

# Mineração de Dados Educacionais a partir da Interação de Alunos com uma Plataforma Educacional

Lucas Barreto, Edwin Monteiro, Gabriel Leitão, Thays Bentes e Raimundo Barreto

Universidade Federal do Amazonas

{lucas.barreto, edwin, gabriel.leitao, rbarreto}@icomp.ufam.edu.br

thayskab@hotmail.com

**Abstract.** *The main aim of this paper is to propose a method of data analysis, based on educational data mining and some learning metrics, from the automatic collection of student interaction data with multiple choice questionnaires. These analyzes will serve as a basis for suggesting actions to be taken to improve the teaching/learning process. Experiments were conducted in several classes at a public high school although, in this paper, the focus will be on presenting the analysis of a single class. The experimental results show that the proposed method is promising, since it provides, for each student, the learning metrics, the priorities of the study topics and analysis based on the degree of doubts, the response time and topics with major and minor number of correct answers.*

**Resumo.** *O objetivo principal deste artigo é propor um método de análise de dados, baseado em mineração de dados educacionais e algumas métricas de aprendizagem, a partir da coleta automática de dados da interação de estudantes com questionários de múltipla escolha. Estas análises servirão de base para a sugestão de ações a serem tomadas visando a melhoria do processo de ensino/aprendizagem. Experimentos foram conduzidos em diversas turmas de uma escola pública de nível médio embora, neste artigo, o foco será apresentar as análises de uma única turma. Os resultados experimentais mostram que o método proposto é promissor, uma vez que fornece, para cada estudante, as métricas de aprendizagem, as prioridades dos tópicos de estudo e análises baseadas no grau de dúvidas, no tempo de resposta e nos tópicos com maior e menor número de acertos.*

## 1. Introdução

A sociedade atual é profundamente marcada pela tecnologia, o que implica que os artefatos tecnológicos não devem ser tratados como simples ferramentas, mas como parte integrante da vida das pessoas, moldando o modo como elas se relacionam consigo mesmas e com o mundo ao seu redor [Espírito Santo et al. 2012]. O contexto deste artigo está dentro da análise e descoberta de padrões nos dados de interação dos estudantes com questionários de múltiplas escolhas de uma plataforma educacional. A proposta é que as análises sirvam de base para a geração de ações a serem tomadas por professores e alunos. A principal contribuição deste artigo é a geração de diversas análises sobre os dados, partindo do pressuposto que análises baseadas em métricas de aprendizagem são eficazes. As métricas foram propostas em [Leitão et al. 2020] e estão relacionadas com avaliação formativa, principalmente com questionários de múltipla escolha.

A motivação é que, usualmente, alunos e professores não têm tempo de minerar informações a partir de relatórios de banco de dados. A ideia é entregar análises personalizadas, de tal forma que seja um passo crucial na implementação de melhorias no processo de ensino-aprendizagem. Neste caso, as análises são úteis para que os alunos possam dedicar o seu tempo em assuntos/tópicos que são mais urgentes e que estão muito próximos de atingir o objetivo da aprendizagem. Para os professores, as análises são úteis, pois fornecem melhores subsídios para a tomada de decisões levando em consideração o desempenho da aprendizagem. Isto é justificado porque é trabalhoso ao professor acompanhar uma aula em todas as suas nuances, principalmente considerando que cada aluno possui características particulares de aprendizado, seja pela diferença dos estilos de aprendizagem ou simplesmente pela diferença de velocidade de estudo. Não há dúvidas que analisar os diversos dados, e tomar a melhor decisão, sem a ajuda da tecnologia, não é uma tarefa fácil. Portanto, o problema a ser considerado neste artigo busca responder a seguinte pergunta: *É possível gerar análises de aprendizagem, tanto para alunos quanto para professores, a partir de dados coletados da interação dos alunos com questionários de múltipla escolha de uma plataforma educacional?*

O trabalho de [Paiva et al. 2019] descreve uma forma de avaliação baseada em analíticos de aprendizagem, mineração de dados e visualização de dados (DataViz). O trabalho de [Govindarajan et al. 2015] faz uso de dados acadêmicos com a finalidade de reduzir taxas de insucesso de aprendizado. O trabalho de [Martin e Ndoye 2016] usa técnicas de analítico de aprendizagem para avaliação formativa e somativa de estudantes em cursos online. Entretanto, estes não especificam nenhuma métrica de avaliação da aprendizagem diferentes da tradicional.

