

Data Mart para Análise Comparativa de Dados do IDEB em Municípios da Microrregião do Pajeú em Pernambuco

Lana Yara do N. Bem¹, Valson da S. Pereira¹, Ellen Souza¹

¹Universidade Federal Rural de Pernambuco (UFRPE)
Caixa Postal 063 – 56.900-000 – Serra Talhada – PE – Brazil

{lanayarabem, valson.pereira}@gmail.com, ellen.ramos@ufrpe.br

Abstract. *The Index of Development of Basic Education (IDEB) is a measure that makes a positive relationship between performance and proficiency of elementary schools in Brazil. This evaluation method has brought a new way of analyzing data from basic education of the municipalities, since it evaluates mainly the learning of the student throughout his school life. Due to the importance of this index, this work presents a data mart for comparative analysis of IDEB data from municipalities of the Pajeu microregion, located in the semiarid state of Pernambuco, with the goal of support decision making of municipal managers and contribute to the elaboration of educational policies.*

Resumo. *O Índice de Desenvolvimento da Educação Básica (IDEB) é um índice que faz um relacionamento de forma positiva sobre informações de rendimento escolar e proficiência de escolas do ensino básico do Brasil. Sua maneira de avaliação trouxe uma nova forma de analisar os dados da educação básica dos municípios, pois avalia principalmente o aprendizado do aluno ao longo de sua vida escolar. Devido a essa importância do IDEB, este trabalho tem como objetivo apresentar um data mart para análise comparativa com base nos dados do IDEB e suas metas, entre municípios da microrregião do Pajeú, situada no semiárido pernambucano, com o propósito de apoiar a tomada de decisão de gestores municipais e contribuir para elaboração de políticas educacionais.*

1. Introdução

A educação básica no Brasil vem sofrendo profundas mudanças desde a promulgação da Constituição Federal de 1988, que possibilitou a criação da Lei de Diretrizes e Bases da Educação Nacional. Nesta lei, é declarada de maneira explícita que o poder público municipal e estadual são responsáveis pela educação básica (Cury et al., 2002).

Para tanto, no ano de 2007, o governo federal criou o Índice de Desenvolvimento da Educação Básica (IDEB), que segundo Fernandes (2007a), tem a finalidade de detectar escolas e redes de ensino que apresentam um baixo desempenho, serve também para o acompanhamento da evolução temporal dessas instituições, por meio de metas estabelecidas.

Diante deste contexto, através da utilização de um *data mart* para armazenamento dos dados históricos, é possível ter uma visão dimensional do IDEB e suas metas, que por sua vez, simplifica a tarefa de gerência e análise desses dados, contribuindo para a detecção de desigualdade no desempenho escolar entre os municípios e auxiliando na

tomada de decisão, para que se possa adotar melhores políticas públicas voltadas para a educação.

Dessa maneira, o objetivo deste trabalho é apresentar um *data mart* para uma análise comparativa com base nos dados do IDEB e suas metas, desde o ano de 2007 até 2015, entre municípios da microrregião do Pajeú, situada no semiárido pernambucano, com o propósito de apoiar a tomada de decisão de gestores desses municípios e contribuir para elaboração de políticas educacionais.

Este artigo está organizado da seguinte maneira: na seção 2 é apresentado o referencial teórico sobre *Data Warehouse/Data mart*, IDEB e trabalhos relacionados, na seção 3 está descrito o método de construção do *data mart*, na seção 4 são apresentados os resultados e a seção 5 descreve a conclusão juntamente com os trabalhos futuros.

2. Referencial Teórico

2.1. *Data Warehouse e Data mart*

Data Warehouse (DW) são bancos de dados históricos oriundos de sistemas transacionais das organizações. Segundo Barbieri (2001), os DW's mantêm uma estrutura de dados dimensionais que é destinado ao apoio da tomada de decisão de gestores, possibilitando o processamento analítico por meio de ferramentas específicas. Devido ao custo elevado na construção de um DW, ele pode ser desenvolvido em partes menores, denominada de *data mart* (DM) ou cubo (Ferreira et al., 2010).

