

Indicadores de qualidade do ensino fundamental: o uso das tecnologias de mineração de dados e de visões multidimensionais para apoio à análise e definição de políticas públicas*

Anderson Amendoeira Namen
Sonia Xavier de Almeida Borges
Maria da Glória Schwab Sadala

Resumo

O Ministério da Educação (MEC) possui bases de dados contendo milhões de registros de alunos, professores, diretores e escolas do ensino fundamental. Muitas vezes essas bases não são exploradas em toda a sua potencialidade, devido ao grande volume de dados envolvido e às dificuldades inerentes ao tratamento dessas informações. Este artigo apresenta o potencial de algumas tecnologias de descoberta de conhecimento aplicadas às bases de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). São expostos os conceitos de visão multidimensional e de mineração de dados e apresentam-se as possibilidades de uso dessas ferramentas, que propiciam agilidade e facilidade na obtenção de indicadores de qualidade. São apresentados e discutidos alguns indicadores relacionados ao ensino da Língua Portuguesa, obtidos por intermédio das tecnologias mencionadas.

Palavras-chave: ensino fundamental; ensino de Língua Portuguesa; sistemas de apoio à decisão.

* O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), por intermédio do Programa Observatório da Educação.

Abstract

Elementary school quality indicators: the use of data mining and multidimensional data visualization for decision support and formulation of public policies

The Brazilian Ministry of Education stores data bases which contain millions of records of students, teachers, elementary schools and their principals. Despite of its great importance, these data bases are not explored in its total potentiality. The great volume of data creates some difficulties related to the manipulation of the information. This paper shows some knowledge discovery technologies that can be used with education data bases. Multidimensional data visualization and data mining concepts are presented. The possibilities and potentialities of both technologies provide agility and facilitate getting quality indicators. Some indicators related to Portuguese Language learning, which were obtained through the use of these technologies, are presented.

Keywords: elementary school; Portuguese Language teaching; decision support systems.

Introdução

Em 2007, foi lançado o Plano de Desenvolvimento da Educação (PDE), definindo uma série de metas a serem alcançadas relacionadas ao fortalecimento da educação básica. O Censo Escolar da Educação Básica, realizado anualmente, é um dos mecanismos que propiciam o acompanhamento das metas definidas pelo PDE (Brasília. Inep, 2009). Além do censo, periodicamente são realizadas avaliações em nível nacional (Saeb e Prova Brasil) visando à coleta de indicadores de proficiência em Língua Portuguesa e Matemática dos alunos do ensino básico.

As informações sobre escolas, diretores, professores e alunos, coletadas por intermédio do censo escolar e das avaliações periódicas, são armazenadas em uma ampla base de dados disponibilizada a pesquisadores pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).

Avaliando as bases de dados do Inep, pode-se concluir que estas, apesar de serem importante fonte de informação, muitas vezes não são exploradas em todo o seu potencial, devido às dificuldades inerentes ao grande volume de dados envolvido, que inviabiliza a capacidade humana de sua interpretação. Nesse contexto, o presente trabalho apresenta duas tecnologias que podem apoiar a descoberta de conhecimento em bases de dados relacionados à educação: visões multidimensionais de dados e mineração de dados.

Inicialmente são apresentados os conceitos referentes a cada tecnologia e exemplos de aplicações em diferentes áreas de conhecimento. São mencionados, então, alguns aspectos particulares que foram considerados na utilização das bases de dados do Inep, de modo a eliminar inconsistências e possibilitar a construção de indicadores com integridade e confiabilidade. Em seguida, o artigo ilustra a facilidade e agilidade propiciada a pesquisadores na elaboração e criação de relatórios de apoio à decisão. Alguns indicadores ligados ao ensino da Língua Portuguesa são apresentados, demonstrando o potencial dessas tecnologias como ferramenta de apoio à elaboração de políticas públicas voltadas ao ensino fundamental. Finalmente, são traçadas algumas conclusões acerca da utilização das tecnologias mencionadas.

Visão multidimensional de dados e mineração de dados

O processo de extração de conhecimento útil de grandes volumes de dados é conhecido como descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases ou KDD*). Segundo Fayyad et al (1996), *KDD* pode ser definido como o processo não trivial de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis, embutidos nos dados.

Para poder usufruir da *KDD*, é necessário ter como base os denominados *data warehouses*, ou armazéns de dados. A definição oficial dada por Bill Inmon, um dos criadores do conceito, diz que um *data warehouse* é uma coleção de dados históricos e temporais, integrados em torno de um assunto, de forma não volátil, visando ao auxílio ao processo de tomada de decisão (Inmon, 1997). Há a possibilidade também de se utilizar um *data mart*. Este consiste em um subconjunto de um *data warehouse*, no qual uma porção resumida ou altamente focalizada dos dados é colocada em um banco de dados separado, destinado a uma população específica de usuários.

Visando ilustrar os conceitos de *data warehouse* e *data mart*, poderíamos imaginar um *data warehouse* contendo os dados de todos os alunos, professores, diretores e escolas, assim como os resultados das provas de Língua Portuguesa e Matemática aplicadas em diversos anos para os alunos do ensino básico. Um *data mart*, por sua vez, poderia ser o subconjunto contendo somente os dados do Estado do Rio de Janeiro, com foco nos resultados de proficiência dos alunos em Língua Portuguesa.

O processo de construção de um *data warehouse* (ou de um *data mart*) envolve uma série de etapas. Essas etapas consistem na seleção, passando pela limpeza e transformação, e, finalmente, no carregamento dos dados a partir de uma ou de diversas fontes (Elmasri; Navathe, 2011). Na próxima seção, essas etapas são apresentadas com maior detalhe, com as considerações específicas feitas para a construção do *data mart* contendo os dados ligados ao ensino de Língua Portuguesa no Estado do Rio de Janeiro.

Uma vez construído, o *data warehouse* (ou *data mart*) propicia a construção de relatórios em que é possível a visualização multidimensional, permitindo que os dados sejam vistos a partir de diferentes perspectivas, também chamadas dimensões. Relatórios tradicionais são formatados em uma ou no máximo duas dimensões; nas visões multidimensionais de dados, duas, três ou mais dimensões podem ser apresentadas, aumentando a perspectiva e a visão do problema por parte do tomador de decisão. Outra motivação para a visualização multidimensional de dados é a possibilidade de realizar agregações de dados de diversas formas. No caso do ensino de Língua Portuguesa, seria possível, por exemplo, visualizar as médias dos alunos agrupadas de acordo com seus perfis; ou do perfil dos seus professores; ou então, de acordo com o perfil dos diretores de suas escolas; ou ainda, visualizar as médias agrupadas considerando cada combinação dos perfis de ambas as dimensões (alunos, professores e diretores).

As Figuras 1, 2 e 3 ilustram as afirmações acima. A Figura 1 apresenta modelos de relatórios tradicionais. São apresentados dois relatórios: o primeiro, contendo as médias das notas dos alunos da 8ª série (9º ano) do ensino fundamental do Estado do Rio de Janeiro, na Prova Brasil (Língua Portuguesa) do ano de 2007, de acordo com o grupo étnico dos alunos; o segundo, contendo os mesmos dados, mas agrupados de acordo com a experiência do professor.

