

## **SOFTWARES DE MINERAÇÃO DE TEXTO NA ANÁLISE DE PRODUÇÕES TEXTUAIS DE ESTUDANTES DO ENSINO FUNDAMENTAL: POSSIBILIDADES INTERDISCIPLINARES**

Luciana Gonçalves de Oliveira  
Breno Fabrício Terra Azevedo  
Cleidiane Basílio Almeida Gomes

### **Resumo**

Algumas ferramentas de mineração de texto podem auxiliar na análise de produções textuais, oportunizando ações pedagógicas interdisciplinares, de forma a relacionar conteúdos de diversas áreas do conhecimento. Ao mesmo tempo, viabilizam a contagem de termos e parágrafos, sendo possível verificar quais palavras mais se repetem num texto. O objetivo geral deste trabalho consiste em analisar o potencial das ferramentas de mineração de texto como uma tecnologia capaz de auxiliar docentes na análise das produções textuais dos estudantes. Foram utilizadas duas ferramentas de mineração de texto, Sobek e WordCounter. Os resultados foram analisados de forma qualitativa, observando-se as potencialidades de cada software, chegando-se a conclusão que ambos auxiliam o docente na análise e correção dos textos elaborados pelos estudantes, viabilizando um trabalho didático dentro do conceito de interdisciplinaridade.

**Palavras-chave:** Produções Textuais. Mineração de Texto. Tecnologias Digitais. Interdisciplinaridade.

### **1 Introdução**

O professor deve incentivar a produção textual estimulando os alunos a exteriorizar ideias, fazer autocrítica a fatos que ocorrem no seu entorno e levantar pontos que necessitam de esclarecimento adicional. Segundo Neves et. al. (2015, p. 219), “o contexto histórico-cultural em que a linguagem escrita está inserida entrelaça-se com os sentidos dos textos produzidos, bem como com as relações que os sujeitos estabelecem com esses textos e entre si”.

O presente trabalho surgiu a partir da indagação: “As ferramentas de mineração de textos podem auxiliar na análise das produções textuais de alunos do Ensino Fundamental?” Visando obter resposta para esse questionamento utilizou-se as ferramentas Sobek e *WordCounter* para analisar a produção textual dos alunos do 4.º ano do ensino fundamental de uma escola municipal.

A mineração de texto é uma área de conhecimento que permite ao usuário analisar uma série de documentos por meio de ferramentas que extraem dos mesmos informações úteis. Mineração de texto e mineração de dados possuem muitas semelhanças no que tange a arquitetura de sistemas (FELDMAN;SANGER, 2007).

Este artigo tem como objetivo geral analisar o potencial das ferramentas de mineração de texto como uma tecnologia auxiliar nas produções textuais dos estudantes. Os objetivos específicos consistem em: analisar se os alunos estão produzindo textos de acordo com o tema proposto; identificar quais as palavras chave que mais se repetem nas produções textuais dos alunos; analisar a possibilidade de conexão de uma palavra com outras palavras que se repetem.

## **2 Referencial Teórico**

Nessa seção são apresentados os principais conceitos que nortearam a pesquisa. Com isso, descreve-se abaixo a fundamentação teórica sobre texto e linguagem, utilização de tecnologias educacionais e mineração de textos.

### **2.1 Texto e Linguagem**

Segundo Bakhtin (2003), o texto é uma construção de signos desenvolvida pelo sujeito para expressar suas ideias e opiniões, como resultado de suas interações com outras pessoas dentro de uma cultura. Desta forma, compreende-se o sujeito que escreve como um ser social, com o anseio de ter suas expressões compreendidas por outros sujeitos.

Sobre a linguagem, Rizzo (1986), afirma que é a necessidade de comunicação entre os indivíduos, de forma que a linguagem escrita reflete, de certa forma, a linguagem oral. É importante desenvolver as habilidades relacionadas à capacidade de produzir textos, tendo a escrita como instrumento de transmissão no processo de dialogicidade entre as pessoas.

Para Koch (1998) é preciso pensar a linguagem como lugar de interação entre os homens, onde constroem-se identidades e há a constituição de papéis, encarando-a não apenas como representação do mundo e do pensamento, mas também como forma de interação social. Em função disto, faz-se necessário o desenvolvimento das competências relativas à escrita desde as séries iniciais, contextualizando o trabalho pedagógico e considerando os aspectos do avanço tecnológico que influenciam nas formas de comunicação e interação social.