A criação de um DW se inicia a partir da extração, transformação e carga (fase também conhecida como ETL ou ETC) dos dados que estão armazenados em bancos de dados transacionais. De acordo com Kanashiro (2007), nesta etapa é realizada a extração de uma ou mais fontes de dados que irão compor o DW/DM, logo após é feita padronização nas inconsistências e variações detectadas nesses dados, por meio da transformação, e por fim os dados são carregados em um banco de dados dimensional.

O modelo de dados dimensional mais utilizado na construção de um DW/DM é o modelo estrela (*star schema*) que segundo Kimball e Ross (2011), é um modelo que pode ser implementado em sistemas de banco de dados relacionais e não necessita de normalização dos dados, o que permite um melhor desempenho com relação às consultas. O modelo estrela é formado por uma tabela central denominada de tabela fato, na qual estão as métricas a serem analisadas, e as tabelas de dimensão que são tabelas que complementam a fato e possuem atributos de negócio (Júnior et al., 2015). Na criação de um DM/DW podem ser utilizadas abordagens de construção que possuem um maior aspecto de integração, denominada de abordagem monolítica, ou uma abordagem que prioriza a disponibilização mais rápida do DM/DW que consiste em uma abordagem evolutiva (Barbieri, 2001).

2.1.1. Aplicações *Online Analytical Processing* (OLAP)

Após a fase de ETL com os dados já carregados no modelo dimensional do DW/DM, os gestores podem utilizar aplicações *Online Analytical Processing* (OLAP), que de acordo com Colaço Jr. (2004), tem como objetivo fornecer um conjunto de técnicas e ferramentas

para possibilitar o tratamento de informações contidas em um grande volume de dados, como por exemplo, um DM/DW.

Segundo Fortulan e Gonçalves Filho (2005), aplicações OLAP devem apresentar as seguintes características:

- devem possibilitar uma visão multidimensional dos dados;
- suportar a realização de operações matemática complexas;
- criar agregados de dimensões;
- prever e analisar tendências;
- implementar cenários a partir de hipóteses;
- realizar cálculos e manipulação de dados entre dimensões;

Esse tipo de aplicação apoia gestores em análises comparativas e tomada de decisão, pois pode fornecer dentre outras funcionalidades: relatórios, *dashboards*, consultas pré definidas e *ad-hoc*, de acordo com a necessidade de consulta do gestor.

Para Barbieri (2001), existem basicamente três tipos de arquitetura para aplicações OLAP:

- **cliente gordo:** a aplicação OLAP é armazenada na estação de trabalho do cliente;
- **cliente magro:** utiliza a arquitetura de cliente/servidor, na qual a aplicação OLAP permanece hospedada em um servidor local;
- **cliente servidor/web:** utiliza a mesma abordagem do cliente magro, porém com a utilização de um navegador pela parte do cliente.

2.2. Índice de Desenvolvimento da Educação Básica (IDEB)

O IDEB é um índice que faz um relacionamento de forma positiva sobre informações de rendimento escolar e proficiência de escolas do ensino básico, por meio de avaliações como a Prova Brasil e o SAEB (Fernandes, 2007b).

A forma de avaliação do IDEB trouxe uma nova forma de analisar os dados da educação básica dos municípios, pois avalia principalmente o aprendizado do aluno ao longo de sua vida escolar e não só processos de ensino e gerenciais como feito anteriormente. Antes do IDEB, era adotado como medida de contingência apenas a expansão de aspectos, tais como, mais horas aula, mais disciplinas (Soares, 2011).

O valor do IDEB é calculado em uma escala de 0 a 10, que segundo Da Paz (2010), é uma “combinação de indicadores de fluxo (aprovação, repetência e evasão) e de proficiência (Prova Brasil e Saeb)”. O cálculo consiste no produto entre a média da proficiência dos alunos da escola pela taxa média de aprovação (Fernandes e Gremaud, 2009). Sua avaliação ocorre a cada dois anos em escolas de todos os municípios brasileiros.