## 2. Referencial Teórico

### 2.1. Learning Analytics (Análise da Aprendizagem)

Este artigo emprega a mesma a definição de *Learning Analytics*, ou Análise da Aprendizagem, estabelecida na chamada de artigos da primeira *Conference on Learning Analytics and Knowledge* (LAK 2011) e adotada pela *Society for Learning Analytics Research* (SoLAR), ou seja, *a Análise da Aprendizagem é a medição, coleta, análise e geração de relatórios sobre os alunos e seus contextos, com o objetivo de compreender e otimizar a aprendizagem e os ambientes em que ela ocorre*. Esta definição pode ser usada para cobrir a maioria das pesquisas sobre educação. Entretanto, esta definição está associada com duas suposições principais: (i) que a análise da aprendizagem faz uso de dados que podem ser processados por computadores preexistentes; e (ii) que suas técnicas podem ser usadas para lidar com um “grande volume de dados”, os quais seriam impraticáveis serem tratados manualmente [Ferguson 2012].

### 2.2. Plataforma Educacional

O trabalho de [Leitao 2017] apresenta uma plataforma educacional apoiada por tecnologia, chamada SIGNAL, para auxiliar no processo de ensino-aprendizagem. Durante a interação do educando com a plataforma, o ambiente coleta dados que serão minerados e analisados por métricas de aprendizagem com a finalidade de prover *feedback* do desempenho estudantil. A Figura 1 apresenta a arquitetura da plataforma educacional, que é dividida em 5 módulos: Compositor, Servidor, Player, Inteligência e Analytics. O

módulo Compositor permite ao professor elaborar e cadastrar conteúdos educacionais, como aulas e avaliações que podem conter Objetos de Aprendizagem (OA), por exemplo apresentações, planilhas e vídeos. Tanto as aulas quanto os OA são armazenados em bancos de dados específicos. O Servidor é o módulo que gerencia salas de aula virtuais, estabelece a conexão entre os dispositivos dos usuários e consolida os dados capturados. É a partir deste módulo que professores e estudantes acessam as interfaces geridas por Player e Analytics. O módulo Player tem duas funções: exibir o conteúdo preparado pelo professor de modo que o aluno possa estudar e resolver o conteúdo avaliativo; e coletar os dados resultantes da interação do aluno. Após a coleta de dados, o módulo Analytics calcula as métricas de aprendizagem, analisa os dados minerados e extrai informações úteis, por exemplo, o número de vezes que um aluno alterou a resposta de uma questão e quanto tempo foi preciso. A partir das informações identificadas nos padrões de dados, o módulo apresenta um *feedback* no formato de gráficos baseado nas métricas adotadas e envia os dados gerados ao módulo Inteligência que faz recomendações de objetos de aprendizagem ou notifica o que deve ser tido como prioridade para o estudante.