### **2.2 Utilização de Tecnologias Educacionais**

Johnson (2001) afirma que o uso do computador muda a maneira de escrever e que a forma como se processa o pensamento passa a acontecer à medida que a ação de escrever vai se desenrolando. Num espaço em que as tecnologias estão cada vez mais presentes no dia a dia torna-se clara a compreensão de que “[...] um dos desafios à educação o repensar sobre novas propostas educativas que superem a instrução ditada pelo livro didático, centrada no dizer do professor e na passividade do aluno” (VALENTE;ALMEIDA;GERALDINE, 2017, p.458).

Cabe ressaltar que não pode-se excluir ou minimizar o papel do professor diante da adoção destas, ou de quaisquer outras ferramentas. O docente é o mediador da relação entre o aluno e as Tecnologias Digitais de Informação e Comunicação no processo de construção de conhecimento (ALVES et al, 2017).

A utilização de ferramentas de mineração de texto pode auxiliar o docente, uma vez que a partir das mesmas é possível extrair informações de um texto, como os termos mais citados, verificação das conexões de termos e observar se o texto está fluído (KLEMMANN et. al, 2009).

## **2.3 Mineração de Textos**

Para Aranha e Passos (2006), a mineração de textos consiste em extrair dados textuais de grandes volumes de documentos com objetivos específicos. Trata-se de ferramentas que obtém dados dos textos, facilitando a organização de informações para um determinado fim.

Optou-se por utilizar as ferramentas de mineração de texto Sobek e *WordCounter* devido a alguns recursos oferecidos pelas ferramentas serem úteis ao trabalho e possibilitar uma análise diferenciada do mesmo texto. Ambas ferramentas possuem funcionamento *online*, são gratuitas e oferecem funcionalidades como contagem de termos e apresentação de palavras-chave.

### **2.3.1 Sobek**

A ferramenta Sobek foi desenvolvida no Programa de Pós-Graduação em Informática na Educação na Universidade Federal do Rio Grande do Sul, para auxiliar os professores no acompanhamento da escrita colaborativa dos alunos, com a capacidade de extrair termos frequentes nos textos e identificar a relação

existentes entre termos, como expõe Macedo et al. (2009).

Segundo Kleimann et al. (2009), o Sobek também foi adaptado para utilização com os próprios alunos com a finalidade de auxiliá-los no processo de produção textual.

Disponível por meio do link: [<http://sobek.ufrgs.br/>](http://sobek.ufrgs.br/), apresenta os seguintes recursos:

- Editar, acrescentar ou remover nodo (conceitos);
- Salvar grafo (permite trabalhar com este posteriormente);
- Salvar imagem do grafo;
- Identificar a relação entre os conceitos;
- Opções de idiomas: português/ inglês;
- *Stopwords*;
- Selecionar a frequência mínima de conceitos;
- Selecionar o número médio em que devem aparecer os conceitos.

Por meio do link citado anteriormente é possível fazer o download da ferramenta ou utilizar o Sobek *online*.

### **2.3.2 WordCounter**

Trata-se de uma ferramenta *online* e gratuita para contar palavras, frases, parágrafos e sinais em qualquer idioma. Também é possível verificar a densidade de uma palavra, ou seja, quantas vezes ela se repete no texto. De acordo com Klemann; Reategui; Rapkiewicz (2011), a ferramenta é muito útil para autores de textos, pois mostra as palavras redundantes numa lista permitindo uma escrita mais eficiente.

A ferramenta está disponível por meio do *link* [<http://pt.wordcounter360.com/>](http://pt.wordcounter360.com/). É possível inserir um texto inteiro ou trechos no campo próprio para contagem. Automaticamente, o contador emite o resultado da quantidade de palavras, total de caracteres, caracteres sem espaço, frases, parágrafos e média de palavras / frase.

Simultaneamente, em outro campo denominado “densidade de palavras-chave”, são mostradas as palavras que se repetem em ordem decrescente de acordo com a quantidade de vezes que a palavra consta no texto e qual o percentual

de participação em relação a todas as palavras do texto.

### 3 Metodologia

Foi proposta uma atividade numa turma do 4º ano do Ensino Fundamental, com 15 alunos, de uma escola pública do município de Campos dos Goytacazes no Estado do Rio de Janeiro. A atividade consistia, inicialmente, no que foi denominada “dinâmica do abraço”, onde os alunos, após lerem um texto sobre o referido tema, abraçaram-se. Em seguida, foi proposto que a turma escrevesse como foi essa experiência e o que sentiram ao abraçar.

