

Padrões Semanais de Engajamento em AVA e Evasão em Disciplinas EaD: uma Análise Longitudinal com Big Data Educacional

Giovanna Mendes Garbácio¹, Filipe Dwan Pereira¹, Mario Antonio Pessoa Santos², Tyagi Lima², Debora Dizio Romano², Luiz Rodrigues³, Rafael Ferreira Mello⁴

¹ Universidade Federal de Roraima (UFRR),

²Cogna Educação S.A., SP, Brasil

³Universidade Tecnológica Federal do Paraná (UTFPR)

⁴Universidade Federal Rural de Pernambuco (UFRPE)

giovanna.garbacio@gmail.com

Resumo. A evasão na Educação a Distância é um desafio complexo, e seus sinais de descontinuidade podem emergir via padrões semanais de engajamento em Ambientes Virtuais de Aprendizagem (AVA). Este estudo realiza uma análise longitudinal com Big Data Educacional a partir de logs de 39.372 estudantes de graduação, totalizando 224.309 observações semanais. Por meio de Equações de Estimção Generalizadas (GEE), extraíram-se indicadores interpretáveis capazes de modelar o engajamento e o risco de evasão em uma disciplina ofertada em 31 cursos. Os resultados evidenciam padrões comportamentais associados à evasão; destaca-se, por exemplo, que calouros apresentam chance 2,11 vezes maior de evasão da disciplina quando comparados aos veteranos.

Abstract. Dropout in distance education is a complex challenge, and its signs of discontinuity may emerge through weekly engagement patterns in Virtual Learning Environments. This study conducts a longitudinal analysis with Educational Big Data based on logs from 39,372 undergraduate students, totaling 224,309 weekly observations. Using Generalized Estimating Equations (GEE), interpretable indicators were extracted to model engagement and the risk of dropout in a subject offered across 31 degree programs. The results reveal behavioral patterns associated with dropout, such as that first-year students were 2.11 times more likely to drop out of the subject when compared to other students.

1. Introdução

O aumento do uso de Ambientes Virtuais de Aprendizagem (AVA) gera grandes volumes de dados sobre as trajetórias dos estudantes [Ahmadi et al. 2023, Pereira et al. 2025]. No contexto de *Learning Analytics*, a análise desses dados permite observar sinais de afastamento acadêmico antes que se consolidem como evasão do curso [Palanci et al. 2024]. Em cursos EaD organizados em AVA, esse afastamento pode se manifestar pela redução ou interrupção da participação do estudante em um componente curricular. Assim, comportamentos como a diminuição de acessos, a não realização de avaliações e a interrupção

das interações com materiais oferecem indícios para analisar a evasão no nível da disciplina, considerando que esses padrões podem variar entre componentes curriculares [Barbare et al. 2025, Alzahrani et al. 2025, Prenkaj et al. 2020].

Para analisar esses padrões de interação a partir de *logs* de AVA, estudos sobre evasão frequentemente recorrem a métricas quantitativas, como frequência de *login*, número de acessos e tempo total gasto na plataforma (*time-on-task*) [Ahmadi et al. 2023, Raj and G 2022]. Porém, métricas centradas exclusivamente no volume de acesso ou no tempo total tendem a tratar comportamentos distintos como equivalentes, como consumir materiais didáticos de diferentes tipos, realizar exercícios e questionários ou revisitar conteúdos ao longo da disciplina [Arizmendi et al. 2023, Goh 2026]. Além disso, elas consideram que os padrões de interação não são estáveis nem homogêneos entre os estudantes ou ao longo do tempo, podendo assumir formas distintas conforme a experiência acadêmica do aluno e seu momento na trajetória da graduação [McKenna et al. 2024].

Essa limitação dialoga com a noção de *Time Spent Learning Well* (TSLW), métrica proposta para qualificar o tempo de uso a partir de sua contribuição esperada para a aprendizagem [Salvador 2024]. Essa perspectiva se aproxima da literatura sobre tempo de aprendizagem, que distingue o tempo dedicado a uma tarefa do tempo efetivamente associado ao alcance de critérios de aprendizagem [Gettinger 1984]. Em AVAs, estudos recentes reforçam essa necessidade ao mostrar que o tempo de estudo pode assumir interpretações distintas conforme o tipo de material acessado, a fase do percurso e a consistência das interações ao longo do curso [Hsu et al. 2023, Han 2026, Goh 2026].

Esse contexto mostra que ainda há espaço para abordagens que integrem, em dados de larga escala, indicadores semanais de engajamento, progressão acadêmica, composição do tempo por tipo de conteúdo e perfil acadêmico do estudante. Essa lacuna é particularmente relevante, uma vez que a transformação de registros institucionais brutos em atributos comportamentais interpretáveis e acionáveis permanece um desafio em aplicações de Big Data Educacional voltadas à evasão [Pereira et al. 2025].

Diante dessa lacuna, este trabalho propõe uma análise longitudinal e interpretável da evasão em disciplinas de cursos EaD de graduação a partir de *logs* de AVA disponibilizados pela Cogna Educação S.A., um dos maiores grupos educacionais privados brasileiros. Para isso, utilizam-se Equações de Estimação Generalizadas (GEE), abordagem adequada para modelar, em dados longitudinais com observações correlacionadas, o desfecho binário de evasão da disciplina [Huang 2022]. O objetivo é investigar como comportamentos semanais de engajamento no AVA e indicadores derivados do TSLW se associam à evasão da disciplina ao responder às seguintes questões de pesquisa (QP):

- **QP1:** Quais padrões semanais de interação caracterizam o engajamento dos estudantes em disciplinas de EaD organizadas em AVA?
- **QP2:** Como indicadores semanais de acesso, regularidade, revisitas, progressão acadêmica e participação em atividades se associam à evasão da disciplina?
- **QP3:** O *Time Spent Learning Well* (TSLW) acrescenta informação à análise da evasão em relação às métricas tradicionais de engajamento?
- **QP4:** Como os padrões semanais de engajamento variam entre estudantes calouros e veteranos em disciplinas EaD?

