

A AVALIAÇÃO DA PRODUÇÃO TEXTUAL NOS VESTIBULARES E OUTROS CONCURSOS: A QUESTÃO DA SUBJETIVIDADE

ELIANA AMARANTE DE MENDONÇA MENDES*

Recebido: 25 abr. 2012

Aprovado: 6 jun. 2012

* Doutora em Filologia e Língua Portuguesa pela Universidade de São Paulo. Professora Titular em Estudos Linguísticos do Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais. Belo Horizonte, MG, Brasil. E-mail: mendes@ufmg.br

Resumo: O objetivo deste artigo é refletir sobre o problema da subjetividade na avaliação de redações em vestibulares e outros concursos e demonstrar a inviabilidade de se manter tal tipo de prova, dada a impossibilidade de se fazer justiça aos candidatos. Apresento uma sucinta revisão bibliográfica sobre o tema, alguns dados históricos sobre a avaliação através dos tempos e o estado da arte no Brasil. Análise o sistema de avaliação de redações do ENEM, mostrando sua insustentabilidade. Apoio-me, teoricamente, em Bakhtin e em especialistas em avaliação. Apresento, também, uma pequena pesquisa empírica sobre o sistema de avaliação de redações do ENEM.

Palavras-chave: Subjetividade. Avaliação da produção escrita. Insustentabilidade. ENEM.

ASSESSING WRITING IN SCREENING PROCESSES: THE SUBJECTIVITY ISSUE

Abstract: I intend in this paper to offer some reflections on the problem of subjectivity in the writing assessment, in the examinations for access to the University and other public competitions in Brazil, and to demonstrate the unfeasibility of keeping such type of test, given the impossibility of doing justice to the candidates. I present a brief history of assessment over time, a succinct bibliographical summary on the subject and some data on the state of the art in Brazil. I analyze the system of writing assessment of the ENEM, showing its insustainability. I support myself, theoretically, in Bakhtin and specialists in assessment and evaluation. I present, also, a small empirical research on the system of writing evaluation of the ENEM.

Key words: Subjectivity. Writing assessment. Insustainability. ENEM

1 O VALOR DA REDAÇÃO NUMA CULTURA QUE SE QUER LETRADA

Há consenso em nossa sociedade acerca da importância de saber redigir. Isso se evidencia na presença da redação nos currículos de todos os níveis de ensino, até mesmo em currículos de cursos superiores. Em decorrência, temos também, em praticamente todos os concursos vestibulares das universidades e em concursos públicos para os mais diversos cargos, nas mais diversas instituições, provas de redação e de questões abertas, discursivas, consideradas por todos indispensáveis.

Problemas relativos às políticas nacionais de avaliação têm sido muito discutidos pelos especialistas. Existem vários e importantes trabalhos sobre

avaliação, que focalizam as grandes avaliações nacionais e internacionais. Encontram-se também, em número considerável, trabalhos sobre a avaliação formativa, aquela que se desenvolve na sala de aula, e sobre os problemas encontrados nas redações de vestibulandos. Embora tragam reflexões importantes, tais trabalhos não trazem contribuição para a compreensão dos processos de avaliação somativa, que é o que me interessa aqui. Faltam, portanto, estudos específicos sobre a avaliação da produção textual escrita, produzida no contexto de vestibulares e concursos públicos.

2 SOBRE A AVALIAÇÃO DE REDAÇÕES

A avaliação é sempre uma tarefa complexa e delicada. No caso da avaliação de redações, reveste-se de complexidade ainda maior, principalmente quando circunscrita a contextos de seleção e classificação de candidatos em vestibulares e concursos públicos. Nesse cenário de grande competição por uma vaga, qualquer valor decimal a maior ou a menor pode levar à classificação ou desclassificação do candidato. Quanto maior o número de redações, maior a complexidade, pois participam do processo subjetividades, valores, gostos e conhecimentos dos organizadores, dos avaliadores e dos candidatos. Nesse contexto, é preciso evitar que o candidato seja tratado como participante de uma loteria ou de um júri, sem direito de defesa e contraditório.

Apesar de toda essa complexidade, são muito raras as pesquisas específicas sobre o assunto no Brasil. São essas lacunas que me levam a propor uma reflexão sobre a questão da subjetividade na avaliação da produção textual nos vestibulares e concursos públicos, no entendimento de que é necessário dar mais transparência a esse assunto, que costuma ser tratado como um tabu, a portas fechadas.

