. *Science*, v. 315, n. 5811, p. 464–465, 2007. <https://doi.org/10.1126/science.1134513>

CONNOR, C. M. et al. Testing the impact of child characteristics × instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, v. 46, n. 3, p. 189–221, 2011a. <https://doi.org/10.1598/RRQ.46.3.1>

CONNOR, C. M. et al. Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. *Journal of*

*Research on Educational Effectiveness*, v. 4, n. 3, p. 173–207, 2011b. <https://doi.org/10.1080/19345747.2010.510179>

CONNOR, C. M. et al. A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science*, v. 24, n. 8, 1408–1419. <https://doi.org/10.1177/0956797612472204>

CONNOR, C. M. et al. Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development*, v. 80, n. 1, p. 77–100, 2009. <https://doi.org/10.1111/j.1467-8624.2008.01247.x>

DEBRAY, T. P.; MOONS, K. G.; RILEY, R. D. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival a comparison of new and existing tests. *Research Synthesis Methods*, v. 9, n. 1, p. 41–50, 2018. <https://doi.org/10.1002/jrsm.1266>

DEMIR, Ü. The effect of using negative knowledge-based intelligent tutoring system evaluator software to the academic success in English language education. *Pedagogies: An International Journal*, v. 15, n. 4, p. 245–261, 2020. <https://doi.org/10.1080/1554480X.2019.1706522>

DEUNK, M. I. et al. Effective differentiation practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, v. 24, p. 31–54, 2018. <https://doi.org/10.1016/j.edurev.2018.02.002>

DEWAELE, J.-M. Individual differences in second language acquisition. In: RITCHIE, W. C.; BHATIA, T. K. (Eds.). *The new handbook of second language acquisition*. Emerald, 2009. p. 623–646.

DU BOULAY, B. Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems*, v. 31, n. 6, p. 76–81, 2016. <https://doi.org/10.1109/MIS.2016.93>

EGGER, M. et al. Bias in meta-analysis detected by a simple, graphical test. *The BMJ*, v. 315, n. 7109, p. 629–634, 1997. <https://doi.org/10.1136/bmj.315.7109.629>

ESMAEILI, Z.; SHAHROKHI, M. The impact of Memrise application on Iranian EFL learners' collocation learning and retention. *International Journal of Language Education*, v. 4, n. 2, p. 221–233, 2020. <https://doi.org/10.26858/ijole.v4i2.10672>

FEHR, C. N. et al. The effects of individualized, online vocabulary instruction on picture vocabulary scores: An efficacy study. *Computer Assisted Language Learning*, v. 25, n. 1, p. 87–102, 2012. <https://doi.org/10.1080/09588221.2011.586640>

FLOREA, A. M.; RADU, S. Artificial intelligence and education. In: *2019 22nd International Conference on Control Systems and Computer Science (CSCS)* (pp. 381–382), 2019, maio 28–30. IEEE. <https://doi.org/10.1109/CSCS.2019.00069>

GHUMRA, F. *OpenAI GPT-3, the most powerful language model: An overview*. e-Infochips, 2022, março 30. <https://www.einfochips.com/blog/openai-gpt-3-the-most-powerful-language-model-an-overview/>

GMI. *AI in education market size & share, growth forecast 2022-2030*. Global Market Insights Inc., 2022. <https://www.gminsights.com/industry-analysis/artificial-intelligence-ai-in-education-market>

GODWIN-JONES, R. Big data and language learning: Opportunities and challenges. *Language Learning & Technology*, v. 25, n. 1, p. 4–19, 2021. <http://hdl.handle.net/10125/44747>

GONZÁLEZ-BRENES, J. P.; MOSTOW, J. Classifying dialogue in high-dimensional space. *ACM Transactions on Speech and Language Processing*, v. 7, n. 3, p. 1–15, 2011. <https://doi.org/10.1145/1966407.1966413>

HASSANI, K. et al. Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interactive Learning Environments*, v. 24, n. 1, p. 252–271, 2016. <https://doi.org/10.1080/10494820.2013.846265>

HECHLER, E. et al. *Deploying AI in the enterprise: IT approaches for design, devops, governance, change management, blockchain, and quantum computing*. Springer, 2020.