A principal contribuição deste estudo está em articular Big Data Educacional, granularidade semanal, análise no nível da disciplina e modelagem longitudinal interpretável

para investigar a evasão em disciplinas EaD de graduação. A partir de *logs* de AVA em larga escala, o trabalho operacionaliza indicadores semanais de ausência, progressão, participação, perfil acadêmico e composição do tempo por tipo de conteúdo, esta última representada pelo TSLW. Os resultados oferecem evidências interpretáveis sobre fatores associados à evasão, destacando a importância da dinâmica semanal do engajamento para estratégias de acompanhamento acadêmico.

2. Trabalhos Relacionados

A literatura já consolidou um conjunto de indicadores comportamentais baseados em *logs*, abrangendo uso da plataforma, acesso a materiais, realização de atividades, avaliações e comunicação [Ahmadi et al. 2023, Palanci et al. 2024]. Esses trabalhos consolidam os *logs* como fonte relevante para representar indicadores comportamentais de engajamento, mas geralmente utilizam esses sinais para a identificação, classificação ou predição de estados de engajamento e de risco único, sem considerar aspectos longitudinais entre padrões de acesso, regularidade, revisitas, progressão, participação e evasão da disciplina [Bellarhmouch et al. 2025, Raj and G 2022, Barbare et al. 2025].

A perspectiva de TSLW aprofunda essa limitação ao mostrar que a quantidade de tempo dedicada ao estudo nem sempre é suficiente para interpretar o engajamento. [Hsu et al. 2023] distinguem tempo total, tempo por tipo de material e tempo em diferentes fases de preparação para exames, enquanto [Han 2026] diferencia abordagens de estudo orientadas à preparação e à avaliação a partir de dados de *clickstream*. [Goh 2026], por sua vez, distingue persistência e consistência, reforçando que o valor analítico do tempo depende de sua distribuição ao longo do percurso. Esses trabalhos aprofundam o uso de *logs* ao tratar o tempo como medida composta, mas concentram-se principalmente em desempenho, estratégias de estudo ou consistência. Assim, permanece uma lacuna no contexto de evasão em disciplinas EaD que considere o TSLW, por exemplo, de uma decomposição semanal do tempo registrado no AVA por categorias de conteúdo.

A lacuna anterior é relevante porque a granularidade temporal é uma decisão central na construção de indicadores a partir de *logs*. [Alzahrani et al. 2025] propõem perfis semanais de engajamento, enquanto [Arizmendi et al. 2023] argumentam que atributos derivados de *logs* devem considerar a ação registrada, o conteúdo, a especificidade do comportamento e a janela temporal de observação. [Tamada et al. 2022] reforçam essa perspectiva ao demonstrar a contribuição de avaliar modelos em diferentes marcos de avanço do curso. A literatura também indica que esses padrões variam conforme a trajetória acadêmica do estudante. [McKenna et al. 2024] mostram diferenças no uso de recursos online entre anos da graduação, enquanto [Hsu et al. 2023] identificam ritmos distintos de estudo. Ou seja, diferentemente de análises centradas apenas no curso ou em agregações finais, é imperativo combinar granularidade semanal e unidade aluno-disciplina para comparar padrões de engajamento entre calouros e veteranos.

Por fim, destaca-se que estudos sobre evasão têm priorizado aumentar a compreensibilidade dos modelos. [Ramos et al. 2018] comparam classificadores para previsão de evasão em cursos EaD e observam desempenho competitivo da Regressão Logística, destacando uma alternativa menos opaca que redes neurais. [Silva et al. 2024] analisam métodos de IA Explicável aplicados à evasão e mostram que explicações baseadas em regras podem ser mais acessíveis a gestores e professores. [Kittur et al. 2022] propõem um

escore de engajamento em cursos online, combinando medidas temporais e relativas ao comportamento da turma para distinguir estudantes que concluem ou abandonam o curso.

Apesar desses avanços, parte das abordagens interpreta modelos preditivos após o treinamento ou sintetiza o engajamento em escores finais. Em contraste, esta pesquisa adota uma alternativa interpretável desde a modelagem (GEE) para lidar com observações correlacionadas, desfecho binário e inferência populacional em dados longitudinais [Pekár and Brabec 2018, Huang 2022]. Com isso, contribui-se para uma análise que preserva a estrutura semanal dos *logs* e estima associações entre indicadores de engajamento e de evasão, avançando o entendimento sobre padrões de interação em AVAs ao longo do tempo e de acordo com comportamentos particulares dos estudantes.

3. Método

Este estudo adota uma abordagem quantitativa, retrospectiva e longitudinal, baseada em Big Data Educacional proveniente de *logs* de interação de estudantes em um AVA de cursos EaD de graduação. A unidade de análise é a observação semanal de cada estudante em uma disciplina, o que permite investigar associações entre padrões de interação, progressão acadêmica, distribuição do tempo por tipo de conteúdo e evasão da disciplina.

Os dados utilizados foram extraídos de *logs* de acesso a uma plataforma AVA de cursos EaD de graduação de uma instituição privada de ensino superior brasileira. Esses registros compõem uma base educacional de larga escala por reunirem eventos digitais de navegação, com marcação temporal e associação a metadados acadêmicos da estrutura curricular. Os registros brutos correspondem a eventos de navegação realizados pelos estudantes e incluem identificador interno pseudonimizado do estudante, data e horário do acesso e URL visitada. A partir das URLs registradas e de metadados institucionais, cada evento foi associado ao respectivo curso, disciplina, unidade, aula e bloco de conteúdo acessado, enquanto eventos sem correspondência com tal estrutura foram descartados.

A análise foi conduzida na disciplina de Empreendedorismo e Educação, ofertada durante fevereiro e março de 2026 a estudantes de 31 cursos de graduação EaD. A disciplina ocorreu ao longo de nove semanas (0 a 8), organizada em unidades, aulas, materiais didáticos e atividades no AVA. Nesta oferta, calouros realizaram avaliação final presencial, enquanto veteranos realizaram avaliação final online, diferença considerada na interpretação dos nossos resultados. Esse recorte foi adotado para reduzir a heterogeneidade curricular e avaliativa, mantendo estudantes expostos à mesma estrutura de disciplina, conteúdos e atividades, ainda que vinculados a diferentes cursos de graduação.