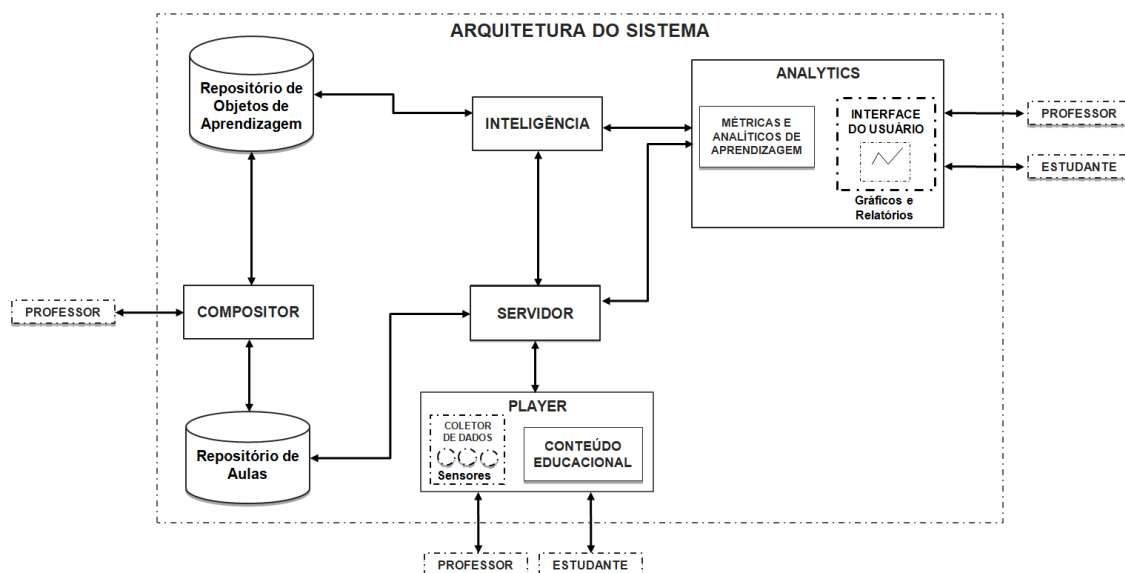


Figura 1. Arquitetura da Plataforma Educacional estruturada em módulos

### 2.3. Métricas de Aprendizagem

A plataforma educacional torna possível a captura automática dos dados de interação do aluno com os objetos de aprendizagem, particularmente, com os questionários de múltipla escolha. Os dados da interação possibilitam calcular as métricas de aprendizagem. Os detalhes das métricas estão fora do escopo deste artigo, mas podem ser consultadas em [Leitão et al. 2020]. Entretanto, as métricas consideradas neste artigo, de forma resumida, são: (i) **Média Tradicional** que é a pontuação baseada no total de questões assinaladas corretamente; (ii) **Média Ponderada** que considera que as alternativas possuem um peso, onde 1 indica totalmente errada e 4 indica totalmente correta. Esta métrica indica o quão próximo ou distante a alternativa marcada está da resposta correta; (iii) **Grau de Dúvida** que corresponde a quantidade de vezes que o aluno alterou a marcação de uma resposta no questionário; se aluno marcou somente uma vez, então o grau de dúvida é

zero; (iv) **Prioridade** que indica a prioridade de estudos de uma matéria baseando-se na Nota Tradicional e na Nota Ponderada; (v) **Tempo de resposta** que apresenta o tempo acumulado para responder cada questão; (vi) **Grau de Assertividade** que é uma métrica que relaciona o total de respostas corretas pela quantidade de vezes que o aluno trocou de alternativas em um questionário; e (vii) **Nível de Compreensão** que é uma métrica que quantifica a relação entre os índices de dificuldade do tópico, a assertividade e o peso relativo ao tempo de resposta.

### 3. Método Proposto

O método proposto consiste em uma abordagem para a análise da aprendizagem baseada em métricas específicas (veja Seção 2.3), a partir da coleta automática de dados da interação dos alunos, enquanto estes respondem questionários de múltipla escolha de disciplinas como Língua Portuguesa e Literatura, Matemática e outras destacadas na Seção 4. Esta coleta é possibilitada pela plataforma educacional apresentada na Seção 2.2. Estas análises são úteis para a geração de ações a serem tomadas, tanto para professores quanto para alunos. Há análises que são relativamente mais simples e diretas, que consistem em orientar o aluno a simplesmente estudar as matérias que teve o desempenho no questionário abaixo da média. Para ajudar ainda mais os alunos, esta ação pode ser baseada na métrica “Prioridade” (veja as Seções 2.3 e 4.1.2), que sugere uma ordem de tópicos ou matérias que o aluno poderia estudar, e que é baseada tanto na sua nota tradicional quanto na sua nota ponderada, ou seja, também considera o quão perto o aluno chegou de acertar a resposta. Outras análises, que são um pouco mais complexas, baseiam-se em outras métricas, por exemplo: identificar os alunos que tiveram muitas dúvidas em determinado tópico; ou os alunos que responderam em um tempo muito elevado ou muito baixo; ou os tópicos que tiveram um número muito baixo de acertos; e assim por diante. O método proposto neste artigo está baseado em [Leitao 2017] e consiste nas seguintes etapas:

1. compor um questionário para cada matéria, e a junção de vários questionários formam um simulado (Compositor);
2. aplicar o simulado para os alunos da turma (Player);
3. durante a aplicação do simulado, coletar os dados da interação dos alunos com a plataforma educacional e armazenar em um log (Servidor);
4. calcular as métricas de aprendizagem a partir do log (Analytics);
5. analisar as métricas visando a geração das análises da aprendizagem.

O objetivo deste artigo é a implementação e experimentação dos itens 4 e 5 (acima). Por limitação de espaço, os outros itens estão fora do escopo. A geração das métricas, tabulações, ordenações, seleções, categorizações, geração de gráficos e demais estatísticas, foram todas implementadas na linguagem de programação Python 3 utilizando as bibliotecas numpy, pandas e pymongo.