HEDGES, L. V. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, v. 6, n. 2, p. 107–128, 1981. <https://doi.org/10.3102/10769986006002107>

HEILMAN, M. et al. Personalization of reading passages improves vocabulary acquisition. *International Journal of Artificial Intelligence in Education*, v. 20, n. 1, p. 73–98, 2010. <https://doi.org/10.3233/JAI-2010-0003>

HOLMES, W.; TUOMI, I. State of the art and practice in AI in education. *European Journal of Education*, v. 57, n. 4, p. 542–570, 2022. <https://doi.org/10.1111/ejed.12533>

HOX, J. J.; MOERBEEK, M.; VAN DE SCHOOT, R. *Multilevel analysis: Techniques and applications*. 3. ed. Routledge, 2018.

HUANG, W.; HEW, K. F.; FRYER, L. K. Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, v. 38, n. 1, p. 237–257, 2022. <https://doi.org/10.1111/jcal.12610>

HWANG, S. Examining the effects of artificial intelligence on elementary students' mathematics achievement: A meta-analysis. *Sustainability*, v. 14, n. 20, p. 13185, 2022. <https://doi.org/10.3390/su142013185>

INTERNATIONAL READING ASSOCIATION. *Making a difference means making it different: Honoring children's rights to excellent reading instruction*. International Reading Association, 2000.

JIA, J. et al. Effects of a vocabulary acquisition and assessment system on students' performance in a blended learning class for English subject. *Computers & Education*, v. 58, n. 1, p. 63–76, 2012. <https://doi.org/10.1016/j.compedu.2011.08.002>

JOHNSON, K. The “general” study of expertise. In: JOHNSON, K. (Ed.). *Expertise in second language learning and teaching*. Palgrave Macmillan, 2005. p. 11–33.

JOHNSON, R. K.; SWAIN, M. From core to content: Bridging the L2 proficiency gap in late immersion. *Language and Education*, v. 8, n. 4, p. 211–229, 1994. <https://doi.org/10.1080/09500789409541392>

KANDLHOFER, M. et al. Artificial intelligence and computer science in education: From kindergarten to university. In: *2016 IEEE Frontiers in Education Conference*, Erie, PA, 12 out. 2016. p. 1–9. IEEE. <https://doi.org/10.1109/FIE.2016.7757410>

KESSLER, G. Computer Assisted Language Learning. In: HINKEL, E. (Ed.). *Handbook of practical second language teaching and learning*. New York: Routledge, 2023. p. 173–183.

KUSUMADEWI, H.; WIDYASTUTI, M. The effects of using Duolingo towards students' vocabulary mastery. *Indonesian Journal of English Teaching*, v. 7, n. 2, p. 172–186, 2018. <https://doi.org/10.15642/ijet2.2018.7.2.172-186>

LEE, H. et al. The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, v. 33, n. 2, p. 124–140, 2020. <https://doi.org/10.1080/08957347.2020.1732383>

LEE, H. et al. Simple view of second language reading: A meta-analytic structural equation modeling approach. *Scientific Studies of Reading*, v. 26, n. 6, p. 585–603, 2022. <https://doi.org/10.1080/10888438.2022.2087526>

LEE, H. et al. Advancing CALL research via data mining techniques: Unearthing hidden groups of learners in a corpus-based L2 vocabulary learning experiment. *ReCALL*, v. 31, n. 2, p. 135–149, 2019. <https://doi.org/10.1017/S0958344018000162>

LEE, H. et al. The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, v. 40, n. 5, p. 721–753, 2019. <https://doi.org/10.1093/applin/amy012>

LEE, H.; LEE, J. H. Social robots for English language teaching. *ELT Journal*, v. 76, n. 1, p. 119–124, 2022. <https://doi.org/10.1093/elt/ccab041>

LEE, J. H.; LEE, H. Teachers' verbal lexical explanation for second language vocabulary learning: A meta-analysis. *Language Learning*, v. 72, n. 2, p. 576–612, 2022. <https://doi.org/10.1111/lang.12493>

LEE, J. H.; MACARO, E. Investigating age in the use of L1 or English-only instruction: Vocabulary acquisition by Korean EFL learners. *The Modern Language Journal*, v. 97, n. 4, p. 887–901, 2013. <https://doi.org/10.1111/j.1540-4781.2013.12044.x>

LEE, K.-F.; QIUfan, C. *AI 2041: Ten visions for our future*. London: WH Allen, 2021.