Grupo étnico Aluno	Média
Branco(a)	238,74
Amarelo(a)	232,81
Pardo(a)	229,64
Indígena	228,57
Preto(a)	223,86
Média Geral	231,56

(a) Média por raça/cor

Professor (Experiência Lecionando)	Média
Há mais de 20 anos	233,27
De 15 a 20 anos	231,96
De 10 a 15 anos	230,34
De 6 a 9 anos	231,42
De 3 a 5 anos	227,34
De 1 a 2 anos	226,99
Há menos de 1 ano	223,12
Média Geral	231,56

(b) Média por experiência

Figura 1 – Médias de Proficiência em Língua Portuguesa dos Alunos da 8ª Série do Ensino Fundamental do Estado do Rio de Janeiro com Visualização Unidimensional

Na Figura 2, as duas dimensões são exibidas em conjunto, com uma perspectiva mais ampla. Nota-se que tanto a dimensão do professor (com foco em sua experiência) quanto a dimensão do aluno (e seu grupo étnico) são apresentadas no mesmo relatório. Os valores agregados das médias são listados para cada dimensão e para as duas em conjunto. Finalmente, na Figura 3, mais uma dimensão de dados é adicionada ao relatório: diretores de escolas e o agrupamento por sexo. Conforme mencionado, novas agregações são feitas, com cálculos das médias para cada dimensão e para todas as combinações entre as diferentes dimensões. Isso possibilita

ao tomador de decisão avaliar, em um mesmo relatório, a informação agregada sobre diferentes perspectivas. É importante salientar que, neste ponto do artigo, não se objetiva analisar mais profundamente os dados dos relatórios, mas sim apresentar o potencial da tecnologia envolvida.

		PROFESSOR (Experiência lecionando)							Média Geral
		Há mais de 20 anos	De 15 a 20 anos	De 10 a 15 anos	De 6 a 9 anos	De 3 a 5 anos	De 1 a 2 anos	Há menos de 1 ano	
A L U N O	Branco(a)	240,84	238,61	237,84	238,85	232,61	233,33	227,31	238,74
	Amarelo(a)	233,86	234,56	232,14	230,57	229,65	226,93	230,95	232,81
	Pardo(a)	231,12	230,22	228,55	229,53	225,27	225,38	222,66	229,64
	Indígena	229,54	229,22	226,34	229,83	231,48	222,11	217,41	228,57
	Preto(a)	225,57	224,20	221,92	223,10	221,82	221,34	217,37	223,86
	Média Geral	233,27	231,96	230,34	231,42	227,34	226,99	223,12	231,56

Figura 2 – Médias de Proficiência em Língua Portuguesa dos Alunos da 8ª Série do Ensino Fundamental do Estado do Rio de Janeiro com Visualização em Duas Dimensões

		PROFESSOR (Experiência lecionando)							Média (Diretor-sexo x Aluno-etnia)	
		Há mais de 20 anos	De 15 a 20 anos	De 10 a 15 anos	De 6 a 9 anos	De 3 a 5 anos	De 1 a 2 anos	Há menos de 1 ano		
D I R E T O R	F e m i n i n o	Branco(a)	240,15	237,67	236,26	239,14	231,54	226,94	224,60	237,78
		Amarelo(a)	232,40	234,29	229,81	229,61	228,84	223,33	229,20	231,47
		Pardo(a)	230,93	229,99	228,22	229,37	226,10	221,62	221,46	229,42
		Indígena	229,47	230,16	226,26	228,99	230,01	218,72	216,77	228,49
		Preto(a)	225,17	224,24	221,29	223,58	221,24	219,79	214,09	223,49
		Média (Diretor - Feminino)	232,78	231,51	229,34	231,44	227,11	222,90	220,92	230,97
	M A S C U L I N O	Branco(a)	244,67	243,63	244,03	237,41	237,39	253,21	235,47	243,36
		Amarelo(a)	242,14	236,32	242,80	235,94	236,36	236,38	240,80	240,12
		Pardo(a)	232,41	231,60	230,23	230,30	222,75	235,22	226,73	230,94
		Indígena	229,71	222,20	226,42	234,26	244,51	231,92	219,59	228,91
		Preto(a)	227,96	223,96	226,17	219,89	225,00	223,41	235,38	226,00
Média (Diretor - Masculino)	236,16	234,80	235,06	231,23	229,60	238,12	231,19	234,79		
Média Geral	233,29	231,95	230,39	231,41	227,57	226,76	223,12	231,56		

Figura 3 – Médias de Proficiência em Língua Portuguesa dos Alunos da 8ª Série do Ensino Fundamental do Estado do Rio de Janeiro com Visualização em Três Dimensões

Cabe ressaltar que diversas outras combinações de dados (dimensões) são possíveis de serem geradas com bastante facilidade. Por exemplo, ao relatório da Figura 3 poderia ser incorporada mais uma dimensão, por exemplo, a região envolvida. Seria viável, então, a obtenção de um relatório no qual o pesquisador/educador visualizaria as médias dos alunos sob as perspectivas do município, do perfil do diretor da escola, bem como das características dos professores e dos alunos pertencentes à base de dados.

Data warehouses também formam o alicerce para a aplicação da tecnologia de mineração de dados (*Data Mining*). A mineração de dados efetua o casamento da Estatística com a Inteligência Computacional, possibilitando a descoberta de relações ocultas em grandes repositórios de dados e, a partir daí, a inferência de regras para previsão de comportamento futuro e orientação na tomada de decisões (Hirji, 2001).

Exemplo clássico de aplicação de *Data Mining* é o caso em que a cadeia de supermercados *Wall Mart* identificou a tendência, imperceptível num primeiro momento, de que a compra de fraldas estava ligada diretamente à compra de cervejas. Esse caso é citado por Turban, Rainer e Potter (2005). A detecção deste padrão só foi possível por meio da análise computadorizada (*mining*) dos dados das compras efetuadas e das associações entre os produtos vendidos no mesmo tíquete. Essa informação auxiliou a tomada de decisões relacionadas ao *design* das lojas, indicando a necessidade de proximidade entre as seções de bebidas e de produtos para crianças. Assim como o exemplo na área de varejo, diversos padrões ocultos podem ser descobertos ao se aplicar a mineração às bases de dados da educação. Alguns resultados nessa área podem ser encontrados em Namen e Soares (2011).

A descoberta de padrões e tendências escondidos em grandes massas de dados não é um processo trivial. Ela envolve o uso de diferentes tarefas que, por sua vez, se baseiam em diferentes técnicas e algoritmos. O presente trabalho não objetiva detalhar mais profundamente os algoritmos usados, sendo que informações mais detalhadas podem ser encontradas em Hirji (2001), Han e Kamber (2006), MacLennan, Crivat e Tang (2009) e Tan, Steinbach e Kumar (2009).

Bases de dados de educação e aspectos importantes considerados

Conforme mencionado na seção anterior, a construção de um *data warehouse* (ou de um *data mart*) envolve algumas atividades, como a seleção dos dados, sua limpeza, transformação e, finalmente, a carga dos dados a partir de uma ou de diversas fontes. Nesta seção são apresentadas algumas particularidades consideradas na criação do *data mart* contendo os dados relacionados aos alunos do ensino fundamental do Estado do Rio de Janeiro.

A primeira tarefa (seleção dos dados) foi feita a partir da seleção de arquivos da Prova Brasil 2007, disponibilizados pelo Inep em www.inep.gov.br (denominados microdados – Prova Brasil). A cada

dois anos, o governo federal realiza uma prova de Língua Portuguesa e Matemática, denominada Prova Brasil, para todos os alunos da 4ª e 8ª série do ensino fundamental. Além da prova, os alunos respondem um questionário contendo diferentes perguntas relacionadas ao seu perfil socioeconômico. Professores e diretores de escolas também respondem questionários relacionados ao seu perfil e ao da instituição em que trabalham.

Os seguintes arquivos foram selecionados para o presente trabalho: TS_ALUNO (identificação dos alunos juntamente com a proficiência em Língua Portuguesa e Matemática); TS_QUESTION_ALUNO (resposta do questionário aplicado ao aluno de cada série); TS_QUESTION_PROFESSOR (resposta do questionário aplicado ao professor de cada disciplina de cada série); TS_QUESTION_DIRETOR (resposta do questionário do diretor de cada escola); TS_QUESTION_ESCOLA (resposta do questionário aplicado à escola).

Como a pesquisa tinha como foco a análise dos dados do Estado do Rio de Janeiro, foram selecionados, entre os registros existentes nos diferentes arquivos, apenas aqueles com o atributo SIGLA_UF igual ao valor "RJ".