Para a análise dos textos produzidos pelos alunos, orientou-se à turma quanto à utilização da ferramenta Sobek, solicitando que os discentes seguissem os passos seguintes:

- Localização do aplicativo no computador. A fim de auxiliar na localização, o professor realizou a ação projetando-a na lousa.
- Digitação da produção textual no Sobek. A figura 1 ilustra a ação inicial.
- Seleção da opção “Extrair conceitos”. Esta opção permite criar uma base de conceitos através do processo de mineração de textos.

A figura 2 apresenta o grafo gerado após a extração de conceitos.

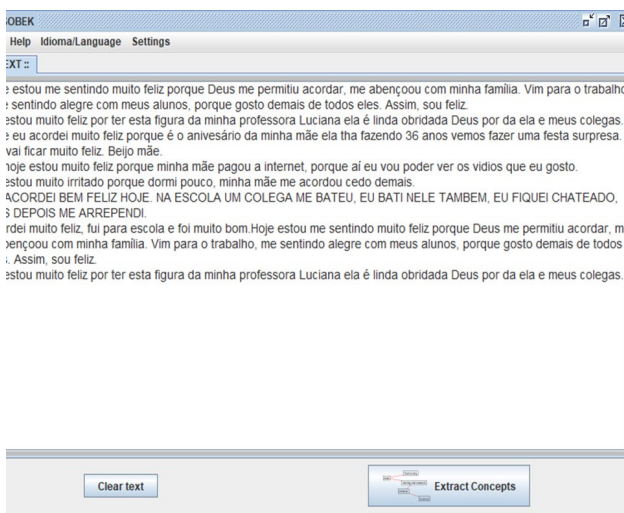


Figura 1- Digitação do texto no Sobek  
Fonte: <http://sobek.ufrgs.br/>

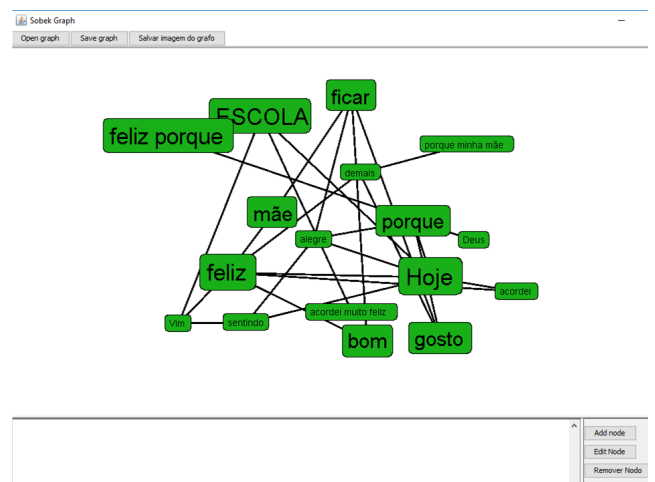


Figura 2- Grafo gerado para o texto analisado  
Fonte: <http://sobek.ufrgs.br/>

A atividade solicitada também propôs, após a produção textual dos alunos em um mesmo arquivo de texto no computador (figura 3), uma análise sobre quais

sentimentos foram mais citados utilizando-se o grafo gerado pelo Sobek (figura 4). Os discentes puderam perceber, inclusive, a relação entre as palavras usando o recurso disponibilizado pela ferramenta que consiste em clicar na palavra para ser demonstrado graficamente a relação entre os conceitos ou termos.