LESSWARE, S. *Why a rules based plus a machine learning hybrid approach provides powerful artificial intelligence for location master data management*. 1Spatial, 26 jul. 2021. Disponível em: <https://1spatial.com/news-events/2021/why-a-rules-based-plus-a-machine-learning-hybrid-approach/>. Acesso em: 22 abr. 2025.

LIN, R. et al. Exploring the effectiveness and moderators of artificial intelligence in the classroom: A meta-analysis. In: YANG, J. et al. (Ed.). *Resilience and Future of Smart Learning*. Springer, 2022. p. 61–66.

LOEWEN, S.; ISBELL, D. R.; SPORN, Z. The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, v. 53, n. 2, p. 209–233, 2020. <https://doi.org/10.1111/flan.12454>

LOVELESS, T.; PARKAS, S.; DUFFETT, A. *High-achieving students in the era of NCLB*. Thomas B. Fordham Institute, 2008.

LUND, B. D.; WANG, T. Chatting about ChatGPT: how may AI and GPT impact academic and libraries? *Library Hi Tech News*, v. 40, n. 3, p. 26–29, 2023. <https://doi.org/10.1108/LHTN-01-2023-0009>

MA, W.; ADESOPE, O. O.; NESBIT, J. C.; LIU, Q. Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, v. 106, n. 4, p. 901–918, 2014. <https://doi.org/10.1037/a0037123>

MAESAROH, A. M. The effectiveness of using “Busuu” in teaching speaking. *Journal Fakultas Keguruan & Ilmu Pendidikan*, v. 2, n. 3, p. 94–98, 2021. Disponível em: <https://www.jurnal.unisa.ac.id/index.php/jfkip/article/view/136>. Acesso em: [colocar a data de acesso, se necessário].

MARTÍNEZ, I. G. et al. Analysing the impact of artificial intelligence and computational sciences on student performance: Systematic review and meta-analysis. *NAER: Journal of New Approaches in Educational Research*, v. 12, n. 1, p. 171–197, 2023. <https://doi.org/10.7821/naer.2023.1.1240>

MCCARTHY, K. S. et al. Improving reading comprehension in Spanish using iSTART-E: A pilot study. *International Journal of Computer-Assisted Language Learning and Teaching*, v. 10, n. 4, p. 66–82, 2020. <https://doi.org/10.4018/IJCALLT.2020100105>

MCNAMARA, D. S. et al. Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, v. 34, n. 2, p. 147–171, 2006. <https://doi.org/10.2190/1RU5-HDTJ-A5C8-JVWE>

MILLS-TETTEY, G. A. et al. Improving child literacy in Africa: Experiments with an automated reading tutor. In: INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES AND DEVELOPMENT (ICTD), 2009, [S.l.]. Proceedings... IEEE, 2009. p. 129–138. <https://doi.org/10.1109/ICTD.2009.5426715>

MOHAMMADZADEH, A.; SARKHOSH, M. The effects of self-regulatory learning through computer-assisted intelligent tutoring system on the improvement of EFL learner's speaking ability. *International Journal of Instruction*, v. 11, n. 2, p. 167–184, 2018. <https://doi.org/10.12973/iji.2018.11212a>

MOSTOW, J. et al. Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research*, v. 29, n. 1, p. 61–117, 2003a. <https://doi.org/10.2190/06AX-QW99-EQ5G-RDCF>

MOSTOW, J. et al. 4-month evaluation of a learner-controlled reading tutor that listens. In: HOLLAND, V. M.; FISHER, F. P. (Ed.). *The path of speech technologies in computer assisted language learning: From research toward practice*. Routledge, 2003b. p. 201–219.

MOSTOW, J.; NELSON-TAYLOR, J.; BECK, J. E. Computer-guided oral reading versus independent practice: Comparison of sustained silent reading to an automated reading tutor that listens. *Journal of Educational Computing Research*, v. 49, n. 2, p. 249–276, 2013. <https://doi.org/10.2190/EC.49.2.g>

MOSTOW, J. Project LISTEN's Reading Tutor. In: CROSSLEY, S. A.; MCNAMARA, D. S. (Ed.). *Adaptive educational technologies for literacy instruction*. Routledge, 2016. p. 263–267.