2.3. Trabalhos Relacionados

A abordagem de Ferreira (2015), realiza uma investigação acerca de fatores que podem influenciar na conclusão do ensino fundamental. Por meio de técnicas de mineração de dados, os microdados públicos do Censo Escolar da Educação Básica do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) são analisados. A análise busca estudar características individuais, familiares e locais, que contribuam para que jovens concluam ou não o ensino fundamental, a partir da hipótese que as características

dos alunos, bem como seu ambiente escolar influenciam fortemente na conclusão desta etapa de ensino.

A proposta de Santos et al. (2014), utilizou dados do Censo Escolar de 2011 obtidos no site do INEP para realizar uma mineração das regiões brasileiras na qual as escolas estão localizadas, para que a partir disso, seja disponibilizada uma visão geral da situação da infraestrutura desses ambientes escolares. O trabalho também fez uma análise dos atributos que formam a base de dados do Censo Escolar para descobrir associações que podem causar impactos na infraestrutura das escolas. A proposta demonstra resultados importantes, como por exemplo que em 90% das escolas do Norte e Nordeste não possuem laboratório de informática, biblioteca e sala de leitura.

Silva et al. (2014), apresenta uma análise de dados sobre a relação entre fatores socioeconômicos e o desempenho dos estudantes do ensino médio na prova do Exame Nacional do Ensino Médio (ENEM). Os dados utilizados foram do desempenho das provas e do questionário socioeconômico do ENEM de 2010 extraídos do INEP. Para buscar a relação dos dados foi realizada uma mineração de dados por meio da tarefa de associação, que dentre outros resultados chegou-se a conclusão que, geralmente alunos de escola pública possuem um desempenho regular perante o exame.

Vidal e Vieira (2011), buscam em seu trabalho observar os comportamentos no IDEB das séries iniciais e finais do ensino fundamental de dez municípios cearenses. Essa observação consiste em investigar questões relativas às gestões de políticas educacionais municipais e quais iniciativas foram tomadas para a melhoria do desempenho desses municípios com relação ao IDEB. O trabalho traça um comparativo entre os municípios analisados com relação a evolução de cada um ao longo dos anos. Como resultado do trabalho são destacados alguns pontos importantes da política educacional a serem melhorados de cada um dos dez municípios. Apesar de ser um trabalho com resultados relevantes, não foi disponibilizado nenhum sistema de informação para apoiar a tomada de decisão.

O trabalho apresentado por Ramos et al. (2015), consiste em um DW para analisar alguns indicadores do IDEB através de clusterização. Na construção do DW foram utilizados dados das escolas que participaram do IDEB entre os anos de 2007 a 2013. A clusterização foi usada para criar grupos de escolas com características similares, como por exemplo por tipo de escola (municipal, estadual, federal), para que fosse possível a análise de dados proposto pelos autores. As análises realizadas se denotam em desempenho no IDEB por tipo de escola e comparativo de fatores externos que podem afetar o desempenho das escolas, como por exemplo, a distribuição de recursos públicos destinados para a educação de cada estado.

Diferentemente dos trabalhos apresentados, a proposta deste artigo tem uma visão mais específica, tanto geograficamente como também analiticamente, pois serão considerados como métricas do *data mart* apenas o valor do IDEB e sua meta, sendo assim, foi desconsiderada a análise de fatores externos que possam ter influência nesses dados. Além disso, foram utilizadas tecnologias distintas dos artigos expostos nesta subseção, como por exemplo, o *Pentaho Data Integration*, *Schema Workbench* e o *Pentaho Business Analytics Platform*, que estão descritos na seção a seguir.

3. Método

Para a implementação do *data mart* foram utilizadas as seguintes tecnologias:

- **Pentaho Data Integration 7.0 (PDI)**: utilizado para a fase de ETL do DM;
- **SQL Power Architect**: ferramenta para fazer a modelagem dimensional em um banco de dados relacional;
- **MySQL**: Sistema de banco de dados relacional que foi adaptado para o modelo dimensional para o armazenamento do DM;
- **Schema Workbench**: ferramenta do *Pentaho* para criação e publicação do cubo OLAP no *BI Server*;
- **Pentaho Business Analytics Platform** ou **Pentaho BI Server**: Plataforma OLAP usada para as consultas dos gestores ao *data mart*.