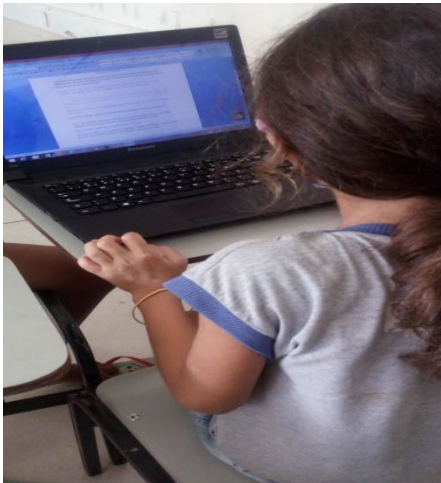


Figura 3- Escrita em um único arquivo de texto  
Fonte: Autoria própria

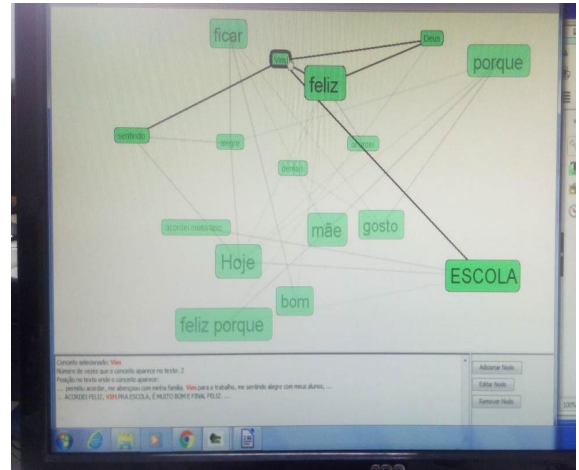


Figura 4- Grafo gerado pelo Sobek  
Fonte: <http://sobek.ufrgs.br/>

Como parte da segunda etapa da atividade, foi proposta à turma a construção de um gráfico com os sentimentos descritos, ao qual foi chamado de Emocionômetro (figura 5).

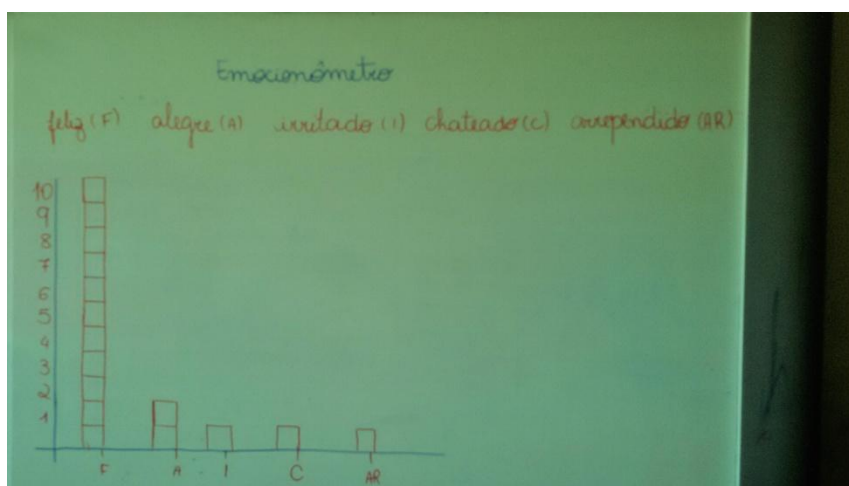


Figura 5- Emocionômetro.  
Fonte: Autoria Própria

A partir da exploração do gráfico (figura 6), foi possível trabalhar a leitura e

interpretação de gráficos, como propõe os Parâmetros Curriculares Nacionais: [...]” fazer com que o aluno venha a construir procedimentos para coletar, organizar, comunicar e interpretar dados, utilizando tabelas, gráficos e representações que aparecem frequentemente em seu dia-a-dia.” (BRASIL 1997, p. 40).

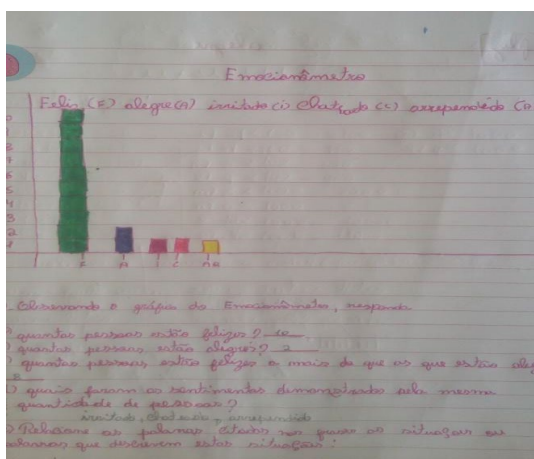


Figura 6- Exploração do gráfico.  
Fonte: Autoria própria

De forma interdisciplinar, as atividades entrelaçaram conteúdos da Língua Portuguesa e da Matemática, com a utilização da ferramenta de mineração de textos como auxiliar ao processo de ensino e aprendizagem, de maneira a garantir o que afirma Fazenda (2008) “[...] interdisciplinaridade sugere um conjunto de relações entre disciplinas abertas sempre a novas relações que se vai descobrindo.” (FAZENDA, 2008, p.18)

Ao mesmo tempo, utilizou-se a mesma produção textual dos alunos e realizou-se a análise do texto com adoção da ferramenta *WordCounter*. O trabalho ocorreu da seguinte forma:

- Acessou-se a ferramenta através do link <<http://pt.wordcounter360.com/>>.
- Copiou-se o texto produzido pelos discentes, o qual foi inserido no campo próprio da ferramenta para que ocorresse a contagem de palavras (figura 7).