3 PEQUENA HISTÓRIA DA AVALIAÇÃO

Segundo Otto Rank (1974, p. 246), a cultura da avaliação tem suas origens nos ritos de passagem, testes para introduzir os adolescentes à idade adulta, certificando a maturidade sexual, mas também introduzindo os adolescentes na cultura, valores e costumes de um grupo. Na Grécia e em Roma, embora não houvesse um sistema formal de avaliação, havia o reconhecimento do mérito dos grandes oradores e as sessões de *declamatio*, em que os estudantes avançados mostravam seus dotes oratórios e eram avaliados.

Na Europa ocidental, os primeiros sistemas de avaliação eram orais: as arguições nas primeiras universidades para aferição dos graus acadêmicos. Só a partir do século XVIII, após a introdução e difusão do papel no ocidente, foram introduzidos os exames escritos nas universidades europeias.

Na China, a cultura escrita, o uso do papel, dos pincéis e os livros impressos existiram muito antes. Por isso foram os chineses também pioneiros na avaliação escrita, que tinha o objetivo de selecionar as melhores pessoas para desempenhar tarefas específicas valorizadas por um grupo. Os exames escritos dos chineses começaram na dinastia Sui, (589 a 618 d. C.). Por volta de 1370, esses exames tinham similaridade com nossas atuais avaliações somativas de produção escrita: composições em prosa e em verso revelavam a extensão da leitura e a profundidade dos conhecimentos do candidato. A avaliação era feita por dois leitores independentes; e um terceiro recebia e conciliava os graus atribuídos. (DUBOIS, 1970, p. 4)

Importante sistema de avaliação formativa foi o jesuítico, pioneiro da avaliação na educação: previa testes escritos para nivelar os alunos ingressantes, para avaliar o conhecimento adquirido e para promover a competição entre o alunado.

Segundo DuBois (1970, p. 10), em meados do século XIX testes orais e escritos já eram rotina na Europa e nos Estados Unidos. Os exames escritos eram reconhecidos como base apropriada para decidir quem receberia graus acadêmicos, poderia exercer determinadas profissões, poderia ser funcionário público, etc. No Brasil a situação era similar. Os exames preparatórios, precursores dos atuais vestibulares, também incluíam exames orais e escritos.

A partir dessa fase, muitos refinamentos foram feitos à avaliação, principalmente quanto aos procedimentos para gradação das notas. Em 1864, segundo Dubois (1970, p.69), George Fischer, da Universidade de Greenwich, coletou exemplares de escritos acadêmicos, organizou-os em uma escala e atribuiu valores de 1 (para os melhores) a 5 (para os piores). Os valores intermediários eram indicados por frações. Com base nessa coletânea de padrões, atribuía-se, por comparação, a nota a qualquer outro trabalho. Pouco a pouco, foram surgindo outros procedimentos para medida do conhecimento através da produção escrita. Até o início do século XX, a regra era basear a avaliação numa tarefa completa, em respostas discursivas longas, orais ou escritas. Conforme DuBois (1970, p. 76-77), só muito recentemente é que foi desenvolvida a noção de medição indireta, cujo exemplo mais difundido são os testes de múltipla escolha. Nos anos de 1920 houve um *boom* de avaliação por testes desse tipo, principalmente porque com eles eliminava-se praticamente todo o favoritismo

pessoal que pudesse influenciar as notas. Na próxima fase, passou-se a adotar também os testes de respostas curtas, que logo se tornou a norma de avaliação nas escolas: reflexo de uma sociedade que queria respostas simples e curtas para problemas complexos. Apesar das vantagens práticas do sistema de múltipla escolha e do de perguntas e respostas curtas, muitos especialistas consideram que o pensamento coerente e a expressão desse pensamento não podem ser medidos por esses tipos de teste. A melhor avaliação deve ser através da escrita de textos completos. É essa postura que levou à adoção, atualmente, da prova de redação, consensualmente considerada indispensável.

4 SOBRE A AVALIAÇÃO

A avaliação é uma tarefa necessária: permite acompanhar o desenvolvimento dos alunos, testar a competência dos professores e a adequação da metodologia e da didática. Serve também para acompanhamento das políticas de educação no país e para selecionar e classificar pessoas para diversos fins.