Os dados do IDEB e suas metas inseridos no *data mart* foram coletados do portal do IDEB/INEP¹. Esta coleta consistiu na extração de dados entre os anos de 2007 a 2015 de municípios da microrregião do Pajeú, situada no semiárido do Pernambuco, que são os seguintes: Afogados da Ingazeira, Brejinho, Calumbi, Carnaíba, Flores, Igaraci, Ingazeira, Itapetim, Quixaba, Santa Cruz da Baixa Verde, Santa Terezinha, São José do Egito, Serra Talhada, Solidão, Tabira, Triunfo e Tuparetama.

A partir dos dados extraídos do portal do IDEB/INEP, os dados foram estruturados no formato JSON para serem carregados no PDI, após isso foi realizada a etapa de extração, transformação e carga dos dados no banco de dados MySQL. Por fim, a aplicação foi disponibilizada para os usuários realizarem suas consultas OLAP através do *Pentaho BI Server*, conforme apresenta a Figura 1, que por sua vez, possui uma arquitetura OLAP cliente servidor/web. Além disso, a aplicação seguiu a abordagem evolutiva para seu desenvolvimento e utilizou o modelo dimensional estrela, como ilustra a Figura 2.

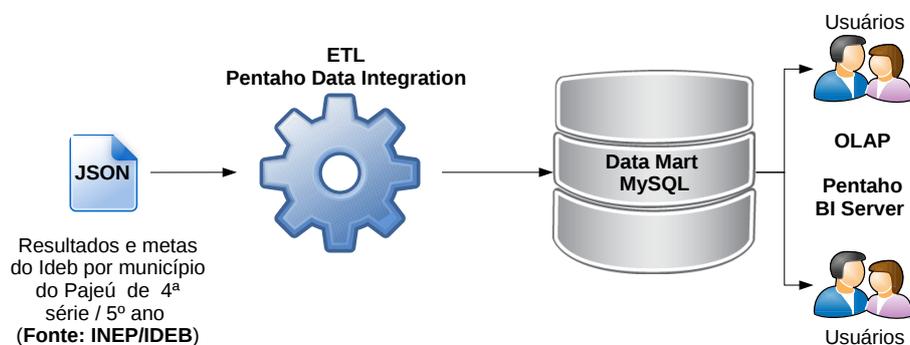


Figura 1. Arquitetura da aplicação.

¹Portal IDEB/INEP: ideb.inep.gov.br

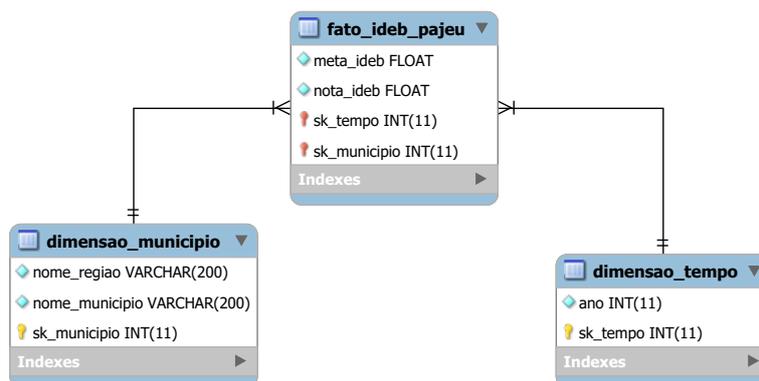


Figura 2. Modelo Dimensional da aplicação.

3.1. Etapa de extração, transformação e carga (ETL)

Os dados do IDEB bem como suas metas foram extraídos do site IDEB/INEP e estruturados em um arquivo JSON como já descrito anteriormente. A partir disso, os municípios foram carregados utilizando o PDI, através do *Step JSON Input*, para serem armazenados na dimensão município por meio do *Step Combination Lookup/Update*, conforme mostra a Figura 3 com o plano de carga.