NAGATA, N. Intelligent computer feedback for second language instruction. *The Modern Language Journal*, v. 77, n. 3, p. 330–339, 1993. <https://doi.org/10.2307/329102>

NAGATA, N. An effective application of natural language processing in second language instruction. *CALICO Journal*, v. 13, n. 1, p. 47–67, 1995. Disponível em: <https://www.jstor.org/stable/24147743>. Acesso em: abr. 2025.

NAGATA, N. Computer vs. workbook instruction in second language acquisition. *CALICO Journal*, v. 14, n. 1, p. 53–75, 1996. Disponível em: <https://www.jstor.org/stable/24157094>. Acesso em: abr. 2025.

NATIONAL INSTITUTES OF HEALTH. *NIH style guide: Age*. U.S. Department of Health & Human Services, 2022. Disponível em: <https://www.nih.gov/nih-style-guide/age>. Acesso em: abr. 2025.

NUNES, A. et al. Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, v. 38, n. 2, p. 599–620, 2022. <https://doi.org/10.1111/jcal.12635>

NURALISAH, A. S.; KAREVIATI, E. The effectiveness of using Memrise application in teaching vocabulary. *Professional Journal of English Education*, v. 3, n. 4, p. 494–500, 2020. <http://dx.doi.org/10.22460/project.v3i4.p494-500>

PENG, P. et al. Simple view of reading in Chinese: A one-stage meta-analytic structural equation modeling. *Review of Educational Research*, v. 91, n. 1, p. 3–33, 2021. <https://doi.org/10.3102/0034654320964198>

POKRIVCAKOVA, S. Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, v. 7, n. 3, p. 135–153, 2019. <https://doi.org/10.2478/jolace-2019-0025>

POULSEN, R. *Tutoring bilingual students with an automated reading tutor that listens: Results of a two-month pilot study*. 2004. Dissertação (Mestrado) — DePaul University, 2004. [Não publicada].

PURWANTO, H. N.; FARIDI, A.; ROZI, F. The effect of Duolingo and SPADA to teach listening to students with different achievement levels. *English Education Journal*, v. 12, n. 1, p. 87–95, 2022. <https://doi.org/10.15294/EEJ.V12I1.54417>

PUZIO, K.; COLBY, G. T.; ALGEO-NICHOLS, D. Differentiated literacy instruction: Boondoggle or best practice? *Review of Educational Research*, v. 90, n. 4, p. 459–498, 2020. <https://doi.org/10.3102/0034654320933536>

RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical linear models: Applications and data analysis methods*. v. 1. Sage, 2002.

REIS, S. M. et al. The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*, v. 48, n. 2, p. 462–501, 2011. <https://doi.org/10.3102/0002831210382891>

ROHIM, R. F.; GUMELAR, R. E.; RIANDI, R. The effect of Memrise application towards students' vocabulary mastery. *Journal of English Language Learning*, v. 6, n. 2, p. 194–198, 2022. <https://doi.org/10.31949/jell.v6i2.3656>

SHARMA, K.; PAPAMITSIOU, Z.; GIANNAKOS, M. Building pipelines for educational data using AI and multimodal analytics: A “grey-box” approach. *British Journal of Educational Technology*, v. 50, n. 6, p. 3004–3031, 2019. <https://doi.org/10.1111/bjet.12854>

SONG, F. et al. Dissemination and publication of research findings: An updated review of related biases. *Health Technology Assessment*, v. 14, n. 8, p. 1–220, 2010. <https://doi.org/10.3310/hta14080>

STATA CORP. Stata statistical software: Release 16. StataCorp LLC, 2019.

TAEBENU, S. F.; KATEMBA, C. V. Vocabulary enhancement through Memrise and Google classroom. *Language Literacy: Journal of Linguistics, Literature and Language Teaching*, v. 5, n. 1, p. 228–241, 2021. <https://doi.org/10.30743/ll.v5i1.3813>

TAO, H. B.; DIAZ, V. R.; GUERRA, Y. M. Artificial intelligence and education: Challenges and disadvantages for the teacher. *Arctic Journal*, v. 72, n. 12, p. 30–50, 2019.