Segundo Davis et al (1987, p. 3-4), existem diversos tipos, processos e modalidades de avaliação. Primeiramente, tem-se que considerar os tipos de medição do conhecimento: o conhecimento pode ser medido diretamente na produção do avaliado, por exemplo, em testes de redação, perguntas e respostas curtas, completar lacunas, etc; pode-se usar também o método de medição indireta, cujo exemplo mais prototípico são os testes de múltipla escolha. Os testes podem ser ainda orais, escritos e práticos (como os aplicados em laboratórios). Há ainda dois grandes processos de avaliação: a avaliação formativa, que permite determinar o grau de conhecimento dos estudantes ou o grau do que eles estão aptos a fazer. Os resultados permitem a previsão das futuras ações para o ensino/aprendizagem. Existe ainda outro processo, a avaliação somativa, de grande importância por seu caráter democrático e meritocrático. É geralmente aplicada no fim de um curso ou de um projeto, usada para atribuir o grau do curso ao estudante e para selecionar e classificar pessoas para diversos fins, como no caso dos vestibulares e outros concursos públicos. Há que se considerar ainda dois tipos de avaliação: Avaliação do processo, que consiste na avaliação do método usado para fazer algo, geralmente envolvendo etapas ou operações, ordenadas ou interdependentes. Por exemplo, o exercício de preescrita, que leva à produção final de um texto. Há também a avaliação do produto, que consiste na avaliação do resultado tangível e estável de uma *performance* ou tarefa. Finalmente, a avaliação pode ser holística, quando consiste na atribuição de uma pontuação única baseada numa avaliação geral da *performance*, não ana-

lisando e pontuando as dimensões individualmente. O produto é considerado ser mais que a soma de suas partes e por isso avalia-se a qualidade do produto final. A avaliação pode ser ainda analítica, no caso em que se segmenta o todo em categorias de critérios que são examinadas individualmente. Esse tipo de avaliação é útil quando há várias dimensões a serem avaliadas num trabalho, como, por exemplo, a redação.

As provas de redação nos nossos processos de avaliação somativa, nos vestibulares e outros concursos, caracterizam-se como avaliação de medição direta, de produto, analítica, na modalidade escrita, geralmente solicitando a escrita de textos longos.

5 A QUESTÃO DA SUBJETIVIDADE

A subjetividade, maior problema da avaliação, é o principal interesse deste trabalho. Na avaliação analítica, a subjetividade já se manifesta na definição das competências a serem consideradas. Por exemplo, há universidades que segmentam a avaliação da redação em 6 competências, já em outras (e também no ENEM) são consideradas 5 competências. Manifesta-se também na definição dos níveis e na pontuação de cada um dos níveis das competências. Essas definições envolvem valores, por natureza subjetivos. A escolha das competências e a pontuação dos níveis podem favorecer alguns candidatos e penalizar outros. Neste trabalho, analiso a questão da subjetividade na avaliação das redações do ENEM.

6 A SUBJETIVIDADE NO MODELO DO ENEM

Nos modelos do ENEM 2010 e 2011, encontram-se evidências do problema da subjetividade na descrição dos níveis das competências. Como se poderá verificar, esses descritores são eivados de modalizadores, apresentam vagueza conceitual e termos obscuros, identificados por mim e sublinhados no documento abaixo. Na versão dos critérios divulgada em 2011, os responsáveis procuraram “traduzi-los”, certamente para torná-los mais claros e operacionalizáveis: a descrição dos níveis das competências foi ampliada e foram acrescentadas observações complementares para cada uma das competências. Em negrito, encontram-se os acréscimos à versão de 2010, seguidos de meus comentários.¹

1 Aspas duplas foram inseridas, no modelo de 2011, em alguns termos e expressões. Presume-se, para indicar aspectos problemáticos que carecem de esclarecimentos.

Simulado 1 ENEM 2011 - REDAÇÃO NO ENEM – CRITÉRIOS DE CORREÇÃO

COMPETÊNCIA I - Demonstrar domínio da norma culta da língua escrita.

1) Demonstra conhecimento “precário” da norma culta: inadequação na escolha da variedade linguística, “graves e frequentes” desvios gramaticais e transgressões “inaceitáveis” das convenções da escrita. **Neste nível o participante emprega expressão muito próxima da modalidade oral e registro inadequado à proposta da prova; demonstra inadequação vocabular; revela conhecimento precário das normas gramaticais e das convenções básicas da escrita.**

. Aqui encontra-se a tradução de ‘precário’ e ‘inaceitáveis’ termos vagos, mas se introduzem as expressões ‘muito próxima’ e ‘básicas’ que, por sua vez, também são vagas.