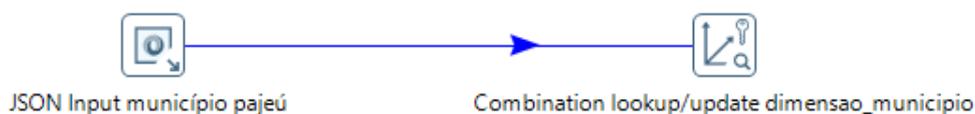


Figura 3. Plano de Carga da Dimensão município.

Após ter carregado os municípios na tabela da dimensão município, foi carregada a dimensão tempo com os anos de 2007 a 2015, também utilizando o *JSON Input* e o *Step Combination Lookup/Update*, como descrito na Figura 4.

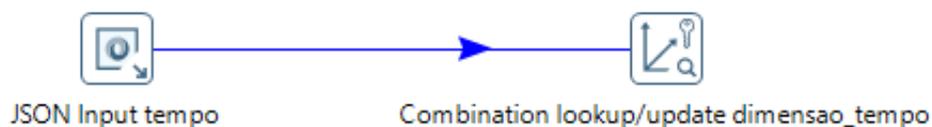


Figura 4. Plano de Carga da Dimensão Tempo.

Depois de carregar a dimensão tempo, foi carregada a tabela de fato. Nela estão as chaves estrangeiras das dimensões tempo e município, além das seguintes métricas: *meta_ideb* e *nota_ideb*. Já com as chaves das dimensões tempo e município recuperadas, os dados da tabela de fato foram carregados, por meio do *Step Table Output*, como ilustra a Figura 5.

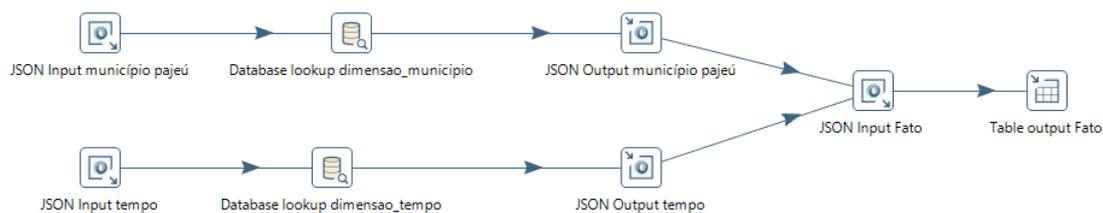


Figura 5. Plano de carga da tabela fato.

3.2. Etapa de criação do cubo

Para criar o cubo OLAP, foi utilizada a ferramenta do Pentaho *Schema Workbench*, que possibilitou o DM ser reconhecido e utilizado pelo *BI Server*, como ilustra a Figura 6.

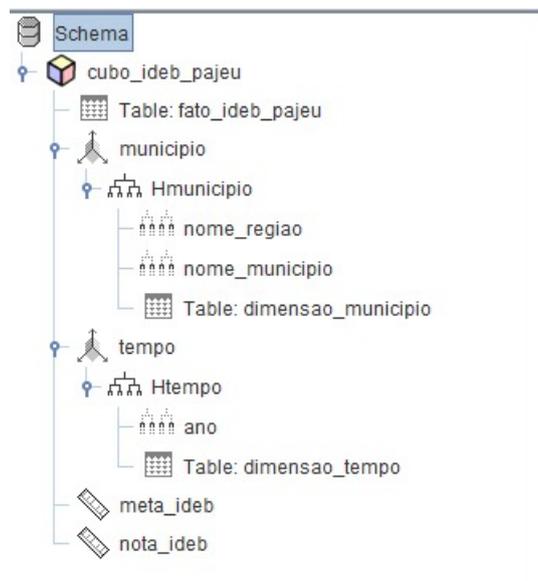


Figura 6. Cubo OLAP criado no *Schema Workbench*

4. Resultado

Após criar e publicar o cubo, por meio da ferramenta *Schema Workbench* e configurar o *BI Server*, foram realizadas consultas demonstrativas com o auxílio da ferramenta OLAP *Pivot4J View (from jPivot Reports)*, que podem contribuir na tomada de decisão dos gestores municipais.