Após os procedimentos de limpeza, filtragem e agregação semanal, a base analítica final foi composta por 224.309 observações semanais, correspondentes a 39.372 estudantes na disciplina analisada. A taxa de evasão observada foi de 13,21%, e 83,96% das unidades analisadas correspondiam a estudantes calouros. Nesta base, 47,43% das observações semanais não apresentaram interação registrada no AVA, enquanto 29,25% registraram realização de quiz na semana.

3.1. Pré-processamento e agregação semanal

Para estimar o tempo de permanência em cada recurso, calculou-se a diferença entre o timestamp de um evento e o timestamp do evento seguinte dentro da mesma sessão de navegação, replicando estudos com *logs* de AVA [Kittur et al. 2022, Goh 2026]. Embora

essa abordagem permita aproximar o tempo dedicado a cada recurso acessado, ela está sujeita a ruídos associados a durações muito curtas ou muito longas.

Para reduzir o efeito de durações atípicas nos indicadores de tempo, foram desconsideradas interações com duração inferior a 5 segundos ou superior a 1 hora e 45 minutos. O limite inferior foi adotado como critério para remover registros possivelmente associados a transições, recarregamentos de página ou acessos acidentais. Já o limite superior corresponde ao percentil 95 da distribuição dos tempos estimados, filtrando durações mais compatíveis com períodos de inatividade ou abandono da sessão aberta.

Após essa filtragem, os eventos válidos foram agregados por estudante, disciplina e semana relativa ao início da disciplina. A agregação semanal foi adotada por permitir representar a evolução do engajamento, preservando variações temporais que poderiam ser suavizadas em métricas acumuladas [Arizmendi et al. 2023, Alzahrani et al. 2025, Tamada et al. 2022]. Após esse procedimento, cada linha da base analítica passou a representar o comportamento de um estudante em uma disciplina durante uma semana relativa ao início da disciplina. Essa organização possibilitou analisar a evolução temporal dos indicadores de interação, progressão acadêmica e consumo de conteúdos.

Por fim, definiu-se como desfecho binário a evasão da disciplina, entendida como um indicador operacional de descontinuidade acadêmica no nível do componente curricular. Considerando que estudos com dados de AVA associam baixa participação, ausência de engajamento e não conclusão de atividades avaliativas ao risco de evasão [Barbare et al. 2025, Arizmendi et al. 2023], foram considerados em situação de evasão os estudantes cuja disciplina apresentou status de cancelamento ou que não possuíam registro de realização da avaliação final.

3.2. Variáveis e Indicadores Semanais

Foram construídos indicadores semanais para representar o comportamento dos estudantes no AVA, incluindo acesso, tempo de uso, progressão na disciplina, revisitação de conteúdos, participação em atividades e perfil acadêmico, em linha com métricas utilizadas em estudos sobre engajamento e evasão em ambientes digitais de aprendizagem [Alzahrani et al. 2025, Ahmadi et al. 2023, McKenna et al. 2024, Barbare et al. 2025]. As semanas foram definidas de forma relativa ao início da disciplina, permitindo comparar trajetórias de engajamento em uma mesma escala temporal. Essa estratégia é compatível com estudos que organizam *logs* educacionais em janelas ou fases de acompanhamento [Hsu et al. 2023, Han 2026, Barbare et al. 2025].

A decomposição do tempo por tipo de conteúdo foi utilizada para representar o TSLW, considerando que a literatura diferencia o tempo total de estudo de sua distribuição entre tipos de materiais e atividades de aprendizagem [Hsu et al. 2023, Han 2026]. Os tempos estimados foram convertidos em proporções do tempo total semanal de cada estudante na disciplina. Os blocos acessados foram agrupados conforme sua função na estrutura da disciplina. Blocos do tipo vídeo foram classificados separadamente, enquanto os demais recursos foram organizados em conceituais, práticos e reflexivos, conforme sua finalidade predominante no material didático. Na modelagem, foram utilizadas as proporções de tempo em conteúdos de vídeo, conceituais e práticos, permitindo comparar a composição do tempo de uso com métricas de engajamento. A Tabela 1 resume as dimensões analisadas e a descrição dos indicadores semanais utilizados no estudo.

Tabela 1. Dimensões e descrição dos indicadores semanais utilizados no estudo

Dimensão	Descrição
Desfecho	Indica evasão da disciplina (0 = conclusão; 1 = evasão).
Temporal	Semana relativa ao início da disciplina (valores: 0–8).
Acesso	Volume e frequência de uso do AVA na semana: quantidade de logs, sessões e dias de acesso (dias: 0–7; demais: ≥ 0).
Ausência	Indica ausência de interação no AVA durante a semana (0 ou 1).
Tempo de uso	Tempo total estimado de permanência nos recursos acessados na semana, em segundos (≥ 0).
TSLW	Proporção do tempo semanal distribuída entre categorias de conteúdo: vídeo, conceitual e prático (valores: 0–1).
Progressão	Avanço do estudante na estrutura da disciplina: cobertura acumulada (0–1), velocidade de progressão (≥ 0), variação entre unidades (≥ 0) e indicador binário de avanço para nova unidade (0 ou 1).
Revisitação	Retorno a conteúdos previamente acessados na semana (≥ 0).
Recursos complementares	Indica acesso a resumos e a conteúdos complementares da disciplina na semana (0 ou 1 cada).
Atividades	Participação semanal em quizzes e avaliações: indicador de realização de quiz (0 ou 1), número de tentativas e quantidade de avaliações distintas (≥ 0).
Perfil do estudante	Indica se o estudante é calouro no curso (0 ou 1).