Em seguida, foi realizado o processo de limpeza. Cabe salientar que a execução dessa tarefa foi essencial para garantir a consistência e integridade dos dados a serem analisados.

O arquivo usado como base para todo o processo de limpeza foi o que continha os dados da proficiência dos alunos nas provas de Língua Portuguesa e Matemática (arquivo TS_ALUNO). Foi identificado que diversos desses registros não possuíam correspondência nos arquivos relacionados aos questionários (TS_QUESTION_ALUNO, TS_QUESTION_PROFESSOR, TS_QUESTION_DIRETOR e TS_QUESTION_ESCOLA). Em outras palavras, foram detectadas situações em que existia o dado contendo o resultado da prova realizada por determinado aluno, mas inexistia o dado de um ou mais questionários correspondendo ao próprio aluno, ou ao professor desse aluno, ou à sua escola e/ou seu diretor. Detalhes técnicos quanto ao método utilizado para a exclusão dos registros podem ser encontrados em Namen e Soares (2011).

Para se ter uma ideia dos volumes de dados envolvidos, inicialmente foram selecionados 278.885 registros com os dados de proficiência dos alunos do Estado do Rio de Janeiro (ano de 2007). Destes, 49.166 (cerca de 17% do total) não possuíam registro correspondente nos arquivos representativos dos questionários, registros esses que foram eliminados. Apesar de um número significativo eliminado, os 229.719 registros remanescentes ainda podem ser considerados um volume adequado, tanto para a obtenção de relatórios contendo visões multidimensionais quanto para a detecção de padrões de associação a partir do uso da mineração de dados.

Ressalta-se que não está sendo desconsiderada a importância dos registros eliminados por intermédio do processo de limpeza. Acredita-se que um estudo mais aprofundado pode ser realizado para identificar os motivos dessas inconsistências: Por que alguns dos questionários não foram preenchidos? Há algum padrão de comportamento específico que

possa indicar o motivo do não preenchimento? Como evitar que essas inconsistências se repitam? Contudo, a abordagem inicial do trabalho manteve o foco nas situações em que todas as dimensões relacionadas às informações coletadas pudessem ser analisadas. E, para isso, foi necessária a exclusão desses registros. Entretanto, o volume de dados remanescente ainda era significativo o suficiente para a detecção de padrões de comportamento e descoberta de conhecimento na base de dados.

Finalmente, foi realizado o processo de transformação dos dados para o carregamento no *data mart*. Na próxima seção será apresentada a interface do sistema de geração de relatórios multidimensionais e como esta facilita a criação de relatórios de apoio à decisão por parte do usuário. É importante salientar que a tarefa de transformação dos dados foi fundamental para garantir a agilidade e facilidade propiciada nessa interface.

A principal e mais trabalhosa tarefa de transformação foi converter os dados referentes às respostas dos questionários em dados que pudessem ser lidos de modo fácil por parte do pesquisador/educador. Para exemplificar, tomemos um dos questionários envolvidos no processo, o do aluno de 4ª série do ensino fundamental. Este é composto por um total de 44 questões. Para acesso a mais detalhes, todos os questionários estão disponíveis no *site* do Inep, no endereço www.inep.gov.br.

Para cada questão, a respectiva resposta pode ser preenchida a partir da utilização de letras do alfabeto. Por exemplo: na primeira pergunta do questionário do aluno da 4ª série, relacionada ao sexo, a letra "A" corresponde à resposta identificando o sexo masculino, enquanto a letra "B" indica o sexo feminino. Do mesmo modo, a questão 2, relacionada ao grupo étnico do estudante, apresenta as seguintes respostas: "Branco(a)" (letra "A"); "Pardo(a)" (letra "B"); "Preto(a)" (letra "C"); "Amarelo(a)" (letra "D"); "Indígena" (letra "E"). Lógica semelhante se replica para todas as 44 perguntas do questionário do aluno. Essa lógica também se aplica aos questionários dos professores, dos diretores e das escolas.

Nos arquivos do Inep, os dados correspondentes às diferentes perguntas dos questionários são armazenados com as letras correspondentes às respectivas respostas. Retornando ao exemplo do questionário do aluno, as 44 perguntas e suas respectivas respostas são armazenadas em um único campo, contendo 44 posições, no qual cada uma das 44 posições é preenchida com a letra do alfabeto correspondente à resposta dada. Por exemplo, a primeira das 44 posições do campo de respostas pode ser preenchida com a letra "A" (sexo masculino) ou a letra "B" (sexo feminino). A segunda posição pode ser preenchida com as letras "A" a "E", cada uma delas identificando diferentes grupos étnicos, e assim sucessivamente. A Figura 4 ilustra a representação de um registro com os dados referentes ao questionário de um aluno. As primeiras 72 posições contêm o código identificador do aluno, sua série e dados sobre a escola: se é federal, estadual ou municipal, se é urbana ou rural e o Estado e o Município onde está localizada. Da 73ª posição em diante, localiza-se o campo contendo as 44 respostas do respectivo aluno. Mais detalhes sobre o leiaute e o significado dos diferentes campos nos arquivos do Inep podem ser obtidos

na publicação Microdados da Prova Brasil 2007: manual do usuário (Brasil. Inep, 2009).

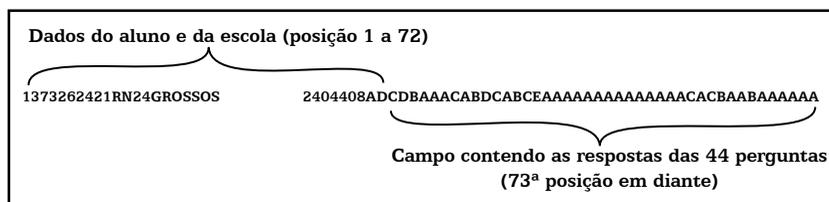


Figura 4 – Representação de um Registro Contendo os Dados do Questionário de um Aluno da 4ª Série

Pode-se concluir que essa representação dificulta qualquer entendimento e interpretação direta dos dados. Por esse motivo, para o caso específico do questionário do aluno da 4ª série, o processo de transformação para o *data mart* exigiu a criação de 44 atributos (campos) diferentes, cada um representando uma determinada pergunta do questionário. Cada um desses campos foi carregado, por sua vez, não com a letra da resposta (que era o dado existente no arquivo), mas sim com o texto da resposta apresentado no questionário. Retornando à primeira pergunta, que versava sobre o sexo do aluno, foi realizada a transformação da letra “A” para o texto “Masculino”, sendo este último o dado carregado no *data mart*. Idem para “B”, que passou a ser “Feminino”.

Não é objetivo deste trabalho entrar em mais detalhes sobre os algoritmos e o procedimento técnico utilizado nessa transformação. Essas informações podem ser encontradas em Namen e Soares (2011). Assim, é importante deixar claro que, sem a realização de um processo de transformação dos dados das bases de dados do Inep, não seria possível a apresentação das informações de forma rápida e bastante clara para os educadores. Entende-se que a criação do *data mart* visa, exatamente, preparar um ambiente em que o tomador de decisão acesse a informação desejada diretamente, sem necessitar realizar quaisquer conversões ou ter qualquer trabalho adicional. A próxima seção ilustrará claramente essas afirmações.

Vantagens de uso das tecnologias

Oferecer um ambiente de fácil uso, permitir a montagem rápida de relatórios, por intermédio de apenas alguns cliques no *mouse*, agilizar a extração da informação e do conhecimento, enfim, aumentar o poder do educador na análise e obtenção de conhecimento das bases de dados de educação. Esse é um dos grandes desafios para os pesquisadores quando se fala na obtenção de indicadores de qualidade na área de educação.

As tecnologias discutidas no presente artigo permitem lidar com o desafio acima exposto, e a presente seção visa deixar claro como estas podem contribuir para a obtenção de conhecimento em grandes bases de

dados de educação, podendo servir de fundamento para futuros trabalhos na área.