### 4. Avaliação Experimental

Uma das principais contribuições deste artigo é apresentar análises da aprendizagem tanto para alunos (veja a Seção 4.1) quanto para professores (veja a Seção 4.2). Para experimentar o método proposto, conduzimos um simulado do ENEM em uma escola pública amazonense de ensino médio com cinco turmas. No total são 420 alunos avaliados. O simulado teve 40 perguntas das matérias Biologia, Física, Geografia, Língua Espanhola,

Língua Inglesa, Língua Portuguesa/Literatura, Matemática e Química. Como a quantidade de dados é muito grande este artigo focará, sem perda de generalidade, na apresentação das análises dos dados de uma única turma.

#### 4.1. Análises da Aprendizagem para Alunos

##### 4.1.1. Análises Baseadas nas Métricas de Aprendizagem

As métricas de aprendizagem a serem tratadas especificamente para os alunos são: Nota Tradicional (NT), Nota Ponderada (NP), Grau de Dúvida, Tempo da Resposta, Grau de Assertividade e Nível de Compreensão. A Tabela 1 apresenta as métricas de um único aluno (ID “522”) em várias matérias. Vale a pena notar que cada questionário de cada matéria é composto por mais de uma questão, assim, os valores apresentados são médios.

Todas as Métricas de um Aluno						
Matéria	NT	NP	Dúvida	Tempo	Assertividade	Compreensão
Biologia	10.0	10.0	20%	60 (64)	83%	81%
Física	4.0	6.5	20%	138 (156)	33%	53%
Geografia	10.0	10.0	0%	19 (250)	100%	21%
L. Espanhola	6.0	8.5	20%	113 (180)	50%	61%
L. Inglesa	4.0	4.0	60%	167 (74)	25%	33%
L. Port/Lit	5.0	7.3	50%	178 (94)	33%	56%
Matemática	1.0	4.8	0%	140 (121)	10%	36%
Química	0.0	2.5	20%	414 (120)	0%	18%
Média	4.6	6.6	24%	142 (127)	42%	45%

**Tabela 1. Métricas de Aprendizagem obtidas para o aluno de ID “522”**

Somente analisando a Tabela 1 pode-se fazer algumas análises, inferências e observações, onde todas essas podem ser a base que permita a tomada de decisão por professores e alunos para um determinado tópico. Analisando a métrica *nota tradicional* do aluno “522”, percebe-se que ele é muito bom em Biologia e Geografia; mediano em Língua Espanhola e Língua Portuguesa; e que não teve bom aproveitamento em Física, Língua Inglesa, Matemática e Química. Na análise da *nota ponderada*, destaca-se a matéria Matemática na qual fica claro que o aluno tem conhecimentos do assunto porque “quase acertou a maioria”, mas por algum motivo acabou marcando a alternativa incorreta.

Com relação ao *grau de dúvidas*, as disciplinas de Língua Inglesa e Língua Portuguesa ocasionaram bastante dúvida conforme o percentual obtido, porém este fator pouco interfere na nota final. Entretanto, vale mencionar o caso contrário, isto é, a matéria Matemática cuja dúvida foi exatamente 0%, todavia o percentual não refletiu em uma nota acima da média. Isso quer dizer que, teoricamente, o aluno estava bem seguro. Este mesmo comportamento se repetiu na matéria Química.

No *tempo de resposta* há duas informações, o tempo médio de resposta do aluno e, entre parêntesis, o tempo médio esperado pelo professor. Percebe-se que há uma grande discrepância nas matérias Geografia, onde a estimativa do professor provavelmente foi exagerada, e Língua Inglesa e Química onde, por outro lado, a estimativa do professor foi provavelmente muito otimista. Essa métrica influencia o nível de compreensão.

A *assertividade* foi muito baixa exatamente em Matemática e Química porque foram as matérias em que o aluno teve as notas mais baixas. Em Biologia, onde o aluno tirou nota máxima, o seu nível de assertividade foi 83% devido aos 20% de grau de dúvida.

O *nível de compreensão* é uma métrica que leva em consideração diversos fatores, incluindo a dificuldade do tópico e a assertividade, além de penalizar o aluno caso ele tenha respondido muito rápido. O nível de compreensão mais baixo foi obtido em Química devido à baixa assertividade. Outro resultado significativo é a compreensão da matéria Geografia onde, apesar do aluno ter tirado a nota máxima, o tempo de resposta foi muito baixo e, por isso, o aluno foi bastante penalizado.