3.3. Modelagem Estatística

A modelagem considerou que uma mesma unidade aluno-disciplina possui múltiplas observações semanais. Como essas observações tendem a apresentar correlação entre si, foram utilizados modelos de GEE, adequados para dados longitudinais e desfechos binários [Pekár and Brabec 2018, Huang 2022, Stellato et al. 2025]. Seja Y_{it} o indicador de evasão de um mesmo estudante na disciplina i na semana t , com $Y_{it} = 1$ para evasão e $Y_{it} = 0$ para conclusão, considerando $\pi_{it} = P(Y_{it} = 1)$, os modelos foram estimados com distribuição binomial e função de ligação logit:

$$\log \left(\frac{\pi_{it}}{1 - \pi_{it}} \right) = \beta_0 + \beta^\top \mathbf{X}_{it}.$$

As variáveis contínuas foram padronizadas por *escore-z*, permitindo interpretar seus coeficientes como a variação na chance de evasão associada ao aumento de um desvio padrão no indicador. As variáveis binárias foram mantidas na escala original 0/1. As observações foram agrupadas pelo identificador aluno-disciplina, e a ordenação temporal foi informada pela semana relativa da disciplina. Neste contexto, foram avaliadas as estruturas de correlação *exchangeable* e AR(1). A estrutura *exchangeable* assume correlação constante entre semanas de uma mesma unidade aluno-disciplina, enquanto a AR(1) assume maior correlação entre semanas temporalmente próximas. Essas estruturas foram comparadas em termos dos modelos detalhados na Tabela 2, incluindo um modelo base

(M_1), um modelo estendido (M_2) com indicadores do TSLW e um modelo complementar (M_3) com interações entre o perfil do estudante calouro e indicadores de engajamento.

Tabela 2. Modelos GEE estimados no estudo

Modelo	Especificação
M_1 Base	Semana relativa, volume de acesso, sessões, dias de acesso, tempo total, ausência, indicadores de progressão acadêmica (cobertura, velocidade, avanço entre unidades), revisitas, recursos complementares, participação em quizzes/avaliações e condição de calouro.
M_2 TSLW	M_1 + proporções padronizadas do tempo semanal em conteúdos de vídeo, conceituais e práticos.
M_3 Calouro	M_2 + interações entre calouro e tempo total, dias de acesso, ausência semanal, cobertura acumulada, realização de quiz e proporção de tempo em vídeo.

Os coeficientes estimados foram convertidos em razões de chances (*odds ratios*, OR), acompanhadas de intervalos de confiança de 95% (IC95%). Assim, valores de OR superiores a 1 indicam maior chance estimada de evasão associada ao indicador analisado, enquanto valores inferiores a 1 indicam menor chance estimada. As análises foram conduzidas no software R versão 4.6.0, com o pacote *geepack* versão 1.3.13.

4. Resultados

4.1. QP1: Padrões semanais de engajamento

Para responder à QP1, foram analisadas estatísticas descritivas das trajetórias semanais de acesso, ausência de interação, participação em quizzes, progressão acumulada, tempo médio entre observações ativas e composição do tempo por tipo de conteúdo. A Figura 1 mostra diferenças entre concluintes e estudantes em situação de evasão já nas primeiras semanas da disciplina. Os concluintes apresentaram, em média, mais dias de acesso, maior participação em quizzes e menor proporção de semanas sem interação registrada, especialmente nas semanas iniciais e intermediárias. Esse padrão sugere maior regularidade de acompanhamento da disciplina entre estudantes que concluíram o componente.

Em relação ao tempo médio entre observações com interação registrada, os concluintes apresentaram valores superiores nas semanas iniciais. A partir da semana 5, próxima ao período de avaliações, observa-se uma aproximação entre os grupos, seguida de queda no tempo médio dos concluintes e maior instabilidade no grupo de evasão nas semanas finais. Esse comportamento deve ser interpretado com cautela, pois o indicador considera apenas semanas com interação registrada e as semanas finais apresentam menor número de observações no grupo de evasão.

A cobertura acumulada indicou avanço progressivo em ambos os grupos, mas com maior cobertura média entre concluintes ao longo de quase todo o período observado. Essa diferença foi mais evidente nas semanas iniciais e intermediárias, sugerindo que a progressão acumulada na estrutura da disciplina distingue melhor os grupos do que medidas isoladas de volume de acesso. Por fim, a composição proporcional do tempo indica predominância de conteúdos em vídeo em ambos os grupos, seguida por conteúdos

conceituais, práticos e reflexivos. No grupo de evasão, observa-se uma redução pontual da proporção de tempo em vídeo próximo ao período avaliativo, acompanhada de maior participação relativa de conteúdos reflexivos e práticos. Esse padrão pode indicar mudanças na forma de navegação em momentos próximos às avaliações.