Dentro dessa perspectiva, inicialmente serão apresentadas algumas figuras para ilustrar o ambiente de geração de relatórios. A Figura 5 apresenta a interface de geração de relatórios que utiliza como base o *data mart* com os dados da Prova Brasil 2007 dos alunos do ensino básico do Estado do Rio de Janeiro, com foco nos resultados de proficiência em Língua Portuguesa. Esse ambiente pertence ao Gerenciador de Banco de Dados SQL Server 2008 e foi utilizado para a confecção de diversos relatórios usando o conceito de visão multidimensional e também minerações de dados.

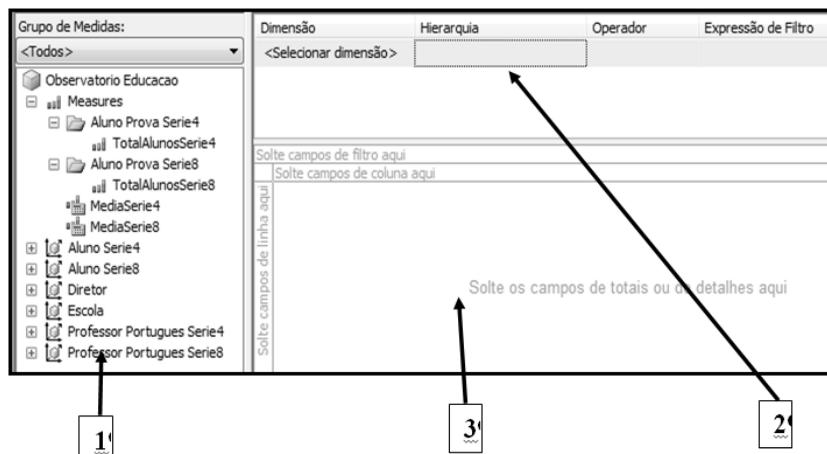


Figura 5 – Interface para Geração de Relatórios

A janela pode ser dividida em três áreas. A área 1, localizada no lado esquerdo, contém os dados que serão agregados, ou medidas (*Measures*). Esses dados são relacionados aos totais de alunos que realizaram as provas tanto da 4ª quanto da 8ª série do ensino fundamental (*TotalAlunosSerie4* e *TotalAlunosSerie8*), bem como as médias das notas nas provas de Língua Portuguesa (*MediaSerie4* e *MediaSerie8*). Na área de número 1 também são encontradas as diferentes dimensões que poderão ser visualizadas quando da criação dos relatórios. Essas dimensões estão relacionadas aos questionários e representam os dados contendo os perfis dos alunos (dimensões *Aluno Serie4* e *Aluno Serie8*), dos professores de Língua Portuguesa (*Professor Portugues Serie4* e *Professor Portugues Serie8*), das escolas (*Escola*) e de seus diretores (*Diretor*).

A área 2, localizada na parte superior à direita, possibilita a inserção de parâmetros para filtragem dos dados. Por exemplo, como o *data mart* possui os dados de todo o Estado do Rio de Janeiro, essa área poderia ser utilizada pelo usuário para fazer a filtragem dos dados somente do município de Niterói. Finalmente, na parte inferior à direita, localiza-se o espaço onde são construídos os relatórios (área 3).

As dimensões apresentadas na área 1 podem ser expandidas na janela. Ao efetuar a expansão (clique no botão +, localizado ao lado

esquerdo do nome de cada dimensão), as perguntas dos questionários podem ser visualizadas. A Figura 6 apresenta a janela com a dimensão correspondente aos alunos da 8ª série expandida (*Aluno Serie8*). É importante salientar que, ao expandir cada dimensão, todas as perguntas do respectivo questionário são apresentadas, apesar de não ser possível a visualização de todas elas na Figura 6, devido às limitações de espaço deste artigo.

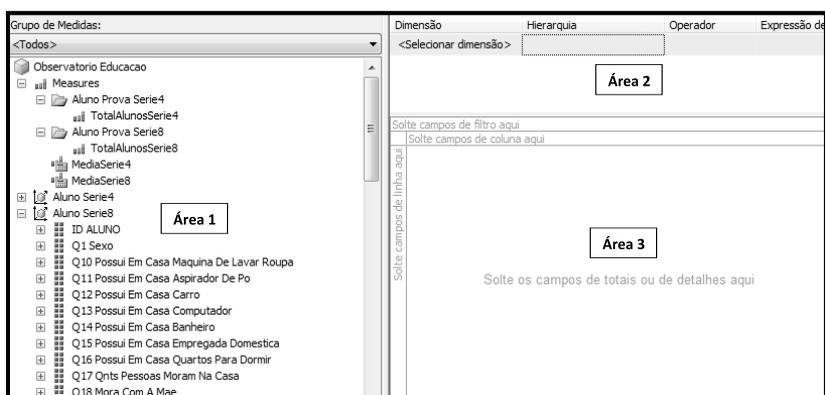


Figura 6 – Expansão da Dimensão Correspondente aos Alunos da 8ª Série do Ensino Fundamental

Nota-se que os atributos da dimensão estão nomeados a partir do seu número, seguido da questão propriamente dita. Por exemplo, o atributo *Q1Sexo* corresponde à primeira pergunta do questionário do aluno da 8ª série, que versa sobre seu sexo. *Q18MoraComAMae*, por sua vez, corresponde ao questionamento se o aluno mora com a mãe (pergunta 18).

Para criar qualquer relatório, basta clicar e arrastar os campos existentes na área 1 para o local de apresentação de relatórios (área 3) ou para a área 2, onde é feita a filtragem de dados. Ressalta-se que a criação de qualquer novo relatório envolvendo uma, duas ou mais dimensões pode ser feita em poucos segundos. Os valores agregados (médias e totais) são calculados automaticamente e apresentados para cada dimensão considerada na geração dos relatórios.

A Figura 7 apresenta um exemplo de relatório que possui as informações das médias dos alunos da 4ª série do ensino fundamental da cidade de Niterói, com visualização a partir da perspectiva do aluno (se ele já foi ou não reprovado) e da perspectiva do professor (se este já sofreu ou não agressão física por parte de algum aluno).

Para obtenção deste relatório, foi necessário, inicialmente, expandir a dimensão Escola, selecionar o atributo com o nome do município (*NO MUNICIPIO*) e arrastá-lo para a área 2 (de filtragem), complementando os dados com o operador de igualdade e o campo expressão de filtro com a cidade de Niterói. Em seguida, os campos *MediaSerie4* e *TotalAlunosSerie4*, referentes a medidas e, por isso, sujeitos a cálculos agregados, foram arrastados para a região central dentro da área 3.

Finalmente, os passos seguintes foram a expansão das duas dimensões consideradas no relatório – *AlunoSerie4* e *ProfessorPortuguesSerie4* – e o posterior arraste do atributo *Q38JaFoiReprovado*, indicador de reprovação ou não do aluno (pergunta 38 do questionário do aluno), para a região indicando o eixo das linhas e do atributo *Q99AgressaoFisicaProfPorAluno*, relacionado ao fato de o professor ter ou não sofrido agressão física por parte de algum aluno (questionário do professor – pergunta 99), para o eixo das colunas, tudo isso dentro da área 3.