2) Demonstra conhecimento “razoável” da norma culta: problemas na escolha da variedade linguística; desvios gramaticais e transgressões das convenções da escrita “pouco aceitáveis” nessa etapa de escolaridade. **Considera-se como “pouco aceitável” o uso que constitua desvio da sintaxe de concordância, regência e colocação, da pontuação e da flexão.**

. Aqui explicita-se o que se deve entender por ‘pouco aceitável’, o que confere maior clareza à instrução.

3) Demonstra “bom” domínio da norma culta (ainda que com “pontuais” desvios gramaticais ou transgressões pontuais das convenções da escrita). **Espera-se que o participante empregue um léxico produtivo, ortografia e pontuação pertinentes, demonstre um bom domínio da sintaxe e das regras de flexão. Neste nível podem ocorrer alguns desvios em relação à norma escrita culta.**

. Não se traduz pontuais e introduz-se a expressão “léxico produtivo”, sem defini-la.

4) Demonstra “muito bom” domínio da norma culta (ainda que com “um ou outro deslize” relativo à norma gramatical ou às convenções da escrita). **O nível 4 é atribuído a textos em que o produtor tenha plena autonomia com relação à utilização da norma escrita culta, ainda que possa apresentar um ou outro deslize.**

. Não se define ‘um ou outro deslize’, expressão vaga.

OBSERVAÇÕES COMPLEMENTARES

Espera-se que, nesta competência, o participante escolha a variedade linguística adequada à situação formal de produção de texto escrito. Na avaliação, devem ser considerados os conhecimentos de língua escrita representados pela utilização da norma culta: adequação ao registro (escolha da variedade linguística adequada ao tipo de texto e à situação formal de interlocução); obediência à norma gramatical (sintaxe de concordância, regência e colocação, pontuação, flexão) e às convenções da escrita (ortografia, acentuação, maiúsculas/minúsculas, etc.).
--

. Aqui, acrescenta-se a ‘adequação ao registro’, um quesito importante, mas que, a rigor, deveria constar da descrição da competência e não da descrição dos níveis.

COMPETÊNCIA II - Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo.

1) Desenvolve “tangencialmente” o tema em um texto que apresenta características do tipo de texto dissertativo--argumentativo; ou apresenta “embrionariamente” o tipo de texto dissertativo-argumentativo (sem “fugir” do tema). **Considera-se “tangencial” o texto que se aproxima apenas por generalidades do tema proposto ou o texto resultante da adaptação forçada de um projeto de texto ao tema selecionado para a redação. Alguns casos podem apontar textos preparados previamente à realização da prova. Outro caso de atribuição do nível 1 é a apresentação de uma argumentação mínima (embrionária) inserida num texto de outra natureza (tal como carta e narrativa), diferente da estrutura dissertativa proposta.**

. A tradução apresentada para ‘tangencial’ não é esclarecedora, uma vez que se introduz ‘que se aproxima apenas por generalidades’, explicação obscura. A explicação a ‘argumentação insuficiente’ e ‘texto de outra natureza’ são esclarecedoras. Entretanto, entendo que a menção a ‘textos preparados previamente’, embora se saiba que isso costuma ocorrer, não é pertinente, pois trata-se de uma suposição de comprovação impossível.

2) Desenvolve “razoavelmente” o tema, a partir de considerações próximas do senso comum ou por meio de paráfrases dos textos-estímulo, e domina precária/razoavelmente [“precariamente” ou “razoavelmente”] o tipo de texto dissertativo-argumentativo. **O nível 2 indica a presença de um texto construído com base em considerações próximas do senso comum, com pouca reflexão do participante e muita repetição de chavões e clichês; a estrutura é pouco encadeada e sem progressão. Nesse caso, o participante revela ausência de projeto e demonstra tentar discutir o tema à medida que vai escrevendo.**

. A menção, a ‘chavões e clichês’, traduzindo ‘próximas ao senso comum’ é esclarecedora bem como a menção a ‘sem de progressão’. Substitui-se ‘precária/razoavelmente’ por ‘precariamente ou “razoavelmente’’. Considerando que o termo ‘precário’ é usado para a caracterização do nível 1 e ‘razoável’ para o nível 3, o uso simultâneo dos dois termos evidencia uma indefinição quanto à característica deste nível.