Padrões semanais de interação, progressão e composição do tempo

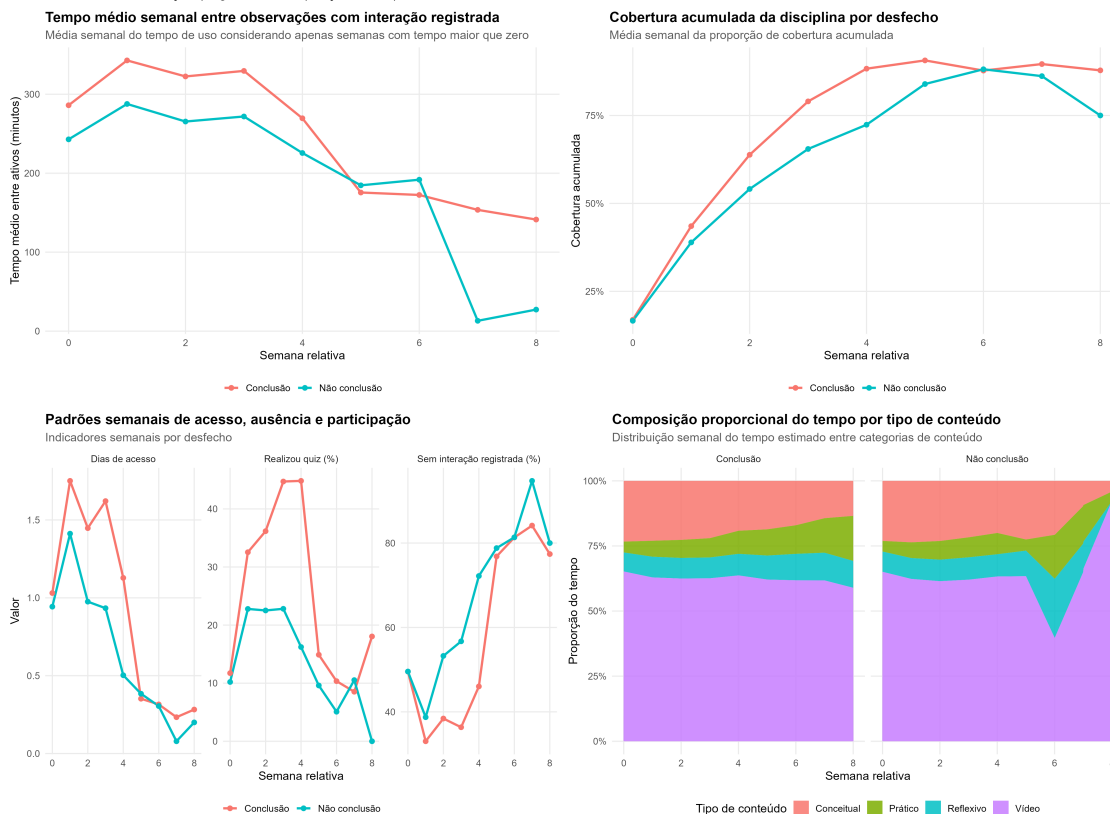


Figura 1. Padrões semanais de interação, progressão e composição do tempo por desfecho da disciplina. O tempo médio considera apenas observações semanais com interação registrada

4.2. QP2: Indicadores associados à evasão

Para responder à QP2, foi analisado o modelo base M_1 . Na comparação das estruturas de correlação, os modelos com AR(1) apresentaram maior adequação do que as versões *exchangeable*, logo, os resultados reportados utilizam a estrutura AR(1).

Os indicadores com associações mais consistentes com a evasão foram os relacionados à progressão e à continuidade na disciplina. A cobertura acumulada apresentou associação negativa com a evasão (OR = 0,9864; IC95%: 0,9835–0,9893), assim como o avanço para uma nova unidade (OR = 0,9934; IC95%: 0,9924–0,9944). Em contraste, semanas sem interação registrada apresentaram associação positiva com a evasão (OR = 1,0074; IC95%: 1,0052–1,0096). Por outro lado, velocidade de progressão e variação entre unidades acessadas apresentaram OR ligeiramente superiores a 1, indicando que deslocamentos mais rápidos pela estrutura da disciplina podem refletir navegação fragmentada, e não necessariamente progressão sustentada.

A participação em quizzes também se mostrou associada a uma menor chance estimada de evasão (realização de quiz na semana ($OR = 0,9965$; $IC95\%: 0,9950-0,9980$) e número de tentativas de quiz ($OR = 0,9981$; $IC95\%: 0,9969-0,9993$)). Entre as métricas de acesso, tempo total e número de sessões apresentaram associações negativas de pequena magnitude, enquanto quantidade de *logs* e dias de acesso não se mostraram indicadores consistentes. Em conjunto, os resultados sugerem que a evasão esteve mais associada à ausência semanal, à menor progressão acumulada e à menor participação em atividades do que ao volume bruto de eventos no AVA.

4.3. QP3: Contribuição do TSLW

Para responder à QP3, comparou-se o modelo base M_1 ao modelo estendido M_2 , mantendo a abordagem longitudinal por GEE adotada na modelagem estatística [Liang and Zeger 1986, Pekár and Brabec 2018] e acrescentando indicadores de TSLW baseados na decomposição do tempo por tipo de conteúdo. Como mostra a Figura 2, a inclusão da decomposição do tempo por tipo de conteúdo produziu estimativas semelhantes às do modelo base, pois os principais indicadores tradicionais mantiveram direção e magnitude próximas após a inclusão do TSLW. A principal diferença ocorreu no indicador de ausência semanal, cujo OR diminuiu de 1,0074 no modelo base para 1,0040 no modelo com TSLW, com intervalo de confiança passando a tangenciar 1, embora permanesse associado ao aumento da chance de evasão e sem alterar a interpretação geral dos resultados. De forma coerente com esse resultado, a diferença de ajuste entre os modelos também foi pequena. Na estrutura AR(1), o modelo com TSLW apresentou $QIC = 173.660,91$, enquanto o modelo base apresentou $QIC = 173.660,98$, sugerindo que a decomposição do tempo por tipo de conteúdo acrescentou informação limitada ao ajuste global do modelo.

Apesar da similaridade, a relevância teórica dos indicadores de TSLW justifica sua investigação [Hsu et al. 2023, Han 2026]. Entre os indicadores de TSLW, a proporção de tempo em conteúdos práticos apresentou associação negativa de pequena magnitude com a evasão ($OR = 0,9994$; $IC95\%: 0,9989-1,0000$). A proporção de tempo em vídeo teve intervalo de confiança muito próximo de 1, sugerindo evidência fraca de associação, enquanto a proporção de tempo em conteúdos conceituais não apresentou associação estatisticamente significativa. Assim, neste recorte, o TSLW acrescentou informação limitada ao ajuste global do modelo, mas contribuiu para explicitar que a composição do tempo por tipo de conteúdo não altera substancialmente as associações observadas com indicadores de ausência, progressão e participação.

4.4. QP4: Diferenças entre calouros e veteranos

Para responder à QP4, foram comparados os padrões semanais de engajamento entre calouros e veteranos, considerando também o desfecho da disciplina. Como mostra a Figura 3, os grupos apresentaram trajetórias semelhantes de cobertura acumulada, mas diferiram em indicadores de acesso, ausência e participação em quizzes. Entre concluintes, calouros apresentaram maior média de dias de acesso e maior participação em quizzes nas semanas iniciais e intermediárias, enquanto as diferenças de cobertura acumulada entre calouros e veteranos foram pequenas ao longo do período.