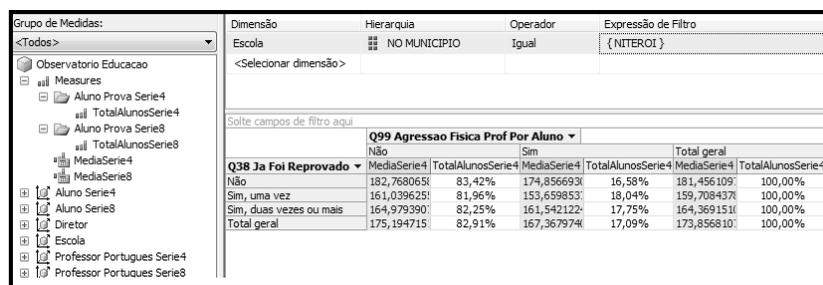


Figura 7 – Exemplo de Relatório Gerado pela Ferramenta

Pode-se perceber que o relatório contém todas as características, já apresentadas, da utilização da tecnologia de visão multidimensional de dados. Mais ainda, fica clara a facilidade de manipulação e criação de relatórios de auxílio à decisão, mesmo que se feito por usuários sem maiores conhecimentos técnicos na área de computação ou bancos de dados. A construção de relatórios que representem os resultados dos alunos nas provas de Língua Portuguesa, abrangendo informações que caracterizam os perfis de alunos, professores, diretores e escolas, seja isoladamente, seja em diferentes combinações das dimensões, é possibilitada a partir de interação com uma interface gráfica simples que provê o cálculo automatizado de médias e totais de alunos, com resposta rápida e eficiente.

Fazendo uma análise inicial do relatório apresentado na Figura 7, pode-se perceber que alunos que nunca foram reprovados tendem a obter melhores resultados (maiores médias) nas provas. Analisando sob a perspectiva da violência, pode-se inferir também que a proporção de alunos que já foram reprovados tendo aulas com professores que já sofreram algum tipo de agressão (18,04% e 17,75%) é maior do que a proporção dos alunos que nunca sofreram reprovação (16,58%). Evidentemente, essa é apenas uma constatação inicial a partir de uma primeira análise, que poderia ser aprofundada para identificar causas, fatores e condições que levaram a esses resultados. Assim, pode-se concluir que a ferramenta permite a fácil geração de relatórios que descortinem novas perspectivas aos educadores e permitam, conforme já mencionado, a descoberta de conhecimento a partir das bases de dados de educação.

Apresentação de alguns indicadores relacionados ao ensino fundamental

Na presente seção, são apresentados apenas alguns indicadores obtidos com o uso das tecnologias apresentadas, devido às limitações inerentes a um artigo acadêmico. Evidentemente, o universo de trabalho e as possibilidades de apresentação de outros indicadores são muito mais amplos.

Conforme mencionado anteriormente, a mineração de dados pode ser realizada com a utilização de diferentes técnicas e algoritmos e possibilita a identificação de padrões e de relacionamentos entre diferentes variáveis. Em Namen e Soares (2011), algumas associações relacionadas ao perfil dos alunos da 4ª série do ensino fundamental foram identificadas a partir da utilização do algoritmo computacional denominado *Apriori* – detalhes do algoritmo podem ser encontrados em Han e Kamber (2006), Tan, Steinbach e Kumar (2009) e Elmasri e Navathe (2011). Elas envolviam variáveis como a atuação dos alunos em trabalhos domésticos ou fora de casa, o estímulo de seus pais às atividades de estudo (ou a falta dele) e a faixa etária dos alunos e sua proficiência em Língua Portuguesa.

Outros resultados de minerações foram obtidos a partir da análise de atributos relacionados ao perfil dos professores. Por intermédio da utilização de algoritmos denominados *Classificadores Bayesianos* (veja detalhes em Tan, Steinbach e Kumar, 2009), foram identificadas diversas relações entre as respostas ao questionário do professor (relacionadas ao seu perfil) e as médias de seus alunos, apresentadas na Tabela 1.

A Tabela 1 possui quatro colunas. A primeira e a segunda colunas identificam uma determinada pergunta do questionário e uma resposta a ela, respectivamente. As duas últimas colunas, se preenchidas com X, indicam que a resposta dada privilegia aquele comportamento relacionado à proficiência dos alunos em Língua Portuguesa. Assim, se a terceira coluna estiver marcada, indicará que o fato de um professor ter respondido daquela forma a uma determinada questão favoreceu o fato de que a maioria dos alunos (65% ou mais) obteve resultados nas provas de Língua Portuguesa abaixo da média. Da mesma forma para a quarta coluna, onde o favorecimento seria relacionado a resultados acima da média. Exemplificando, a primeira linha da Tabela 1 indica que os professores que desenvolveram mais de 80% do conteúdo previsto junto aos alunos (pergunta 55 do questionário do professor) tenderam a obter resultados em que a grande maioria de seus discentes obtiveram proficiência em Língua Portuguesa acima da média.

Nota-se que a mineração possibilitou a descoberta de diversos padrões, associando o perfil do professor ao desempenho dos seus alunos. Dentro desse universo, foram selecionadas algumas questões, marcadas em cinza, relacionadas às expectativas dos professores em relação à situação futura de seus alunos. Essas perguntas versam sobre a expectativa do professor quanto ao universo de alunos que concluirão o ensino fundamental (*Q56 Qtd Alunos Conc Ens Fund*), o ensino médio

(Q57 Qtd Alunos Conc Ens Medio) e que entrarão na universidade (Q58 Qtd Alunos Entrarao Universidade).

Tabela 1 – Relações entre Respostas do Questionário do Professor e Desempenho dos Alunos da 4ª Série do Ensino Fundamental do Estado do Rio de Janeiro

(continua)

Pergunta	Resposta	Favorece grande maioria dos alunos abaixo da média	Favorece grande maioria dos alunos acima da média
(Q55) Porcentagem do conteúdo previsto desenvolvido com os alunos	Mais de 80%		X
(Q57) Expectativa do professor quanto ao universo de alunos que concluíram o ensino médio	Quase todos os alunos		X
(Q56) Expectativa do professor quanto ao universo de alunos que concluíram o ensino fundamental	Quase todos os alunos		X
(Q55) Porcentagem do conteúdo previsto desenvolvido com os alunos	Entre 40% e 60%	X	
(Q58) Expectativa do professor quanto ao universo de alunos que entrarão na universidade	Um pouco mais da metade dos alunos		X
(Q95) Ocorreram problemas disciplinares causados pelos alunos	Não		X
(Q94) Houve alto índice de faltas por parte dos alunos	Não		X
(Q56) Expectativa do professor quanto ao universo de alunos que concluíram o ensino fundamental	Um pouco mais da metade dos alunos	X	
(Q116) O professor foi agredido verbalmente por algum aluno	Não		X
(Q58) Expectativa do professor quanto ao universo de alunos que entrarão na universidade	Poucos alunos	X	
(Q116) O professor foi agredido verbalmente por algum aluno	Sim	X	
(Q57) Expectativa do professor quanto ao universo de alunos que concluíram o ensino médio	Um pouco menos da metade dos alunos	X	
(Q44) O professor utiliza computadores nesta escola	Sim, utilizo		X
(Q57) Expectativa do professor quanto ao universo de alunos que concluíram o ensino médio	Poucos alunos	X	
(Q44) O professor utiliza computadores nesta escola	Não utilizo porque a escola não tem	X	
(Q95) Ocorreram problemas disciplinares causados pelos alunos	Sim, e foi um problema grave	X	
(Q65) As dificuldades de aprendizagem são decorrentes da indisciplina dos alunos em sala de aula	Discordo		X

Tabela 1 – Relações entre Respostas do Questionário do Professor e Desempenho dos Alunos da 4ª Série do Ensino Fundamental do Estado do Rio de Janeiro

(continuação)