A Figura 7 mostra o gráfico com a média das notas do IDEB referente aos 17 municípios. A Figura 8 descreve um *ranking* por média do IDEB. Pelas figuras, é possível observar, por exemplo, quais cidades estão acima da média geral da Microrregião do Pajeú (4.486), são elas: Carnaíba (5.24), Quixaba (5.24), Tuparetama (5.06), Triunfo (4.86) e São José do Egito (4.74).



Figura 7. Média do IDEB nos municípios entre os anos de 2007 a 2015.

Hmunicipio		Htempo		Measures
(All)	nome_regiao	nome_municipio	(All)	nota_ideb
All municipio.Hmunicipios			All tempo.Htempo	4.486
All municipio.Hmunicipios	Microrregião do Pajeú	Carnaíba	All tempo.Htempo	5.24
		Quixaba	All tempo.Htempo	5.24
		Tuparetama	All tempo.Htempo	5.08
		Triunfo	All tempo.Htempo	4.86
		São José do Egito	All tempo.Htempo	4.74
		Santa Cruz da Baixa Verde	All tempo.Htempo	4.46
		Brejinho	All tempo.Htempo	4.46
		Ingazeira	All tempo.Htempo	4.42
		Afogados da Ingazeira	All tempo.Htempo	4.38
		Solidão	All tempo.Htempo	4.38
		Iguaraci	All tempo.Htempo	4.38
		Tabira	All tempo.Htempo	4.32
		Flores	All tempo.Htempo	4.24
		Itapetim	All tempo.Htempo	4.22
		Santa Terezinha	All tempo.Htempo	4.08
Calumbi	All tempo.Htempo	3.96		
Serra Talhada	All tempo.Htempo	3.82		

Figura 8. Ranking das médias do IDEB entre 2007 e 2015.

A Figura 9 mostra a evolução do IDEB no município de Serra Talhada ao longo dos anos de 2007 a 2015. Nesse caso, é possível observar com mais detalhes pela Figura 10, que os valores do IDEB em Serra Talhada aumentaram consideravelmente num intervalo de 08 anos. O município analisado atingiu um crescimento de 1.6 em sua nota, partindo de 3.1 em 2007 para 4.7 em 2015.

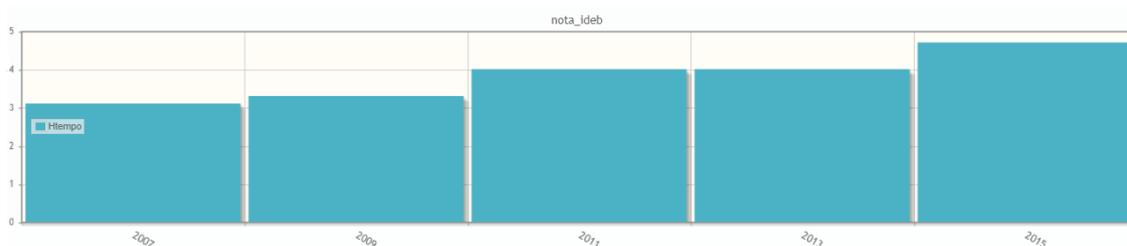


Figura 9. Evolução do IDEB no município de Serra Talhada ao longo dos anos.

		Measures
Hmunicipio	Htempo	↔ nota_ideb
Serra Talhada	2007	3.1
	2009	3.3
	2011	4
	2013	4
	2015	4.7

Figura 10. Notas do IDEB do município de Serra Talhada.

5. Conclusão e Trabalhos Futuros

Diante do trabalho apresentado percebe-se a importância da nota IDEB, não só para a microrregião do Pajeú, mas também para toda a educação básica do Brasil, pois por meio deste indicador, os gestores podem definir políticas educacionais que melhorem a qualidade do ensino básico municipal.

Dentre as diversas formas que os dados do IDEB podem ser analisados pelos gestores, para que estes possam ter um apoio na tomada de decisão, consiste na utilização de um *data mart*, que através de ferramentas OLAP, pode-se facilmente realizar consultas, além de gerar gráficos e relatórios.

Por meio da solução exposta neste artigo, os gestores poderão efetuar consultas e obter relatórios comparativos entre municípios específicos, como também, podem ter uma visão geral da microrregião e perceber se determinado município está com seu IDEB acima ou abaixo da média microrregional.