Entre estudantes em situação de evasão, a ausência de interação foi elevada em ambos os perfis, mas apresentou padrões distintos. Veteranos mantiveram níveis altos

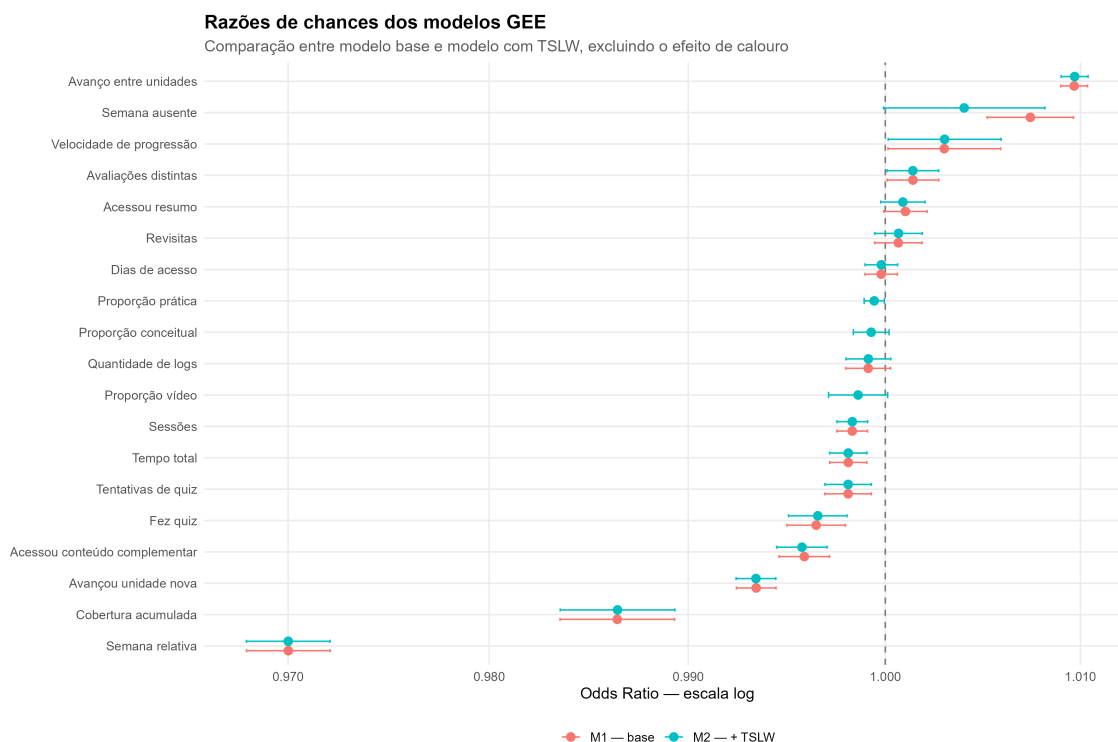


Figura 2. Comparação das razões de chances entre o modelo base e o modelo com TSLW. O efeito de calouro foi omitido da figura para melhorar a visualização dos demais indicadores

de ausência desde as primeiras semanas, enquanto calouros em evasão apresentaram aumento mais acentuado da ausência nas semanas intermediárias e finais. A participação em quizzes também foi menor entre estudantes em evasão, especialmente entre calouros, indicando menor continuidade no acompanhamento das atividades avaliativas.

Descritivamente, estudantes que evadiram apresentaram maior percentual de semanas sem interação e menor participação em quizzes em ambos os perfis. Entre calouros, a ausência semanal foi de 52,60% no grupo de evasão e 43,67% entre concluintes; a participação em quizzes foi de 18,61% e 32,13%, respectivamente. Entre veteranos, a ausência foi elevada nos dois grupos, com 63,99% entre não concluintes e 60,14% entre concluintes. Nos modelos GEE, a variável *calouro* apresentou associação consistente com a chance estimada de evasão. No M2, calouros apresentaram 2,11 vezes a chance estimada de evasão em comparação aos veteranos (OR = 2,11; IC95%: 1,93–2,30), controlados os demais indicadores semanais. O M3 com interações indicou heterogeneidade entre perfis acadêmicos. As interações entre *calouro* e ausência semanal (OR = 0,9885; IC95%: 0,9795–0,9976) e entre *calouro* e cobertura acumulada (OR = 0,9809; IC95%: 0,9756–0,9863) foram estatisticamente significativas, revelando que esses indicadores se associam de forma distinta à evasão entre calouros e veteranos.