Pergunta	Resposta	Favorece grande maioria dos alunos abaixo da média	Favorece grande maioria dos alunos acima da média
(Q45) O professor utiliza internet nesta escola	Sim, utilizo		X
(Q45) O professor utiliza internet nesta escola	Não utilizo porque a escola não tem	X	
(Q115) O professor foi ameaçado por algum aluno	Não		X
(Q89) Ocorreram problemas de inexistência de professores para algumas disciplinas ou séries Disciplina	Não		X
(Q58) Expectativa do professor quanto ao universo de alunos que entrarão na universidade	Quase todos os alunos		X
(Q30) Frequência de atividades de LP que permitem aos alunos conversar sobre contos, crônicas, poesias e romances	Mais de uma vez por mês		X
(Q58) Expectativa do professor quanto ao universo de alunos que entrarão na universidade	Não sei	X	
(Q65) As dificuldades de aprendizagem são decorrentes da indisciplina dos alunos em sala de aula	Concordo	X	
(Q61) As dificuldades de aprendizagem são decorrentes de insegurança física na escola	Discordo		X
(Q29) Frequência de atividades de LP permitem aos alunos ler contos, crônicas, poesias e romances	Mais de uma vez por mês		X
(Q89) Ocorreram problemas de inexistência de professores para algumas disciplinas ou séries Disciplina	Sim, e foi um problema grave	X	
(Q115) O professor foi ameaçado por algum aluno	Sim	X	
(Q94) Houve alto índice de faltas por parte dos alunos	Sim, e foi um problema grave	X	
(Q131) Avaliação dos livros didáticos na(s) disciplina(s) que ministra	Ótimo		X
(Q131) Avaliação dos livros didáticos na(s) disciplina(s) que ministra	Ruim	X	
(Q105) Houve agressão física a alunos por aluno	Não		X
(Q61) As dificuldade de aprendizagem são decorrentes de insegurança física na escola	Concordo	X	
(Q30) Frequência de atividades de LP que permitem aos alunos conversar sobre contos, crônicas, poesias e romances	Entre uma vez ou duas vezes por bimestre	X	

Tabela 1 – Relações entre Respostas do Questionário do Professor e Desempenho dos Alunos da 4ª Série do Ensino Fundamental do Estado do Rio de Janeiro

(continuação)

Pergunta	Resposta	Favorece grande maioria dos alunos abaixo da média	Favorece grande maioria dos alunos acima da média
(Q58) Expectativa do professor quanto ao universo de alunos que entrarão na universidade	Um pouco menos da metade dos alunos		X
(Q29) Frequência de atividades de LP permitem aos alunos ler contos, crônicas, poesias e romances	Entre uma vez ou duas vezes por bimestre	X	
(Q56) Expectativa do professor quanto ao universo de alunos que concluíram o ensino fundamental	Um pouco menos da metade dos alunos	X	
(Q4) Nível máximo de escolaridades até a graduação	Ensino Superior		X
(Q105) Houve agressão física a alunos por aluno	Sim	X	
(Q118) O professor foi vítima de furto na escola	Não		X
(Q4) Nível máximo de escolaridades até a graduação	Ensino Médio ou menos	X	
(Q118) O professor foi vítima de furto na escola	Sim	X	
(Q94) Houve alto índice de faltas por parte dos alunos	Sim, mas não foi um problema grave	X	
(Q130) O livro didático escolhido foi o recebido	Sim		X
(Q131) Avaliação dos livros didáticos na(s) disciplina(s) que ministra	Razoável	X	
(Q55) Porcentagem do conteúdo previsto desenvolvido com os alunos	Menos de 40%	X	
(Q57) Expectativa do professor quanto ao universo de alunos que concluíram o ensino médio	Não sei	X	
(Q121) Houve alunos que frequentaram as suas aulas sob efeito de drogas ilícitas	Não		X
(Q131) Avaliação dos livros didáticos na(s) disciplina(s) que ministra	Bom		X
(Q130) O livro didático escolhido foi o recebido	Resposta nula	X	
(Q55) Porcentagem do conteúdo previsto desenvolvido com os alunos	Entre 60% e 80%	X	
(Q121) Houve alunos que frequentaram as suas aulas sob efeito de drogas ilícitas	Sim	X	
(Q56) Expectativa do professor quanto ao universo de alunos que concluíram o ensino fundamental	Não sei	X	

Tabela 1 – Relações entre Respostas do Questionário do Professor e Desempenho dos Alunos da 4ª Série do Ensino Fundamental do Estado do Rio de Janeiro

(conclusão)

Pergunta	Resposta	Favorece grande maioria dos alunos abaixo da média	Favorece grande maioria dos alunos acima da média
(Q30) Frequência de atividades de LP que permitem aos alunos conversar sobre contos, crônicas, poesias e romances	Nunca	X	
(Q29) Frequência de atividades de LP permitem aos alunos ler contos, crônicas, poesias e romances	Ausente	X	
(Q118) O professor foi vítima de furto na escola	Resposta nula	X	
(Q130) O livro didático escolhido foi o recebido	Não	X	

Fazendo uma análise mais apurada dessas informações, poderia se afirmar que, quanto maiores as expectativas dos professores quanto à futura formação educacional de seus alunos, melhores os resultados obtidos. Para corroborar esse fato, usou-se o gerador de relatórios, apresentado na seção anterior, como ferramenta complementar, gerando-se o relatório apresentado na Figura 8, que foca suas lentes sobre a pergunta 57.

Expectativa quanto à conclusão do Ensino Médio (questão 57)	Média	Total de Alunos
Quase todos os alunos	183,35	31,10%
Um pouco mais da metade dos alunos	176,74	37,09%
Não sei	172,42	9,08%
Um pouco menos da metade dos alunos	171,29	16,14%
Poucos alunos	168,15	6,59%
Média geral	176,96	100,00%

Figura 8 – Médias de Proficiência em Língua Portuguesa (Estado do Rio de Janeiro – 2007) de acordo com a Expectativa do Professor em relação à Conclusão do Ensino Médio pelos seus Alunos

As informações aqui obtidas demonstram como as expectativas dos professores afetam o desempenho dos alunos. Os primeiros trabalhos que abordam esse fenômeno podem ser encontrados em Rosenthal e Jacobson (1966, 1968). Segundo os autores, professores que têm uma visão positiva dos alunos tendem a estimular o lado bom desses alunos, que tendem a obter melhores resultados; inversamente, professores que não têm apreço por seus alunos adotam posturas que acabam por comprometer negativamente o desempenho dos educandos.

Trabalhos mais recentes abordam também esse fenômeno. Earp (2009) fala sobre as crenças e as formas de agir dos professores de escolas

cariocas. Giagaglia e Penteadó (2006) mencionam a responsabilidade dos orientadores educacionais ao tratar sigilosamente algumas informações de alunos, visando não criar possíveis expectativas negativas junto aos professores. Alves e Passador (2011) concluem que escolas cujos alunos têm maior nível socioeconômico obtiveram melhor desempenho nas avaliações do Inep. Apesar de apontarem diversos outros fatores, e não as expectativas dos professores em relação aos seus alunos (e a relação com seu nível socioeconômico), estas últimas poderiam também compor o conjunto de fatores causais relacionados à conclusão obtida.

Na Tabela 1 também é possível observar que foram identificados alguns padrões de comportamento relacionados às atividades didáticas realizadas com os alunos. Algumas linhas da tabela apresentam tendências relacionadas à realização de atividades de leitura e conversas sobre contos, crônicas, poesias e romances (perguntas 29 e 30 do questionário do professor). Dentro dessa perspectiva, despertou-se o interesse de realizar uma análise do comportamento das médias dos alunos, de acordo com a frequência de realização de diferentes atividades didáticas por parte dos professores de Língua Portuguesa, frequência essa informada a partir das respostas destes últimos às questões 24 a 33. Na Figura 9, podem ser visualizadas diferentes perspectivas em relação às atividades realizadas com os alunos de Língua Portuguesa, sua frequência e a proficiência dos alunos.