TSIRIGA, V.; VIRVOU, M. Evaluating the intelligent features of a web-based intelligent computer assisted language learning system. *International Journal on Artificial Intelligence Tools*, v. 13, n. 2, p. 411–425, 2004. <https://doi.org/10.1142/S0218213004001600>

VANTASSEL-BASKA, J.; WOOD, S. The integrated curriculum model (ICM). *Learning and Individual Differences*, v. 20, n. 4, p. 345–357, 2010. <https://doi.org/10.1016/j.lindif.2009.12.006>

VESSELINOV, R.; GREGO, J. *Duolingo effectiveness study*: Final report. 2012. Disponível em: [https://www.languagezen.com/pt/about/english/Duolingo\\_Efficacy\\_Study.pdf](https://www.languagezen.com/pt/about/english/Duolingo_Efficacy_Study.pdf)

VIECHTBAUER, W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, v. 36, n. 3, p. 1–48, 2010. <https://doi.org/10.18637/jss.v036.i03>

VIECHTBAUER, W.; CHEUNG, M. W.-L. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, v. 1, n. 2, p. 112–125, 2010. <https://doi.org/10.1002/jrsm.11>

VYGOTSKY, L. S. *Mind in society*: The development of higher psychological processes. Cambridge: Harvard University Press, 1978.

WANG, M. C.; WALBERG, H. J. (Eds.). *Adapting instruction to individual differences*. McCutchan, 1985.

WASHBURNE, C. W. (Ed.). *Adapting the schools to individual differences* (24th yearbook of the National Society for the Study of Education, Part 2). University of Chicago Press, 1925.

WAXMAN, H. C.; ALFORD, B. L.; BROWN, D. B. Individualized instruction. In: HATTIE, J.; ANDERMAN, E. M. (Eds.). *International guide to student achievement*. Routledge, 2013. p. 405–407.

WIJEKUMAR, K. K.; MEYER, B. J. F.; LEI, P. Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, v. 60, p. 987–1013, 2012. <https://doi.org/10.1007/s11423-012-9263-4>

WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J. *Data mining: Practical machine learning tools and techniques*. 4. ed. Morgan Kaufmann, 2016.

WOLTMAN, H. et al. An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, v. 8, n. 1, p. 52–69, 2012. <https://doi.org/10.20982/tqmp.08.1.p052>

XU, Y. et al. Artificial intelligence: a powerful paradigm for scientific research. *The Innovation*, v. 2, n. 4, p. 100179, 2021. <https://doi.org/10.1016/j.xinn.2021.100179>

ZAMZAMI, M. F. *The effect of using Duolingo application through students' vocabulary mastery*. 2019. [Tese de bacharelado não publicada]. Syarif Hidayatullah State Islamic University of Jakarta.

ZHENG, L. et al. The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments*, v. 31, n. 9, p. 5650–5664, 2023. <https://doi.org/10.1080/10494820.2021.2015693>.

## Apêndice

### Análise de Subgrupos para a Pergunta de Pesquisa 3

A Tabela A e a Figura A apresentam os resultados da análise de subgrupos para amostras entre grupos.