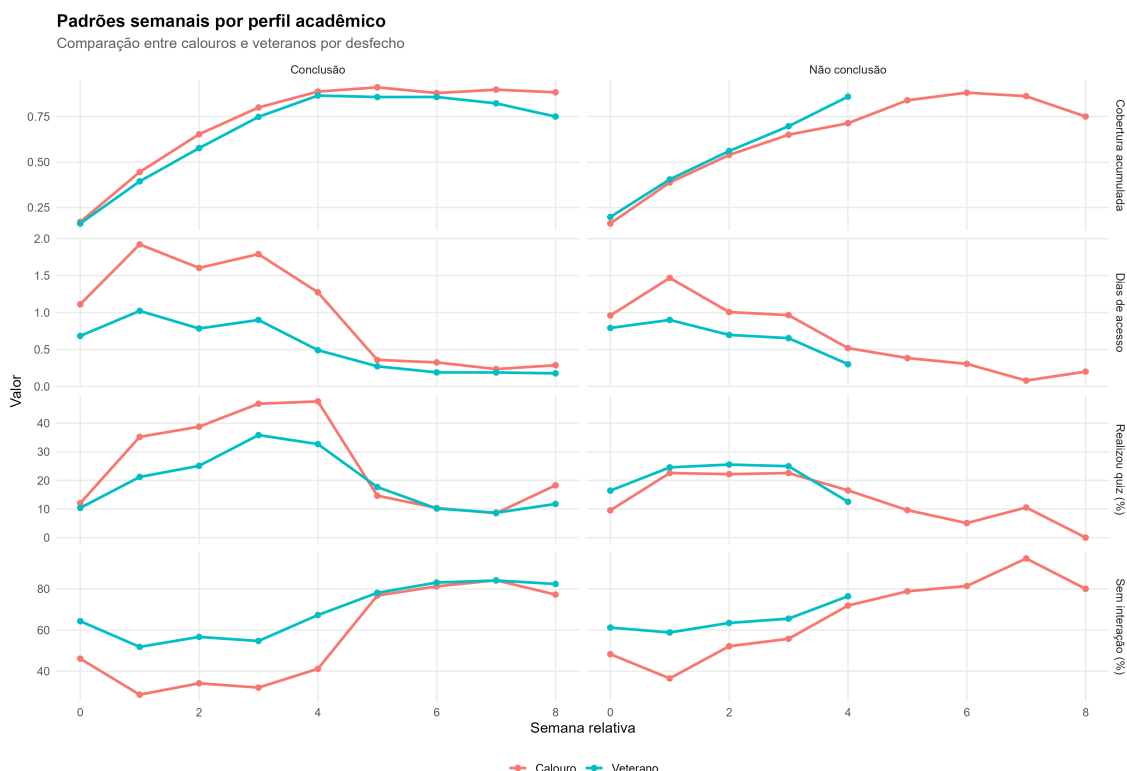


Figura 3. Padrões semanais de engajamento por perfil acadêmico, desfecho da disciplina e semana relativa

5. Discussão

Os resultados indicam que a evasão esteve mais associada a sinais de descontinuidade e baixa progressão do que ao volume bruto de acesso ao AVA. Cobertura acumulada, participação em quizzes e ausência semanal apresentaram associações mais consistentes com o desfecho, enquanto quantidade de *logs* e dias de acesso, isoladamente, não se destacaram da mesma forma. Esse achado converge com a crítica de que métricas centradas apenas em volume podem tratar como equivalentes interações qualitativamente distintas [Ahmadi et al. 2023, Goh 2026], reforçando a utilidade de indicadores semanais interpretáveis para o acompanhamento acadêmico.

Por outro lado, a decomposição do tempo por tipo de conteúdo, operacionalizada pelo TSLW, acrescentou informação limitada ao modelo base e não alterou substancialmente as associações observadas com ausência, progressão e participação. Esse resultado complementa estudos que destacam a relevância de distinguir o tempo total de estudo de sua distribuição entre tipos de materiais, atividades e fases do percurso de aprendizagem [Hsu et al. 2023, Han 2026], ao indicar que a contribuição dessa decomposição pode variar conforme o desfecho analisado. Ainda assim, o TSLW mantém relevância analítica ao qualificar o tempo de uso pela natureza do conteúdo acessado. Uma possível explicação para sua contribuição limitada é a estrutura relativamente padronizada da disciplina, que pode ter reduzido a variabilidade da composição do tempo entre os estudantes. Em disciplinas com maior autonomia de navegação ou diversidade de recursos, sua contribuição pode se manifestar de forma diferente.

O perfil acadêmico do estudante também se mostrou relevante. Calouros apresentaram 2,11 vezes a chance estimada de evasão em comparação aos veteranos, mesmo após controle pelos demais indicadores semanais. Esse resultado se aproxima de estudos que indicam variações no uso de recursos online conforme o momento do estudante na trajetória acadêmica [McKenna et al. 2024]. No entanto, os descritivos também mostraram ausência elevada entre veteranos, indicando que o afastamento acadêmico não se restringe aos ingressantes. Dessa forma, os resultados sugerem que estratégias de acompanhamento em EaD devem combinar ações específicas para calouros, possivelmente relacionadas à adaptação ao ambiente virtual e à rotina de estudos, com monitoramento contínuo de padrões de ausência e baixa progressão entre estudantes em diferentes etapas do curso.

Nesse contexto, os achados deste artigo têm contribuição especialmente relevante por seu diferencial metodológico. O uso de GEE permitiu analisar dados longitudinais em larga escala considerando a correlação entre observações semanais de um mesmo estudante na disciplina. Diferentemente de abordagens voltadas principalmente à classificação de estudantes em risco [Bellarmouch et al. 2025, Raj and G 2022, Barbare et al. 2025], a modelagem adotada favorece a interpretação de associações populacionais entre indicadores de engajamento e evasão [Liang and Zeger 1986, Pekár and Brabec 2018, Huang 2022]. Essa característica é relevante para aplicações em LA, nas quais os resultados precisam ser compreensíveis para apoiar decisões pedagógicas e institucionais.

Ainda, velocidade de progressão e variação entre unidades apresentaram associações em direção distinta da esperada. Esses indicadores podem representar formas fragmentadas de navegação, nas quais o estudante se desloca rapidamente entre conteúdos sem consolidar sua cobertura. Esse padrão difere da progressão gradual observada entre concluintes e sugere que avançar rapidamente na estrutura da disciplina não necessariamente representa engajamento sustentado. Isso dialoga com estudos que diferenciam intensidade, persistência e consistência do engajamento em AVA [Goh 2026] e com a recomendação de considerar a ação registrada, o conteúdo e a janela temporal na construção de atributos derivados de *logs* [Arizmendi et al. 2023]. Assim, os resultados reforçam que indicadores de progressão devem ser analisados em conjunto com medidas de cobertura, regularidade e participação.

Os achados devem ser interpretados considerando que a análise abrangeu uma única disciplina de uma instituição privada, com predominância de calouros e modalidades distintas de avaliação final por perfil. Além disso, o tempo de permanência foi estimado pela diferença entre eventos consecutivos, procedimento sujeito a ruídos mesmo após a aplicação dos limiares de duração. Adicionalmente, os modelos estimam associações entre indicadores semanais e evasão, não permitindo inferir relações causais.

Para trabalhos futuros, recomenda-se ampliar a análise para múltiplas disciplinas, períodos de oferta e instituições, permitindo avaliar se os padrões identificados se replicam em diferentes estruturas curriculares, modelos avaliativos e amostras. Também seria relevante comparar disciplinas com diferentes graus de autonomia de navegação e diversidade de recursos, nas quais a variabilidade do TSLW e demais comportamentos entre estudantes pode ser maior. Por fim, a inclusão do curso como variável de controle ou fator de estratificação pode revelar heterogeneidades relevantes para estratégias de acompanhamento acadêmico mais sensíveis ao contexto institucional.

Agradecimentos

Agradecemos à Cogna Educação S.A. pelo apoio institucional e pela disponibilização da infraestrutura e dos dados educacionais necessários à realização deste estudo, bem como aos especialistas da organização que contribuíram para sua viabilização.