3) Desenvolve “bem” o tema, mesmo apresentando argumentos previsíveis, e domina “bem” o tipo de texto dissertativo-argumentativo, com indícios de autoria. **O nível 3 indica um bom desenvolvimento do tema, como consequência de sua correta compreensão, ainda que com argumentos previsíveis que não impedem, contudo,**

o surgimento de indícios de autoria. Evidencia-se um bom domínio das técnicas dissertativo-argumentativas.

. É esclarecedora a menção a um bom domínio do texto dissertativo-argumentativo. ‘Argumentos previsíveis’, no entanto, só podem ser objetivamente avaliados, caso se trate de repetição dos argumentos do texto constante da proposta da questão. A expressão ‘indícios de autoria’ é obscura, de difícil compreensão.

4) Desenvolve “muito bem” o tema, a partir de um projeto pessoal de texto e de um repertório cultural produtivo; domina “muito bem” o tipo de texto dissertativo-argumentativo. **Este nível indica o desenvolvimento seguro e autônomo do tema, a partir de um repertório cultural produtor e de considerações que fogem ao senso comum. O domínio da estrutura dissertativo-argumentativa é muito bom. O projeto de texto do participante define-se com base na compreensão plena da proposta e é desenvolvido segundo a articulação de diferentes perspectivas para a defesa do ponto de vista. Revela ainda marcas de autoria.**

. As explicações para ‘muito bem’ são ainda complexas: ‘repertório cultural produtor’, ‘fugir ao senso comum’ e ‘marcas de autoria’ são conceitos cuja interpretação vai depender muito dos pontos de vista do avaliador.

OBSERVAÇÕES COMPLEMENTARES

A Competência II tem o eixo na compreensão do tema, a respeito do qual se pede uma reflexão por escrito, sob a forma de texto dissertativo-argumentativo. Na dissertação, o enunciador apresenta explicitamente sua opinião, valendo-se do recurso dos argumentos de apoio para comprovar suas hipóteses e tese e assegurar o desenvolvimento de seu projeto de texto. Esta competência envolve dois grandes momentos: o da leitura/interpretação da proposta e o da compreensão transposta para o projeto de texto. O participante deve exercer simultaneamente o papel de leitor da proposta e produtor/leitor de seu próprio texto. Os critérios de avaliação da competência procuram verificar o desempenho do participante em uma situação formal de interlocução em que determinado tema é discutido. Devem ser considerados os aspectos: a) relativos ao tema (compreensão da proposta e desenvolvimento do tema a partir de um projeto de texto); b) relativos à estrutura do texto dissertativo-argumentativo (encadeamento e progressão temática). O texto que não desenvolve o tema proposto ou que não se enquadra no tipo solicitado (dissertativo-argumentativo) será “Desconsiderado”, em virtude de o participante não ter compreendido a proposição (quanto ao tema ou à tipologia) ou ter decidido produzir outro texto (quanto ao tema ou à tipologia).

. Aqui descreve-se o texto ideal, mas não há contribuição para a elucidação dos níveis e de como distingui-los e graduá-los.

COMPETÊNCIA III - Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.

1) Apresenta informações, fatos e opiniões “precariamente” relacionados ao tema. **O nível 1 indica textos com problemas de coerência decorrentes da maneira como as informações, os fatos e as opiniões relacionam-se entre si ou com o tema; ou problemas de coerência de textos que, explicitamente, procuram adequar ao tema proposto o projeto desenvolvido para outro tema (caso comentado na Competência II, nível 1). Nesse nível não há seleção de argumentos.**

. As explicações mencionam ‘problemas de coerência’, sem explicitar como quantificá-los para efeito de pontuação.

2) Apresenta informações, fatos e opiniões “razoavelmente” relacionados a um embrião de projeto de texto e/ou limita-se a reproduzir os elementos fornecidos pela proposta de redação. **O nível 2 indica textos que apresentam uma estrutura fragmentada ou circular, com eventuais contradições; também se aplica a textos que se limitam a reproduzir ou a parafrasear os argumentos, fatos e opiniões constantes da proposta de redação.**

. Aqui a menção a ‘estrutura fragmentada ou circular’ é esclarecedora, bem como a menção a ‘paráfrase de argumentos e fatos constantes da proposta de redação’.