Há diversas ações que podem ser sugeridas para este aluno. Por exemplo, a partir da nota ponderada percebe-se que o aluno está se aproximando das respostas corretas em algumas disciplinas, como Física, Língua Espanhola, Língua Portuguesa e Matemática. Neste caso, a recomendação poderia ser uma maior dedicação à essas matérias porque o esforço para obter melhores resultados não seria muito elevado (veja a Seção 4.1.2). Outra possível recomendação indicaria ao aluno uma maior dedicação aos tópicos de Língua Inglesa, pois apresenta a maior taxa de dúvida (60%). Vale destacar que o tempo gasto para resolver as questões de Língua Inglesa, Língua Portuguesa e Química está muito elevado. Uma recomendação para a prática de mais exercícios pode melhorar a agilidade do estudante. Por fim, o tempo de resposta do aluno também pode gerar recomendações para o professor, caso o tempo das respostas esteja subestimado ou superestimado.

Análise das Prioridades de Estudos (Matérias)			
ID Aluno	Prioridade 1	Prioridade 2	Prioridade 3
077	Química (3.6)	<b>Matemática (3.5)</b>	Física (3.4)
095	Química (3.0)	L.Port/Lit (2.8)	L.Inglesa (2.4)
148	<b>Matemática (5.4)</b>	Física (4.2)	Química (2.7)
157	<b>Matemática (4.0)</b>	Química (3.9)	Física (3.0)
166	L.Port/Lit (4.1)	Física (3.5)	<b>Matemática (3.4)</b>
246	<b>Matemática (4.7)</b>	Física (4.5)	L.Port/Lit (4.4)
264	<b>Matemática (4.7)</b>	L.Port/Lit (4.2)	Química (3.9)
282	<b>Matemática (4.2)</b>	Geografia (3.3)	Química (3.2)
291	<b>Matemática (5.0)</b>	Física (4.4)	L.Espanhola (4.2)
317	Física (3.4)	L.Espanhola (3.3)	<b>Matemática (3.2)</b>
371	Química (3.9)	<b>Matemática (3.6)</b>	L.Port/Lit (3.6)
390	Física (4.4)	L.Port/Lit (3.9)	<b>Matemática (3.8)</b>
513	L.Port/Lit (3.6)	Física (3.6)	Química (3.6)
522	<b>Matemática (4.3)</b>	Física (3.9)	L.Port/Lit (3.6)
531	<b>Matemática (4.0)</b>	Química (3.6)	L.Port/Lit (3.2)
596	<b>Matemática (4.7)</b>	L.Port/Lit (3.5)	L.Espanhola (3.0)
602	Química (3.2)	Física (3.0)	L.Port/Lit (3.0)
649	<b>Matemática (4.3)</b>	L.Port/Lit (4.0)	Química (3.9)
685	L.Port/Lit (4.4)	<b>Matemática (4.0)</b>	Física (3.9)
765	Física (4.2)	L.Port/Lit (4.0)	<b>Matemática (3.9)</b>
809	Química (4.5)	<b>Matemática (3.9)</b>	Física (3.6)
872	L.Port/Lit (2.9)	L.Espanhola (2.8)	Geografia (2.4)
907	Química (4.5)	<b>Matemática (4.4)</b>	Física (3.4)
964	Física (4.4)	<b>Matemática (4.0)</b>	L.Port/Lit (3.9)

Tabela 2. Análise das Prioridades de Estudos (Matérias) a partir das métricas

#### 4.1.2. Análises Baseadas na Prioridade de Estudos

A Tabela 2 descreve as três matérias prioritárias para cada aluno da turma. Tais informações são muito importantes para fornecer um melhor direcionamento aos estudantes. Neste caso particular, as prioridades dependem do desempenho de cada aluno no respectivo simulado. Por exemplo, para o aluno “077” as prioridades são: Química, Matemática e Física, enquanto que para o aluno “872” as prioridades são: Língua Portuguesa/Literatura, Língua Espanhola e Geografia. Uma análise importante que pode ser feita é: há matérias que são recorrentes dentre as prioridades dos alunos. Por exemplo, a matéria Matemática está em 20/24 (83%) dos casos, sendo que em 10/24 (42%) está como primeira prioridade. Já as matérias Física e Língua Portuguesa/Literatura estão em 16/24 (67%) dos casos. Independentemente da análise estar focando somente nos alunos, abre-se a possibilidade de estender a análise e colocar estas matérias como prioridade das próprias instituições de ensino.