#### Palavras e caracteres Contador

Bem-vindo ao balcão WordCounter360 ° palavra também chamados de contador / personagem letras contador / temporizador / contador de sinais. Você vai encontrar nesta página web uma ferramenta online gratuita e fácil de calcular o número de letras, sinais, caracteres, palavras, parágrafos e frases em um texto em qualquer idioma.

**211** Palavras **1155** Personagens Excluir

Eu estou muito irritado porque dormi pouco, minha mãe me acordou cedo demais.

EU ACORDEI BEM FELIZ HOJE. NA ESCOLA UM COLEGA ME BATEU, EU BATI NELE TAMBEM, EU FIQUEI CHATEADO, MAS DEPOIS ME ARREPENDEI.

Acordei muito feliz, fui para escola e foi muito bom.

Estou feliz, porque briquei com meu irmão de boneco.Depois comi pão doce e Nescäu eme arrumei para ir pra escola.

ACORDEI FELIZ, VIM PRA ESCOLA, É MUITO BOM E FINAL FELIZ.

É bom ficar alegre, estou alegre hoje porque gosto de estudar, que é melhor que ficar em casa.

frases, parágrafos e a média de palavras/frases. Também foi possível verificar a densidade de palavras-chave, ou seja, os vocábulos mais presentes no texto (figura 8).



Figura 7 - Resultados da análise do texto na ferramenta WordCounter.  
Fonte: <<http://pt.wordcounter360.com/>>

Dessa forma torna-se possível, de acordo com a densidade das palavras chave, compreender se o assunto abordado pelos alunos no texto está contextualizado com o tema proposto pela professora.

#### 4 Resultados e Discussão

A ferramenta *WordCounter* permitiu observar que as três palavras que mais aparecem no texto são “feliz” (11 vezes), “muito” (9 vezes) e “porque” (9 vezes). Em seguida aparecem as palavras “estou” (7 vezes), “minha” (6 vezes), “hoje” (5 vezes) e “acordei” (4 vezes).

As três palavras mais frequentes mostram que a produção dos alunos estava contextualizada com o tema proposto pela professora. Entende-se que os estudantes compreenderam a proposta da escrita sobre a dinâmica do abraço.

A densidade de palavras apresentada no *WordCounter* foi diferente da apresentada no Sobek. Apenas a palavra “feliz” se repetiu na análise de densidade de ambas ferramentas. No Sobek as palavras-chave que mais apareceram em ordem decrescente foram “feliz”, “alegre”, “irritado”, “chateado” e “arrependido”.

A ferramenta de mineração de textos Sobek possibilitou a visualização das palavras dentro das frases em que aparecem. Além de indicar quantas vezes cada conceito se repete, é possível identificar em que contexto, uma vez que há o apontamento das frases, conforme a palavra selecionada.



O Sobek viabilizou a análise da relação entre os termos, ampliando as estratégias de trabalho didático, expandindo as ações pedagógicas, inclusive, para o trabalho de outros conteúdos além da análise da escrita dos alunos. Percebeu-se, por seu aspecto visual semelhante a um mapa conceitual, maior disponibilidade dos alunos e interesse, fazendo com que os discentes realizassem uma discussão oral a respeito da questão afetiva do abraço e sua relação com outros fatores, conforme foi mostrado no grafo.

## **5 Considerações Finais**

O *WordCounter* mostrou-se uma importante estratégia para o uso docente como recurso que auxilia na avaliação e análise das produções escritas dos alunos. Inclusive, no momento da produção, é possível que o docente utilize a ferramenta e observe instantaneamente como está a escrita do aluno orientando quanto a não utilização de caracteres repetitivos e descontextualizados com o tema.

A ferramenta Sobek demonstrou ser um recurso que viabiliza ao professor a análise das produções textuais dos alunos, possibilitando-o à percepção quanto à correspondência entre a escrita e o tema proposto.

O intuito desta pesquisa, que é responder a indagação se as ferramentas de mineração de texto podem auxiliar na análise das produções textuais dos alunos do Ensino Fundamental, foi alcançado, assim como foram alcançados os objetivos específicos. Pode-se considerar satisfatório os resultados para ambas ferramentas.