3) Seleciona informações, fatos, opiniões e argumentos, “relacionando-os” ao seu projeto de texto. **O nível 3 indica textos realizados a partir de um projeto pessoal, com boas seleção e interpretação de informações, fatos, opiniões e argumentos que resultam bem-relacionados a esse projeto.**

Cumpra ainda esclarecer como se sabe se um projeto é pessoal ou não e o que significa de fato ‘boas seleção e interpretação de informações’.

4) Seleciona, interpreta e organiza informações, fatos, opiniões e argumentos, estabelecendo uma relação produtiva entre essa seleção e seu projeto de texto. **O nível 4 indica textos realizados a partir de um projeto pessoal, com excelente seleção, interpretação e organização de informações, fatos, opiniões e argumentos que se relacionam de forma produtora a esse projeto. Observa-se que a seleção de elementos feita pelo participante contribui para a defesa coerente do ponto de vista destacado no seu projeto de texto.**

. Continua difícil distinguir uma boa argumentação de uma excelente argumentação.

OBSERVAÇÕES COMPLEMENTARES

Nesta competência, espera-se que o participante, em uma situação formal de interlocução, saiba selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos pertinentes para defender sua perspectiva sobre o tema proposto. O que será avaliado nos textos produzidos é a coerência na seleção, organização e exposição dos argumentos para a defesa de ponto de vista em relação ao tema e ao projeto de texto desenvolvido.

. Aqui descreve-se um texto ideal, mas não há contribuição para a elucidação dos níveis – como distingui-los e graduá-los –, visando à atribuição de notas diferentes.

COMPETÊNCIA IV - Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação.

1) Articula “precaramente” as partes do texto. **A estrutura de superfície típica do nível 1 é a justaposição de palavras e frases pouco relacionadas, deixando o estabelecimento da coesão a cargo do leitor.**

. É esclarecedora a menção a ‘justaposição de palavras e frases pouco relacionadas, deixando o estabelecimento da coesão a cargo do leitor’.

2) Articula “razoavelmente” as partes do texto, apresentando “problemas frequentes” na utilização dos recursos coesivos. **O texto de nível 2 indica uma grande dificuldade do produtor na utilização da modalidade escrita; mesmo que tente articulá-lo, esbarra em problemas não solucionados durante a escolarização. O participante demonstra não ter conseguido incorporar à sua produção textual as regras que regem a articulação do texto formal.**

. Continua sem explicação o que sejam ‘problemas frequentes’, quão frequentes? Introduce-se, ainda, a expressão ‘grande dificuldade’, sem esclarecê-la. Ademais, parece-me que, nesta competência, a descrição dos níveis 1 e 2 estão invertidas. No meu entender, seguindo essas instruções, um texto de nível 2 é pior que o de nível 1.

3) Articula “bem” as partes do texto (ainda que apresente “problemas pontuais” na utilização dos recursos coesivos). **O texto de nível 3 demonstra domínio dos mecanismos de coesão, embora apresente ainda alguns desvios com relação à utilização desses mecanismos.**

. Aqui se traduz ‘problemas pontuais’ por ‘alguns desvios’, ambas expressões vagas.

4) Articula “muito bem” as partes do texto (ainda que apresente “eventuais deslizes” na utilização de recursos coesivos). **O texto típico do nível 4 indica que o produtor tem plena autonomia com relação à utilização dos mecanismos coesivos da norma escrita culta, ainda que possa apresentar um ou outro deslize.**

. Impossível aqui é distinguir o nível 3 do nível 4. O nível 3 prevê ‘eventuais deslizes’ ou ‘um ou outro deslize’ e no nível 4 são tolerados ‘um ou outro desliz’e’.

OBSERVAÇÕES COMPLEMENTARES

Nesta competência, avaliam-se os recursos coesivos empregados na modalidade escrita, com vista à adequada articulação dos argumentos, fatos e opiniões selecionados para a defesa do ponto de vista sobre o tema proposto. Os mecanismos coesivos são os responsáveis pela construção da argumentação na superfície textual. Assim, é avaliado o uso dos seguintes mecanismos: coesão lexical (uso de sinônimos, hiperônimos, repetição, reiteração, etc.); coesão gramatical (uso de conectivos, tempos verbais, pontuação, sequência temporal, relações anafóricas, conectores intervocabulares, intersentenciais, interparágrafos, etc.). Os níveis 1 e 2 indicam textos semelhantes aos destacados nos níveis 1 e 2 da Competência I. Os produtores desses textos apresentam grande dificuldade em registrar por escrito suas idéias em situação formal. Em relação aos níveis 3 e 4, comprova-se domínio dos mecanismos de coesão.