Matéria	Sem Dúvidas				Dúvidas Aceitáveis				Dúvidas Elevadas			
	Q	S	A	E	Q	S	A	E	Q	S	A	E
Biologia	174	<b>21</b>	3	0	173	15	<b>9</b>	0	171	16	5	<b>3</b>
Física	75	<b>24</b>	0	0	74	18	<b>5</b>	1	74	18	5	<b>1</b>
Geografia	59	<b>23</b>	1	0	58	22	<b>2</b>	0	57	21	1	<b>2</b>
L.Espanhola	153	<b>24</b>	0	0	156	19	<b>5</b>	0	153	24	0	<b>0</b>
L.Inglesa	216	<b>21</b>	3	0	221	16	<b>7</b>	1	222	18	4	<b>2</b>
L.Port/Lit	93	<b>22</b>	1	1	87	18	<b>8</b>	0	85	18	4	<b>2</b>
Matemática	176	<b>22</b>	2	0	177	20	<b>4</b>	0	180	20	2	<b>2</b>
Química	200	<b>21</b>	2	1	199	18	<b>6</b>	0	200	21	2	<b>1</b>

Tabela 3. Análise do Grau de Dúvidas

#### 4.2. Análises da Aprendizagem para as Turmas

##### 4.2.1. Análises Baseadas no Grau de Dúvidas

A Tabela 3 apresenta a análise baseada no grau de dúvidas. Para cada matéria, escolhemos uma questão (Q) que exemplificasse as três análises: (i) questão em que os alunos não tiveram nenhuma dúvida (S); (ii) questão em que os alunos tiveram dúvidas, contudo ainda está em um nível aceitável, ou seja, entre uma e quatro (A); e questão em que os alunos tiveram dúvidas em um nível elevado, ou seja, maiores que quatro (E). Analisando a matéria *Biologia* percebe-se que a questão 174 foi a que teve o maior número de alunos que não tiveram dúvidas, no caso vinte e um alunos. A questão 173 foi a que teve o maior número de alunos com dúvidas aceitáveis, no caso nove alunos. E a questão 171 foi a que teve o maior número de alunos com dúvidas em nível elevado, no caso três alunos. Este mesmo raciocínio é válido para todas as outras matérias.

Os dados dessa turma, em todas as matérias, não são alarmantes, uma vez que o maior número de dúvidas em nível elevado foi de três alunos exatamente na disciplina *Biologia*. Entretanto, uma atenção especial deve ser dada aos alunos que tiveram muitas dúvidas. Há diversas situações que precisam ser avaliadas, por exemplo se a metodologia de ensino está apropriada, se os objetos de aprendizagem têm conseguido ajudar no entendimento do conteúdo, se precisa de mais avaliações formativas, e assim por diante.

Matéria	Muito Rápido				Aceitável				Além do Tempo			
	Q	M	D	A	Q	M	D	A	Q	M	D	A
Biologia	174	<b>1</b>	8	15	173	0	<b>16</b>	8	170	0	3	<b>21</b>
Física	74	<b>4</b>	13	7	77	0	<b>20</b>	4	75	1	9	<b>14</b>
Geografia	58	<b>18</b>	5	1	57	13	<b>11</b>	0	58	18	5	<b>1</b>
L.Espanhola	154	<b>10</b>	13	1	154	10	<b>13</b>	11	155	0	7	<b>17</b>
L.Inglesa	216	<b>1</b>	2	21	217	1	<b>12</b>	11	216	1	2	<b>21</b>
L.Port/Lit	85	<b>2</b>	11	11	93	0	<b>21</b>	3	82	0	0	<b>24</b>
Matemática	182	<b>3</b>	15	6	185	2	<b>19</b>	3	178	0	2	<b>22</b>
Química	199	<b>4</b>	12	8	199	4	<b>12</b>	8	201	0	4	<b>20</b>

Tabela 4. Análise do Tempo de Resposta por questão

#### 4.2.2. Análises Baseadas no Tempo de Resposta

A Tabela 4 apresenta a análise do tempo de resposta de uma turma. Para cada matéria, escolhemos uma questão (Q) que exemplificasse as três análises: (i) questão que os alunos responderam em menos de um quarto do tempo estimado pelo professor (M); (ii) questão que os alunos responderam em um tempo maior do que um quarto, mas dentro do limite de tempo máximo estabelecido pelo professor (D); e (iii) questão que os alunos responderam em um tempo maior do que o tempo máximo estimado pelo professor (A).