(continua)

Atividades de LP permitem aos alunos copiar textos extensos do livro didático ou do quadro negro (Q24)	Média	% Alunos	Atividades de LP permitem aos alunos ler contos, crônicas, poesias ou romances (Q29)	Média	% Alunos
Nunca	179,23	39,4%	Semanalmente	178,37	54,6%
Uma vez por bimestre	177,79	7,4%	Algumas vezes por mês	176,16	36,3%
Uma vez por mês	176,23	10,6%	Uma vez por mês	172,32	6,3%
Algumas vezes por mês	175,35	24,6%	Uma vez por bimestre	171,34	2,2%
Semanalmente	173,90	17,9%	Nunca	169,86	0,5%
Média Geral	176,96	100%	Média Geral	176,96	100%

Atividades de LP permitem aos alunos conversar sobre textos de jornais e revistas (Q25)	Média	% Alunos	Atividades de LP permitem aos alunos conversar sobre contos, crônicas, poesias ou romances (Q30)	Média	% Alunos
Semanalmente	177,22	59,7%	Semanalmente	178,39	49,6%
Algumas vezes por mês	176,97	34,8%	Algumas vezes por mês	176,45	40,5%
Uma vez por bimestre	174,56	1,1%	Uma vez por bimestre	172,84	2,1%
Uma vez por mês	173,67	3,5%	Uma vez por mês	171,39	7,0%
Nunca	171,41	0,9%	Nunca	168,09	0,8%
Média Geral	176,96	100%	Média Geral	176,96	100%

Atividades de LP permitem aos alunos fazer exercícios sobre gramática relacionados com textos de jornais e revistas (Q26)	Média	% Alunos	Atividades de LP permitem aos alunos usar contos, crônicas, poesias ou romances para exercitar aspectos da gramática (Q31)	Média	% Alunos
Algumas vezes por mês	177,83	41,5%	Semanalmente	177,94	49,4%
Semanalmente	176,56	49,6%	Algumas vezes por mês	176,64	41,6%
Uma vez por bimestre	174,93	2,0%	Uma vez por mês	173,04	5,9%
Uma vez por mês	174,80	4,9%	Nunca	172,09	1,4%
Nunca	173,71	1,9%	Uma vez por bimestre	171,59	1,7%
Média Geral	176,96	100%	Média Geral	176,96	100%

Figura 9 – Atividades e Mecanismos Didáticos Utilizados pelos Professores no Ensino de Língua Portuguesa (LP) e Médias Obtidas pelos Alunos da 4ª Série do Ensino Fundamental do Estado do Rio de Janeiro

(conclusão)

Atividades de LP permitem aos alunos automatizar o uso de regras gramaticais (Q27)	Média	% Alunos
Uma vez por bimestre	181,25	2,4%
Nunca	178,25	7,3%
Algumas vezes por mês	177,47	31,6%
Uma vez por mês	176,35	4,7%
Semanalmente	176,35	54,0%
Média Geral	176,96	100%

Atividades de LP permitem aos alunos fixar os nomes de conceitos gramaticais e linguísticos (Q32)	Média	% Alunos
Nunca	178,85	6,0%
Uma vez mês	178,14	6,1%
Algumas vezes por mês	177,99	38,8%
Semanalmente	175,68	46,3%
Uma vez por bimestre	175,47	2,8%
Média Geral	176,96	100%

Atividades de LP permitem aos alunos ler, discutir com colegas e escrever textos relacionados com o desenvolvimento de projeto temático (Q28)	Média	% Alunos
Semanalmente	178,19	49,1%
Uma vez por mês	176,58	7,4%
Nunca	176,29	1,4%
Algumas vezes por mês	175,84	38,0%
Uma vez por bimestre	173,25	4,2%
Média Geral	176,96	100%

Atividades de LP permitem aos alunos discutir um texto, explorando as diferenças entre fatos e opiniões (Q33)	Média	% Alunos
Semanalmente	177,62	73,3%
Algumas vezes por mês	175,50	23,4%
Uma vez por mês	173,10	2,4%
Nunca	171,39	0,3%
Uma vez por bimestre	170,32	0,5%
Média Geral	176,96	100%

Figura 9 – Atividades e Mecanismos Didáticos Utilizados pelos Professores no Ensino de Língua Portuguesa (LP) e Médias Obtidas pelos Alunos da 4ª Série do Ensino Fundamental do Estado do Rio de Janeiro

Analisando as médias, é fácil constatar que determinadas atividades, quando realizadas com maior frequência, impactam positivamente os resultados do processo de ensino-aprendizagem. Esse é o caso das atividades que permitem a discussão e a prática de exercícios relacionados com textos de jornais e revistas ou contos, crônicas, poesias e romances. Por outro lado, maior frequência na aplicação de atividades como cópias extensas de textos no quadro, foco na automatização do uso de regras gramaticais ou fixação dos nomes de conceitos gramaticais e linguísticos indicam piores resultados na qualidade do processo de ensino-aprendizagem.

Conforme mencionado, são apresentados aqui apenas alguns indicadores de qualidade do ensino fundamental, obtidos com o apoio das tecnologias de mineração e visualização multidimensional de dados, demonstrando seu potencial quando aplicadas na área de educação.

Conclusões

O presente artigo teve como principal objetivo apresentar o potencial de uso das tecnologias de mineração e visualização multidimensional de dados e as facilidades que estas podem oferecer ao educador e/ou elaborador de políticas públicas em seu processo de análise e posterior tomada de decisão. É importante deixar claro, contudo, que existem diversas reflexões a serem feitas quanto aos indicadores gerados a partir do uso das tecnologias aqui apresentadas.

Alves e Passador (2011) abordam a limitação quanto à natureza dos dados utilizados. Segundo os autores, o fato de vários alunos não responderem a questão sobre a existência de itens ou serviços em suas

casas leva à reflexão quanto à necessidade de ajustes nos questionários ou uma melhor orientação para o seu preenchimento.

Alves e Soares (2009), por sua vez, observam as limitações da medição do *status* socioeconômico dos estudantes a partir dos questionários a eles aplicados, por exemplo, perguntas relacionadas à posse de itens como telefones, computador e DVD, que inicialmente são acessíveis apenas a grupos com maior poder aquisitivo, ao longo do tempo se tornam acessíveis aos indivíduos das camadas menos favorecidas. Os autores também observam que questões relacionadas ao nível de escolaridade (por exemplo, dos pais dos alunos) não necessariamente compõem um indicador socioeconômico consistente, já que, atualmente, a diferenciação social não se dá apenas pelos anos de estudo, mas também pela diferença da qualidade de educação a que se tem acesso, informação esta difícil de ser avaliada por intermédio de questionários.

Também são levantadas críticas acerca das formas como os indicadores são analisados e divulgados. Alguns impactos negativos podem ocorrer na comunidade escolar quando os resultados são amplamente divulgados na mídia. Ao serem divulgados para a grande massa que desconhece os problemas reais das escolas públicas e seus desafios, os resultados reforçam a imagem negativa que a sociedade tem do ensino público. Efeitos adversos, como resistências na comunidade escolar, gerando dificuldades na utilização dos resultados para ações corretivas, podem ser formados por intermédio dessa abordagem (Vianna, 2003, Ribeiro; Gusmão, 2005).

André (2009) argumenta que, apesar de os testes nacionais cumprirem uma importante função, apresentando a situação do ensino e da aprendizagem nas escolas e chamando a atenção para a necessidade de melhorar a formação dos professores, é preciso cautela para não cair na armadilha de que a aprendizagem fique reduzida ao ensino de conteúdos e ao que for passível de mensuração, deixando de fora aspectos extremamente ricos, como o desenvolvimento de valores, atitudes, convivência social, solidariedade e práticas culturais.

Conforme defendido por Ravitch (2011), o foco de análise e das ações deve ser sempre na melhoria da educação e não simplesmente no aumento das pontuações nas provas de avaliação. É preciso formar jovens com bom embasamento em História, Ciência, Geografia, Matemática, Leitura, e não apenas formar gerações que aprendam a responder testes de múltipla escolha. Além disso, não devem ser replicados no campo da administração educacional os velhos modelos da gestão empresarial, que acreditam que bastaria a introdução de mecanismos de mercado na gestão educacional para que os problemas fossem resolvidos.