Declaração sobre uso de Inteligência Artificial

Reconhecemos o uso de IA Generativa para auxiliar na elaboração e revisão deste artigo, que foi devidamente revisado, e assumimos total responsabilidade por seu conteúdo.

Cuidados Éticos

Este estudo utilizou dados secundários disponibilizados pela Cogna Educação S.A., provenientes de *logs* de interação em AVA e registros acadêmicos institucionais, previamente tratados para fins analíticos e de produção de evidências para a gestão educacional. O tratamento dos dados observou as políticas internas de segurança da informação da organização e a legislação aplicável à proteção de dados pessoais.

A base utilizada não continha identificadores diretos dos estudantes, sendo os registros analisados por identificadores pseudonimizados apenas para possibilitar a agregação longitudinal no nível aluno-disciplina-semana. Os resultados são apresentados de forma agregada, sem exposição de trajetórias individuais ou identificação de estudantes, docentes, polos ou unidades institucionais. Como a pesquisa não envolveu contato direto com participantes, intervenção pedagógica ou coleta primária de dados, foram adotadas medidas para minimizar riscos à privacidade e à confidencialidade, em conformidade com as diretrizes éticas aplicáveis.

Disponibilidade de artefatos

Os artefatos deste estudo serão disponibilizados pelos autores correspondentes.

Referências

- Ahmadi, G., Mohammadi, A., Asadzandi, S., Shah, M., and Mojtahedzadeh, R. (2023). What are the indicators of student engagement in learning management systems? a systematized review of the literature. *The International Review of Research in Open and Distributed Learning*, 24(1):117–136.
- Alzahrani, N., Meccawy, M., Samra, H., and El-Sabagh, H. A. (2025). Identifying weekly student engagement patterns in e-learning via k-means clustering and label-based validation. *Electronics*, 14(15).
- Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., and Gates, K. M. (2023). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods*, 55(6):3026–3054.
- Barbare, L., Jurenoks, A., Rauba, M., and Viskere, Z. (2025). Methodology for analysing lms data to predict student dropout risk in higher education. *ENVIRONMENT. TECHNOLOGY. RESOURCES. Proceedings of the International Scientific and Practical Conference*, 2:57–64.

- Bellarhmouch, Y., Majjate, H., Jeghal, A., Tairi, H., and Benjelloun, N. (2025). Detecting student engagement in an online learning environment using a machine learning algorithm. *Informatics*, 12(2).
- Gettinger, M. (1984). Achievement as a function of time spent in learning and time needed for learning. *American Educational Research Journal*, 21(3):617–628.
- Goh, T.-T. (2026). Learning management system log analytics: the role of persistence and consistency of engagement behaviour on academic success. *Journal of Computers in Education*, 13(1):283–306.
- Han, F. (2026). The relations between study approach, study time, and academic performance in flipped classrooms by questionnaire and clickstream data: To what extent are they consistent? *Journal of Computing in Higher Education*, 38(1):481–503.
- Hsu, C.-Y., Horikoshi, I., Li, H., Majumdar, R., and Ogata, H. (2023). Supporting “time awareness” in self-regulated learning: How do students allocate time during exam preparation? *Smart Learning Environments*, 10(1):21.
- Huang, F. L. (2022). Analyzing cross-sectionally clustered data using generalized estimating equations. *Journal of Educational and Behavioral Statistics*, 47(1):101–125.
- Kittur, J., Bekki, J., and Brunhaver, S. (2022). Development of a student engagement score for online undergraduate engineering courses using learning management system interaction data. *Computer Applications in Engineering Education*, 30(3):661–677.
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.
- McKenna, B. A., Wehr, J. B., and Kopittke, P. M. (2024). Quantifying online engagement at three levels of undergraduate study. *Cogent Education*, 11(1):2345939.
- Palanci, A., Yılmaz, R. M., and Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*, 29(17):22629–22650.
- Pekár, S. and Brabec, M. (2018). Equações de estimação generalizadas: uma abordagem pragmática e flexível para a modelagem glm marginal de dados correlacionados nas ciências comportamentais. *Ethology*, 124(2):86–93.
- Pereira, F. D., Mello, R., Queiroga, E. M., Padilha, I., Zambonin, G., Vital, R., Cassetari, B. d. O., Santos, M. A. P., and Lima, T. (2025). Big data na educação: Predição de evasão estudantil com aprendizado de máquina. In *Anais do XXXVI Simpósio Brasileiro de Informática na Educação*, pages 1805–1814. SBC.
- Prenkaj, B., Velardi, P., Stilo, G., Distant, D., and Faralli, S. (2020). A survey of machine learning approaches for student dropout prediction in online courses. *ACM Computing Surveys*, 53(3).
- Raj, N. S. and G, R. V. (2022). Early prediction of student engagement in virtual learning environments using machine learning techniques. *E-Learning and Digital Media*, 19(6):537–554.
- Ramos, J., Silva, J., Prado, L., Gomes, A., and Rodrigues, R. (2018). Um estudo comparativo de classificadores na previsão da evasão de alunos em ead. In *Anais do XXIX*

Simpósio Brasileiro de Informática na Educação, pages 1463–1472, Porto Alegre, RS, Brasil. SBC.

Salvador, R. (2024). Behind the metric: how we developed “time spent learning well”. Acesso em: 31 maio 2026.

Silva, F., Feitosa, R., Batista, L., and Santana, A. (2024). Análise comparativa de métodos de explicabilidade da inteligência artificial no cenário educacional: um estudo de caso sobre evasão. In *Anais do XXXV Simpósio Brasileiro de Informática na Educação*, pages 2968–2977, Porto Alegre, RS, Brasil. SBC.

Stellato, R. K., van den Bor, R. M., Schipper, M., Lindeboom, M. Y. A., and Eijkemans, M. J. C. (2025). Cohort data with dropout: a simulation study comparing five longitudinal analysis methods. *BMC Medical Research Methodology*, 25(1):103.

Tamada, M. M., Giusti, R., and Netto, J. F. d. M. (2022). Predicting students at risk of dropout in technical course using lms logs. *Electronics*, 11(3).