Como trabalhos futuros pretende-se agregar as demais microrregiões do Pernambuco ao *data mart*, além disso, almeja-se adicionar outros indicadores importantes, como por exemplo, o Índice de Desenvolvimento Humano (IDHM) de cada município pernambucano e por fim, disponibilizar a aplicação em plataforma web para consulta pública.

Referências

- Barbieri, C. (2001). *BI–Business Intelligence– Modelagem & Tecnologia*. Axel Books do Brasil Editora, Rio de Janeiro.
- Colaço Jr., M. (2004). *Projetando sistemas de apoio à decisão baseados em data warehouse*. Axel Books do Brasil Editora, Rio de Janeiro.
- Cury, C. R. J. et al. (2002). A educação básica no Brasil. *Educação e Sociedade*, 23(80):168–200.
- Da Paz, F. M. (2010). O ideb e a qualidade da educação no ensino fundamental: fundamentos, problemas e primeiras análises comparativas. *ETIC-Encontro de Iniciação Científica*, 5(5).

- Fernandes, R. (2007a). *Índice de desenvolvimento da educação básica (IDEB)*. MEC–Ministério da Educação, INEP–Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília.
- Fernandes, R. (2007b). Índice de desenvolvimento da educação básica (ideb): metas intermediárias para a sua trajetória no brasil, estados, municípios e escolas. *INEP/MEC*.
- Fernandes, R. and Gremaud, A. P. (2009). Qualidade da educação: avaliação, indicadores e metas. *Educação básica no Brasil: construindo o país do futuro*. Rio de Janeiro: Elsevier, 1:213–238.
- Ferreira, G. (2015). Investigação acerca dos fatores determinantes para a conclusão do ensino fundamental utilizando mineração de dados educacionais no censo escolar da educação básica do inep 2014. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 4, page 1034.
- Ferreira, J., Miranda, M., Abelha, A., and Machado, J. (2010). O processo etl em sistemas data warehouse. In *INForum*, pages 757–765.
- Fortulan, M. R. and Gonçalves Filho, E. V. (2005). Uma proposta de aplicação de business intelligence no chão-de-fábrica. *Gestão & Produção*, 12(1):55–66.
- Júnior, J. O., Bastos, L., and Kaestner, C. (2015). Uma abordagem de data warehouse educacional para apoio à tomada de decisão. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 4, page 1064.
- Kanashiro, A. (2007). *Um data warehouse de publicações científicas: indexação automática da dimensão tópicos de pesquisa dos data marts*. PhD thesis, Universidade de São Paulo.
- Kimball, R. and Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- Ramos, T. G., Machado, J. C. F., and Cordeiro, B. P. V. (2015). Primary education evaluation in brazil using big data and cluster analysis. In *Procedia Computer Science*, volume 55, pages 1031–1039. Elsevier.
- Santos, M. S., Claro, D. B., de Souza Oliveira Filho, A. M., Paoli, A. R., Barbosa, C. D., Almeida, E. V., de Araújo Carvalho, G. D., de Jesus Santana, H., Bulcão, L. A., Marotta, L. L., et al. (2014). Análise das infraestruturas do censo escolar 2011: uma proposta da disciplina de tópicos em banco de dados. In *III Congresso Brasileiro de Informática na Educação (CBIE) – XXV Simpósio Brasileiro de Informática na Educação (SBIE)*, volume 25, page 124.
- Silva, L. A., Morino, A. H., and Sato, T. M. C. (2014). Prática de mineração de dados no exame nacional do ensino médio. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação (WCBIE)*, volume 3, page 651.
- Soares, J. F. (2011). Análise dos pressupostos educacionais e estatísticos do índice de desenvolvimento da educação básica (ideb). *Encontro de Pesquisa em Educação da Região Sudeste*, 10.
- Vidal, E. M. and Vieira, S. L. (2011). Gestão educacional e resultados no ideb: um estudo de caso em dez municípios cearenses. *Estudos em Avaliação Educacional–Fundação Carlos Chagas*, 22(50):419–434.