Apesar das limitações dos indicadores, acredita-se que eles podem fornecer informações importantes sobre alunos, professores, diretores e escolas. A comparação de desempenho das redes administradas por

diferentes esferas de governo em diferentes regiões, desde que não utilizadas para a geração de um simples esquema de competição, tem sua utilidade para a elaboração de políticas públicas na área de educação (Alves; Passador, 2011).

É importante salientar que as atividades de criação do *data warehouse* ou *data mart* são fundamentais para a utilização consistente das ferramentas de mineração e visão multidimensional. As tarefas de seleção, limpeza, transformação e carga dos dados formam o alicerce que possibilita a posterior descoberta de conhecimento em bancos de dados (*KDD*). Esse processo envolve elevado nível de trabalho e tempo, mas garante posterior facilidade e agilidade para a geração de relatórios de apoio à decisão.

O artigo apresentou alguns indicadores relacionados ao aprendizado de Língua Portuguesa no ensino fundamental. Foi apresentada apenas uma pequena porção do que pode ser gerado em termos de informações que apoiem a descoberta de conhecimento ligado às bases de dados da educação. Optou-se por apresentar os relatórios em formatos tabulares, mas seria perfeitamente viável a sua apresentação também na forma de gráficos, o que poderia auxiliar no processo de visualização e facilitar a identificação de tendências.

Abordando especificamente os relatórios que contêm as relações entre as expectativas dos professores quanto à futura formação educacional de seus alunos e os desempenhos destes, uma análise mais aprofundada poderia ser realizada. A incorporação de outras dimensões, como os perfis de alunos, escolas e professores ao relatório, enriqueceria as informações. Poderiam ser correlacionadas, por exemplo, características das escolas a tendências de expectativas dos professores em relação ao futuro dos alunos. Ou ainda, perfis socioeconômicos dos alunos e as expectativas (positivas ou negativas) a eles direcionadas.

Diversos relatórios multidimensionais poderiam ser gerados comparando os municípios, identificando aspectos como formação e capacitação continuada dos professores, infraestrutura de escolas, enfim, diferentes aspectos (e dimensões) do problema. Além disso, somente foram citadas/apresentadas duas técnicas de mineração de dados: utilização do algoritmo *Apriori* para identificação de padrões de associação entre variáveis e *Classificadores Bayesianos*. Diversos outros algoritmos/técnicas podem ser utilizados para a mineração, como redes neurais, árvores de decisão, análise de *clusters*, entre outros. Evidentemente, conforme já mencionado, devido às limitações de espaço, este artigo não se propõe a apresentar todas essas possibilidades. Desse modo, evidencia-se que as ferramentas aqui apresentadas descortinam diversas oportunidades relacionadas à descoberta de conhecimento em bases de dados da educação.

Referências bibliográficas

ALVES, M. T. G.; SOARES, J. F. Medidas de nível socioeconômico em pesquisas sociais: uma aplicação aos dados de uma pesquisa educacional. *Opinião Pública*, Campinas, v. 15, n. 1, p. 1-30, jun. 2009.

ALVES, T.; PASSADOR, C. S. *Educação pública no Brasil: condições de oferta, nível socioeconômico dos alunos e avaliação*. São Paulo: Annablume; Brasília: Capes, Inep, 2011.

ANDRÉ, M. A complexa relação entre pesquisas e políticas públicas no campo da formação de professores. *Educação*, Porto Alegre, v. 32, n. 3, p. 270-276, set./dez. 2009.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Estudo exploratório sobre o professor brasileiro com base nos resultados do Censo Escolar da Educação Básica 2007*. Brasília, 2009a. Disponível em: <<http://portal.mec.gov.br/dmdocuments/estudoprofessor.pdf>>.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Microdados da Prova Brasil 2007: manual do usuário*. Brasília, 2009b.

EARP, M. L. S. A cultura da repetência em escolas cariocas. Ensaio: *Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 17, n. 65, p. 613-632, out./dez. 2009.

ELMASRI, R.; NAVATHE, S. B. *Sistemas de banco de dados*. São Paulo: Pearson Addison Wesley, 2011.

FAYYAD, U. M. et al. *Advances in knowledge discovery & data mining*. Cambridge: AAAI/MIT, 1996.

GIACAGLIA, L. R. A.; PENTEADO, W. M. A. *Orientação educacional na prática*. São Paulo: Thomson Learning, 2006.

HAN, J.; KAMBER, M. *Data mining: concepts and techniques*. 2th ed. San Francisco: Morgan Kaufmann, 2006.

HIRJI, K. Exploring data mining implementation. *Communications of the ACM*, v. 44, n. 7, July 2001.

INMON, W. H. *Como construir o data warehouse*. Rio de Janeiro: Campus, 1997.

MACLENNAN, J., CRIVAT, B., TANG, Z. *Data Mining with Microsoft SQL Server 2008*. Indianapolis: Wiley Publishing, 2009.

NAMEN, A. A.; SOARES, A. C. S. Mineração de dados relacionados ao aprendizado de Língua Portuguesa: um estudo exploratório. In: Encontro de Modelagem Computacional, 14., 2011, Rio de Janeiro. Encontro de Ciência e Tecnologia de Materiais, 2., 2011, Rio de Janeiro. *Anais...* Rio de Janeiro: Rede Sirius, UERJ, 2011. p. 295-304.

RAVITCH, D. *Vida e morte do grande sistema escolar americano: como os testes padronizados e o modelo de mercado ameaçam a educação*. Porto Alegre: Sulina, 2011.

RIBEIRO, V. M.; GUSMÃO, J. B. Indicadores de qualidade para a mobilização da escola. *Cadernos de Pesquisa*, São Paulo, v. 35, n. 124, p. 227-251, jan./abr. 2005.

ROSENTHAL, R.; JACOBSON, L. *Pygmalion in the classroom: teacher expectation and pupils' intellectual development*. New York: Holt, Rhinehart & Winston, 1968.

ROSENTHAL, R.; JACOBSON, L. Teachers' expectancies: determinants of pupils' IQ gains. *Psychological Report*, v. 19, p. 115-118, 1966.

TAN, P.; STEINBACH, M.; KUMAR, V. *Introdução ao data mining: mineração de dados*. Rio de Janeiro: Ciência Moderna, 2009.

TURBAN, R.; RAINER, R. K.; POTTER, R. E. *Administração de tecnologia da informação: tecnologia e prática*. Rio de Janeiro: Elsevier, 2005.

VIANNA, H. M. Avaliações nacionais em larga escala: análises e propostas. *Estudos em Avaliação Educacional*, São Paulo, n. 27, p. 41-76, jan./jun. 2003.

Anderson Amendoeira Namen, doutor em Engenharia de Sistemas de Computação pela Universidade Federal do Rio de Janeiro (UFRJ), é coordenador institucional de projeto desenvolvido para o Programa Observatório da Educação dentro do Programa de Mestrado em Psicanálise, Saúde e Sociedade da Universidade Veiga de Almeida e professor da Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, RJ, Brasil.
anamen@uva.br

Sonia Xavier de Almeida Borges, doutora em Psicologia da Educação pela Pontifícia Universidade Católica de São Paulo (PUC-SP), é professora do Mestrado Profissional em Psicanálise, Saúde e Sociedade e dos cursos

de graduação em Psicologia e pós-graduação em Teoria Psicanalítica e Prática Clínico-Institucional da Universidade Veiga de Almeida, Rio de Janeiro, RJ, Brasil.

sxaborges@gmail.com

Maria da Glória Schwab Sadala, doutora em Comunicação e Cultura pela Universidade Federal do Rio de Janeiro (UFRJ), atualmente é coordenadora, professora e supervisora do curso de mestrado profissional em Psicanálise, Saúde e Sociedade da Universidade Veiga de Almeida e professora do curso de especialização em Psicologia Clínica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ), Rio de Janeiro, RJ, Brasil.

gloriasadala@uva.br

Recebido em 27 de agosto de 2012.

Aprovado em 23 de agosto de 2013.