Analisando, por exemplo, as questões 58 (Geografia) e 154 (Língua Espanhola), estas foram as que tiveram a maior quantidade de alunos que responderam em um tempo menor do que um quarto do tempo estimado pelo professor. Os tópicos dessas questões são “*Geografia do Amazonas - Clima*” e “*Língua formal*”. Vale a pena notar que esta é uma situação indesejada, uma vez que os alunos estão respondendo em um tempo bastante inferior ao que o professor estimou. Como a quantidade de alunos é relativamente alta, há a possibilidade do professor ter exagerado no tempo máximo estimado. Nesse caso, uma recomendação deve ser enviada para o respectivo professor. As questões 93 (Língua Portuguesa/Literatura) e 77 (Física), foram as que tiveram a maior quantidade de alunos que responderam dentro do tempo esperado. Os tópicos são “*Literatura*” e “*Ondulatória*”, respectivamente. Por último, as questões 82 (Língua Portuguesa/Literatura) e 178 (Matemática) foram as que tiveram a maior quantidade de alunos que responderam além do tempo esperado. Os tópicos são “*Leitura*” e “*Análise Combinatória*”, respectivamente. Neste caso, como foram muitos alunos que extrapolaram o tempo, há a possibilidade do professor ter sido muito otimista no estabelecimento do tempo máximo estimado. Portanto, uma recomendação também deve ser enviada para o respectivo professor.

Um aspecto que vale a pena investigar são os alunos que, embora a grande maioria teve um tempo de resposta baixo, eles, por outro lado, tiveram um tempo de resposta elevado. Analisando os dados dessa turma em particular, descobrimos que o aluno “317” teve esse comportamento na matéria Geografia, questão 58 e tópico “*Geografia do Amazonas - Clima*”. Outro aluno que teve esse mesmo comportamento foi o “371” na matéria Língua Espanhola, questão 154 e tópico “*Língua Formal*”. Outro aspecto que merece a nossa atenção é o caso oposto, ou seja, analisar os casos em que a maioria dos alunos precisaram de um tempo elevado enquanto que alguns poucos precisaram de um tempo muito baixo. Neste caso, deve-se que avaliar a possibilidade de ter havido um “chute”, ou seja,



uma tentativa de acertar a resposta por palpite ao invés de certeza. A Tabela 4 apresenta dois casos em que este comportamento ocorreu. O primeiro ocorreu na matéria Biologia, questão 174 e tópico “*Anatomia do Sistema Nervoso*”; e o segundo foi na matéria Língua Inglesa, questão 216 e tópico “*Leitura*”. Como temos todos os dados, descobrimos que os alunos envolvidos nesta situação são os alunos “317” e “166”.

#### 4.2.3. Análises Baseadas nos Tópicos com o Maior e Menor Números de Acertos

Um simulado é um conjunto de questionários, isto é, um questionário para cada matéria. A ideia desta análise é avaliar os tópicos que tiveram os maiores e menores níveis de acertos. Os resultados das análises estão na Tabela 5. Com relação aos tópicos que tiveram uma maior quantidade de acertos, percebe-se que os tópicos *Divisão anatômica do sistema nervoso*, da matéria Biologia, e o *Literatura - Formas Literárias* da matéria Língua Portuguesa/Literatura, foram os tópicos que tiveram um nível bastante alto de acertos. Por outro lado, quando se considera os tópicos que tiveram um nível de acerto mais baixo, percebe-se que os tópicos *Elementos de coesão* e o *Literatura - Definições e Conceitos*, ambos da matéria Língua Portuguesa/Literatura, foram os tópicos que tiveram o menor número de acertos. Por causa disso, abre-se a possibilidade de sugerir ações diferenciadas para estes tópicos específicos.

Tópicos com Alto Nível de Acerto				Tópicos com Baixo Nível de Acerto			
#	Tópico	Matéria	%	#	Tópico	Matéria	%
90	Sist. nervoso	Biologia	95.8%	48	Coesão	L.Port/Lit	12.5%
46	Literárias	L.Port/Lit	91.7%	49	Literárias	L.Port/Lit	12.5%
80	Linguagens	L.Port/Lit	81.9%	108	Pilhas	Química	16.7%
11	Clima do Am	Geografia	79.2%	35	Leitura	L.Port/Lit	20.8%
8	Relevo do Am	Geografia	75.0%	115	Lit Estrang.	L.Inglesa	25.0%
13	Hidrog. Am	Geografia	75.0%	39	Lit Bras.	L.Port/Lit	25.0%
91	Sist. nervoso	Biologia	75.0%	9	Minerais Am	Geografia	33.3%
10	Geologia Am	Geografia	70.8%	24	NOX	Química	33.3%
34	Leitura	L.Port/Lit	69.8%	89	Hormonal	Biologia	33.3%
107	Pilhas	Química	66.7%	87	Combinatória	Matemát.	40.9%