**Tabela A:** Resultados da Análise de Moderadores (Análise de Subgrupos) para Amostras Entre Grupos (N = 6.020)

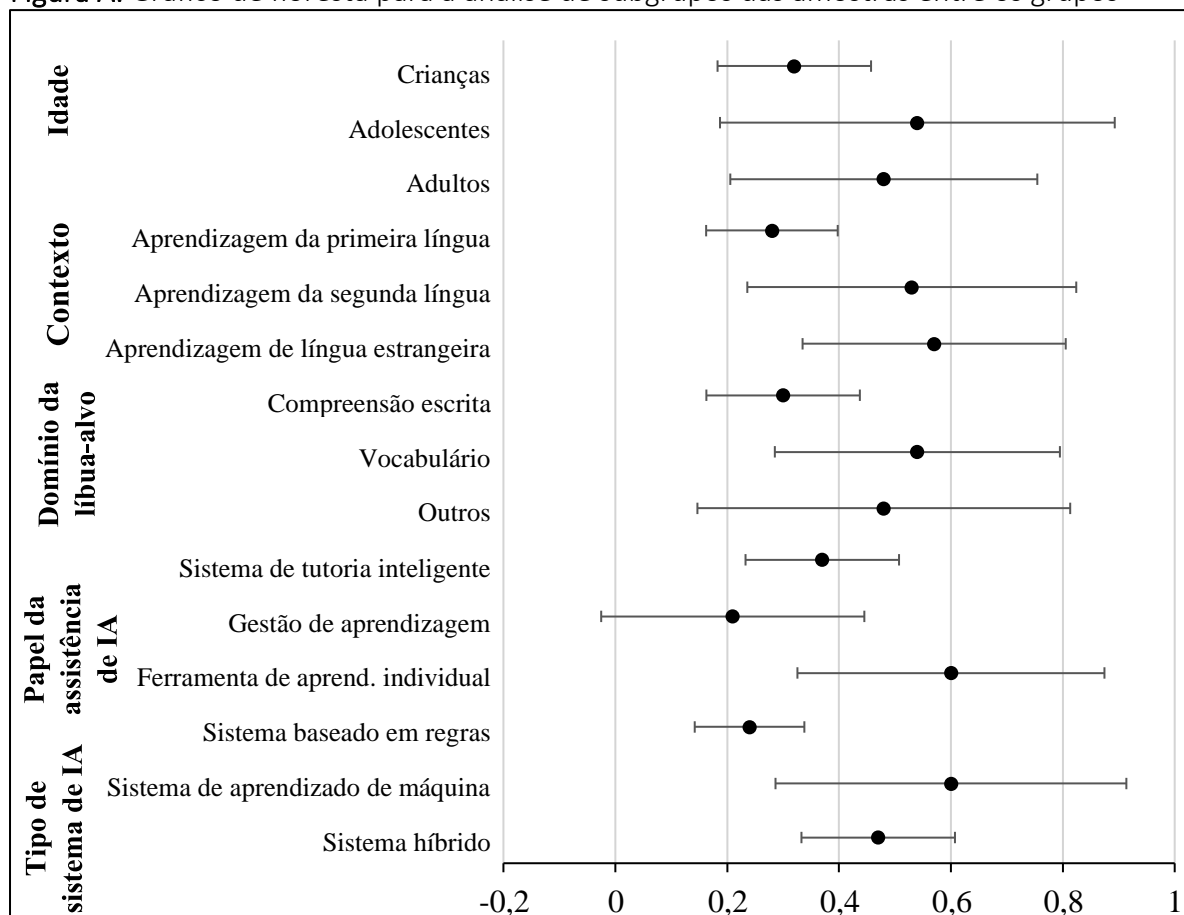
| Variável Independente                 | # do Tamanho do Efeito | # da Amostra | # do Projeto | Média do Tamanho do Efeito | do EP  |
|---------------------------------------|------------------------|--------------|--------------|----------------------------|--------|
| Idade                                 |                        |              |              |                            |        |
| Crianças                              | 60                     | 22           | 6            | 0,32***                    | (0,07) |
| Adolescentes                          | 5                      | 4            | 3            | 0,54**                     | (0,18) |
| Adultos                               | 11                     | 9            | 5            | 0,48***                    | (0,14) |
| Contexto de aprendizagem              |                        |              |              |                            |        |
| Aprendizagem da primeira língua       |                        | 21           | 6            | 0,28***                    | (0,06) |
| Aprendizagem da segunda língua        |                        | 2            | 2            | 0,53***                    | (0,15) |
| Aprendizagem de língua estrangeira    |                        | 12           | 6            | 0,57***                    | (0,12) |
| Domínio da língua-alvo                |                        |              |              |                            |        |
| Compreensão escrita                   | 57                     | 20           | 5            | 0,30***                    | (0,07) |
| Vocabulário                           | 12                     | 8            | 6            | 0,54***                    | (0,13) |
| Outros                                | 7                      | 7            | 4            | 0,48**                     | (0,17) |
| Papel da assistência de IA            |                        |              |              |                            |        |
| Sistema de tutoria inteligente        | 51                     | 20           | 8            | 0,37***                    | (0,07) |
| Gestão de aprendizagem                | 15                     | 7            | 1            | 0,21                       | (0,12) |
| Ferramenta de aprendizagem individual | 10                     | 8            | 4            | 0,60***                    | (0,14) |
| Tipo de sistema de IA                 |                        |              |              |                            |        |
| Sistema baseado em regras             | 43                     | 21           | 10           | 0,24***                    | (0,05) |
| Sistema de aprendizado de máquina     | 6                      | 6            | 2            | 0,60***                    | (0,16) |
| Sistema híbrido                       | 27                     | 8            | 1            | 0,47***                    | (0,07) |