. Tem-se aqui uma aula sobre mecanismos de coesão textual, sem, contudo, esclarecer bem como distinguir os níveis.

COMPETÊNCIA V - Elaborar proposta de intervenção para o problema abordado, demonstrando respeito aos direitos humanos.

1) Elabora proposta “precaricamente” relacionada ao tema, respeitando os direitos humanos. **O nível 1 indica que houve pouca compreensão do tema e que a proposta apresentada apenas o tangencia.**

. Aqui busca-se traduzir ‘proposta precariamente relacionada ao tema’, introduzindo-se ‘pouca compreensão do tema’ e ‘ proposta apenas tangencia o tema’. A vagueza permanece.

2) Elabora proposta “razoavelmente” relacionada ao tema, mas “não articulada” ao texto desenvolvido, respeitando os direitos humanos. **O nível 2 indica textos que apresentam proposta relacionada ao tema, mas sem vinculação ao texto desenvolvido.**

. Traduz-se ‘razoavelmente relacionada ao tema, mas não articulada ao texto’, por ‘relacionada ao tema, mas sem vinculação ao texto desenvolvido’, mantendo-se a dificuldade de entendimento do critério.

3) Elabora proposta “bem-relacionada” ao tema, mas “pouco articulada” à discussão desenvolvida no texto, respeitando os direitos humanos. **O participante preocupa-se em elaborar proposta explícita para a problemática discutida, no entanto, ela resulta pouco articulada à discussão.**

. A explicação não facilita o entendimento e é difícil distinguir o nível 2 do nível 3.

4) Elabora proposta “bem-relacionada” ao tema e “bem-articulada” à discussão desenvolvida no texto, respeitando os direitos humanos. **O participante elabora proposta explícita para a situação-problema e articula-a coerentemente à discussão.**

. A distinção entre os níveis 3 e 4 é também problemática. É difícil entender a diferença entre ‘preocupa-se em elaborar proposta explícita’ e ‘elabora proposta explícita’.

OBSERVAÇÕES COMPLEMENTARES

<p>Nesta competência, procura-se avaliar as possíveis variáveis para solucionar a situação-problema apresentada: a(s) proposta(s) solidária(s) de intervenção do participante, <u>a qualidade</u> e a relação de suas propostas com o projeto desenvolvido por ele sobre o tema, respeitando os direitos humanos. Quando o participante elabora propostas que ferem, de maneira explícita, os direitos humanos, o texto é desconsiderado (D) com base nesta competência. Nos níveis 1 e 2, as propostas ficam, por vezes, subentendidas, isto é, podem ser inferidas pelo leitor sem que tenham sido explicitamente desenvolvidas pelo participante. Nos níveis 3 e 4, verifica-se a preocupação de o autor elaborar proposta efetiva para o tema discutido.</p>
--

. Essas explicações ajudam, mas não esclarecem totalmente o que constitui a qualidade da proposta de texto.

Apesar do meritório esforço no sentido de tentar esclarecer os critérios nessa versão de 2011, e de se constatar alguma melhoria no modelo, no todo a definição dos níveis das competências continua sendo um problema, quer-me parecer, sem solução. Por mais que se tente aprimorar, buscando mais clareza e objetividade, não se chega a um modelo satisfatório, não se consegue evitar as vaguezas, obscuridades e modalizações constantes nos descritores, incompatíveis com a necessária objetividade para se fazer justiça a todos os candidatos.

7 OUTROS PROBLEMAS NA AVALIAÇÃO DO ENEM

Além do problema da subjetividade, outro problema da avaliação analítica é decorrente da própria segmentação: uma vez que o texto é um todo, segmentar a avaliação pode levar à perda da noção do todo. Tomemos como exemplo, as competências II e V do sistema do ENEM. Se o candidato é bem avaliado na competência II, é incoerente ele ser mal avaliado na competência V. Ou seja, uma boa avaliação na competência II implica boa avaliação na competência V, sendo o contrário também verdadeiro. Assim, um candidato que “desenvolve muito bem o tema, a partir de um projeto” (II,4), não pode elaborar proposta “precariedade relacionada ao tema, respeitando os direitos humanos”(V,1). O modelo em questão não evita tal incoerência. Da mesma forma, detecta-se esse tipo de incoerência em relação às competências II e III, e II, III e IV. Portanto, um modelo analítico de avaliação deveria levar em consideração esse problema e desenvolver uma maneira de integrar alguns dos segmentos, estabelecendo relação de implicação entre eles. Isso me leva a questionar o próprio modelo analítico, já que a avaliação por segmentação revela-se não adequada.