Tabela 5. Análise dos Tópicos com Alto e Baixo Nível de Acerto

## 5. Conclusão

Este artigo apresentou um método de análise da aprendizagem baseado em dados de interação de alunos com questionários de múltipla escolha. As análises foram baseadas em um conjunto específico de métricas de aprendizagem [Leitão et al. 2020] e foram direcionadas para professores e alunos. Para os alunos de uma turma, as análises forneceram as métricas de aprendizagem consideradas neste trabalho e as matérias prioritárias para estudo. Para os professores, foram geradas análises baseadas no grau de dúvidas, no tempo de resposta e nos tópicos com maior e menor número de acertos. Como resultado de cada análise foram indicadas possíveis ações a serem tomadas.

Acredita-se ser possível uma melhor avaliação da aprendizagem dos alunos, possibilitando que os professores proporcionem novos direcionamentos e caminhos, sejam

eles para estabelecer novos métodos, ou para reformular os questionamentos trabalhados. Quanto aos alunos, ao não alcançarem em sua totalidade os conteúdos, por não terem obtido compreensão significativa nas matérias, as análises de aprendizagem, e ações sugeridas, poderão ajudar no aumento do engajamento, mas não de forma *ad hoc* e sim direcionada. Além disso, as análises da aprendizagem tecnológicas permitiram identificar, de maneira mais apurada e individualmente, as peculiaridades dos alunos da turma analisada, permitindo aos alunos conhecer suas dificuldades e onde devem dedicar-se mais, para aprimorar seus conhecimentos e otimizar seu aprendizado.

Como trabalhos futuros, pretendemos investigar a geração de agrupamentos. A ideia é agrupar alunos com perfis semelhantes baseando-se nas métricas e atribuir um valor de desempenho geral para o grupo. A proposta é identificar os alunos com desempenho particular destoante do grupo em que foram incluídos. Por exemplo, se um aluno que tem boas notas foi incluído em um grupo que, na maioria, os alunos têm notas ruins, isso indica que esse aluno, apesar de ter notas consideradas boas, em pouco tempo pode vir a estar com notas ruins também. Por outro lado, se um aluno com notas ruins for colocado em um grupo com notas consideradas boas, tudo indica que esse aluno em pouco tempo pode ter boas notas também.

### **Agradecimentos**

Esta pesquisa, conforme previsto no Art. 48 do decreto nº 6.008/2006, foi parcialmente financiada pela Samsung Eletrônica da Amazônia Ltda, nos termos da Lei Federal nº 8.387/1991, através de convênio nº 003/2019, firmado com o ICOMP/UFAM. Os autores também agradecem o apoio financeiro da Fundação de Amparo à Pesquisa do Estado do Amazonas através do Projeto 122/2018 (UNIVERSAL).

### **Referências**

- Espírito Santo, J., Castelano, L., e Almeida, J. (2012). Uso de tecnologias na prática docente: Um estudo de caso no contexto de uma escola pública do interior do rio de janeiro. Em *II Congresso Intern. TIC e Educação*, páginas 1023–1031. Univ. Lisboa.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6):304–317.
- Govindarajan, K., Kumar, V. S., Boulanger, D., et al. (2015). Learning analytics solution for reducing learners' course failure rate. Em *IEEE seventh international conference on technology for education (T4E)*, páginas 83–90. IEEE.
- Leitao, G. (2017). Uma plataforma de suporte ao docente no contexto da educação digital. Master's thesis, Programa de Pós-graduação em Informática da Universidade Federal do Amazonas.
- Leitão, G., Colonna, J., Monteiro, E., Oliveira, E., e Barreto, R. (2020). New metrics for learning evaluation in digital education platforms. <https://arxiv.org/abs/2006.14711>.
- Martin, F. e Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice*, 13(3):7.
- Paiva, R., Bittencourt, I., e Lemos, W. (2019). Helping teachers visualize students' performance. Em *Simp. Bras. de Informática na Educação*, volume 30, página 1731.