Nota: Quatro variáveis de controle (a) duração da instrução, (b) ano de publicação, (c) desenho do estudo e (d) confiabilidade do teste, foram controladas; nenhuma foi estatisticamente significativa ( $p > 0,05$ ).

\*\*\*  $p < 0,001$ , \*\*  $p < 0,01$

Fonte: Elaboração própria (2024).

Figura A: Gráfico de floresta para a análise de subgrupos das amostras entre os grupos



Fonte: Elaboração própria (2024).

### Análise dos subgrupos

O mais importante, para uma descrição geral, foi que encontramos que a aprendizagem de línguas individualizada guiada por IA teve efeitos de pequeno porte ( $0,20 \leq d < 0,50$ ) ou de porte médio ( $0,50 \leq d < 0,80$ ) no desenvolvimento linguístico dos alunos em várias variáveis categóricas de cinco moderadores. Considerando que não houve tamanhos de efeito marginais (ou seja,  $d < 0,20$ ), foi concluído que a aprendizagem de línguas individualizada guiada por IA foi eficaz de forma geral. Em relação às diferenças estatísticas entre os valores dos moderadores, não conseguimos identificar nenhuma delas com base nos intervalos de confiança de 95% (ver Figura A).

Para diferentes faixas etárias, os resultados indicaram que a aprendizagem de línguas individualizada orientada por IA teve um efeito de porte médio ( $d = 0,54$ ,  $EP = 0,18$ ) para estudantes adolescentes, mas um efeito de pequeno porte tanto para crianças ( $d = 0,32$ ,  $EP = 0,07$ ) quanto para aprendizes adultos ( $d = 0,48$ ,  $EP = 0,14$ ). Quanto aos contextos de aprendizagem de línguas, encontramos efeitos de porte médio tanto para o aprendizado de segunda língua ( $d = 0,53$ ,  $EP = 0,15$ ) quanto para o aprendizado de língua estrangeira ( $d = 0,57$ ,

EP = 0,12). No entanto, houve apenas um efeito de pequeno porte para contextos de aprendizagem de língua materna ( $d = 0,28$ , EP = 0,06). Quanto aos domínios da língua-alvo, o tamanho médio do efeito foi médio para o aprimoramento do conhecimento de vocabulário dos alunos ( $d = 0,54$ , EP = 0,13), mas foi pequeno para a melhoria das habilidades de leitura dos alunos ( $d = 0,30$ , EP = 0,07) e para os outros domínios ( $d = 0,48$ , EP = 0,17).

Quanto aos papéis e tipos de sistemas de IA, identificamos que as ferramentas de aprendizagem individual tiveram um tamanho médio de efeito médio ( $d = 0,60$ , EP = 0,14) e que os dois papéis restantes, sistemas de tutoria inteligente ( $d = 0,37$ , EP = 0,07) e sistemas de gerenciamento de aprendizagem ( $d = 0,21$ , EP = 0,12), tiveram tamanhos de efeito médio de pequeno porte.

De acordo com o tipo de sistema de IA, os resultados revelaram que os sistemas de IA com sistemas de aprendizado de máquina tiveram um tamanho médio de efeito médio ( $d = 0,60$ , EP = 0,16), enquanto os outros tipos tiveram tamanhos de efeito médio pequeno, nomeadamente, os sistemas baseados em regras ( $d = 0,24$ , EP = 0,05) e os sistemas híbridos ( $d = 0,47$ , EP = 0,07).