A pesquisa identificou que as ferramentas de mineração de texto podem auxiliar o docente na correção das produções textuais dos alunos. Caso haja a necessidade de análise de dezenas ou centenas de produções textuais, as ferramentas podem ser muito úteis.

O presente trabalho, pautado na experiência com utilização de duas ferramentas de mineração de texto para análise da produção textual dos alunos do ensino fundamental, também pôde ressaltar a importância de inserir nas práticas pedagógicas as tecnologias digitais como aliadas no processo de ensino e aprendizagem.

## 6 Referências Bibliográficas

ALVES, V. Q.; MOREIRA, R.; DIAS, P. R. A.. **Arte e Filosofia**: contextualizando a prática docente com Tecnologias de Comunicação Digital na Educação Básica. *Integração de Tecnologias na Educação: Práticas Inovadoras na Educação Básica*, [S.l.], v. 1, p. 28 - 35, nov. 2017. Disponível em: <<https://publicacoes.rexlab.ufsc.br/index.php/ebook-integracao/article/view/104>>. Acesso em: 10 maio 2018.

ARANHA, C. PASSOS, E. A Tecnologia de Mineração de Textos. **RESI-Revista Eletrônica de Sistemas de Informação**, n.2, 2006. Disponível em:><http://www.spell.org.br/documentos/ver/26518/a-tecnologia-de-mineracao-de-textos><. Acesso em 28 abril 2018.

BAKHTIN, M. **Estética da criação verbal**. 4 ed. São Paulo: Martins Fontes, 2003.

BRASIL. S.E.F. **Parâmetros curriculares nacionais** : matemática / Secretaria de Educação Fundamental. – Brasília : MEC/SEF, 1997.

FAZENDA, Ivani (Org.). **O que é interdisciplinaridade?** São Paulo: Cortez, 2008.

FELDMAN, R. SANGER J. **The text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data**. Cambridge, MA: Cambridge University Press, 2007.

JOHNSON, S. **Cultura da Interface**: Como o computador transforma nossa maneira de criar e comunicar. Rio de Janeiro, Jorge Zahar, 2001.

KLEMMANN, Angela. **Texto e leitor: aspectos cognitivos da leitura**. 12. ed. São Paulo: Pontes, 2009.

\_\_\_\_\_, LORENZATTI, A., REATEGUI, E. O Emprego da Ferramenta de Mineração de Textos SOBEK como Apoio à Produção Textual. Em *Anais do Simpósio Brasileiro de Informática na Educação*, 2009, Florianópolis.

KLEIMAN, M.; REATEGUI, E.; RAPKIEWICZ, C. Análise de Ferramentas de Mineração de Textos para Apoio à Produção Textual. *Anais do XXII SBIE – XVII – WIE*. Aracaju, Nov. 2011. Disponível em: <<http://www.br-ie.org/pub/index.php/sbie/article/view/1866>>. Acesso em: 28 abril 2018.

KOCH, I. G. V. **A inter-ação pela linguagem**. São Paulo: Contexto, 1992.

MACEDO, A.; REATEGUI, E.; LORENZATTI, A.; BEHAR, P. Using Text-Mining to Support the Evaluation of Texts Produced Collaboratively. *Education and Technology for a Better World: Selected papers of the 9th World Conference on Computers in Education*, Bento Gonçalves, Brazil. Springer, 2009.

NEVES, V.F.A.; GOUVÊA, M.C.S.; CASTANHEIRA, M.L. O letramento e o brincar em processos de socialização na educação infantil. **Revista Brasileira de**



**Educação – v.20, n.º 60, p.215-244, jan-mar, 2015.**

RIZZO, G. **Educação Pré-Escolar**. Rio de Janeiro, 4ª ed. Francisco Alves, 1986.

VALENTE, J.A.; BIANCONCINI DE ALMEIDA, M. E.; FLOGI SERPA GERALDINI, A.. Metodologias ativas: das concepções às práticas em distintos níveis de ensino.

**Revista Diálogo Educacional, [S.l.], v. 17, n. 52, p. 455-478, jun. 2017. ISSN 1981-416X.**

Disponível

em:

<<https://periodicos.pucpr.br/index.php/dialogoeducacional/article/view/9900/12386>>.

Acesso em: 06 abr. 2018.

VEER, R. V. D.; VALSINER, J. **Vygotsky** - Uma síntese. São Paulo: Loyola,