Cumpra ainda notar que esse sistema de avaliação é uma mistura de avaliação analítica com um teste de múltipla escolha, com quatro opções de resposta, sendo imposto ao avaliador optar por uma delas. Segundo especialistas, modalizadores, termos vagos e termos obscuros devem ser evitados em testes de múltipla escolha.

Os responsáveis por essa prova nos vestibulares, tendo consciência dos problemas, procuram minorá-los de diversas formas. Por exemplo, adotando o sistema de dupla correção. Esse sistema, embora não elimine toda subjetividade, a minimiza. Como a experiência mostra que são raríssimos os casos de atribuição da mesma nota pelas duplas de avaliadores, costuma-se estabelecer uma margem de tolerância: se a diferença é de até X pontos, vale a nota maior (prevalece a subjetividade de quem atribuiu nota maior); se a diferença é maior que X pontos, tira-se a média das duas notas (divide-se a responsabilidade pelas duas subjetividades); se a diferença é ainda maior, discutem os dois avaliadores para se chegar a um consenso (prevalece aí a subjetividade do avaliador que argumentar melhor). Costuma-se também usar da intervenção de um árbitro, profissional mais experiente, que vai decidir sobre a nota (vale aí a subjetividade do árbitro). Como se vê, apesar dos esforços das equipes, não se fica livre das interferências da subjetividade.

8 A AVALIAÇÃO E BAKHTIN

Como vimos, desde que existem sistemas de avaliação, a subjetividade implicada no processo preocupa. Muito antes de Bakhtin. Com Bakhtin, no entanto, temos que pensar a subjetividade de um modo diferente. Não mais no subjetivismo idealista, do domínio do sujeito, do pré-domínio do sujeito, ou seja um sujeito ideal como o literato, sujeito da estilística, ou como o sujeito cognoscente de Piaget, nem o falante ouvinte ideal de Chomsky. Também não mais no objetivismo abstrato, do domínio da estrutura, do não domínio do sujeito, do não sujeito. Bakhtin postula o co-domínio do sujeito. O sujeito bakhtiniano é dialógico – isto significa ser constituído por um outro, social, histórico e ideológico. A constituição do eu se dá a partir da constituição do outro. Dentre os desdobramentos dessa concepção de Bakhtin, recorro a alguns aspectos que podem me ajudar a pensar a subjetividade na avaliação de redações.

Para Bakhtin, o texto não traz um sentido completo e estável, o sentido é construído também no ato de leitura, pelo leitor. Em suas palavras, “A compreensão é uma forma de diálogo; ela está para a enunciação assim como uma réplica está para a outra no diálogo. Compreender é opor à palavra do locutor uma contra-palavra.” (BAKHTIN, 1981, p. 9) Assim sendo, um enunciado é um ato de linguagem cujos contornos permitem e solicitam que um outro realize uma apreciação valorativa com relação àquilo que falamos ou escrevemos. Nossas palavras se dirigem à apreciação de um outro, tanto quanto suas palavras se dirigem à nossa apreciação. Sob essa ótica, a avaliação do texto deve ser conduzida como um processo cooperativo e dialógico, em que educadores e aprendizes se conheçam através da própria avaliação. Essa possibilidade só existe, no entanto, para a avaliação formativa, em sala de aula. No caso da avaliação somativa, a avaliação é um encontro entre autor e leitor com o objetivo de apreciar para comparar, classificar, selecionar e pontuar. E sem a possibilidade de uma interlocução que dê conta de conferir o que cada um compreendeu a respeito do discurso do outro. Aí se encontra a maior dificuldade, se não a impossibilidade, da avaliação somativa de textos. A apreciação, nos termos bakhtinianos, não é passível de ser convertida em uma pontuação e classificação, uma vez que não existem parâmetros objetivos que viabilizem tal medição. Para Bakhtin (2010, p. 39),

[...] cada texto (como enunciado) é algo individual, único e singular, e nisso reside todo o seu sentido (a sua intenção em prol da qual ele foi criado). É aquilo que nele tem relação com a verdade, com a bondade, com a beleza, com a história.

Entretanto, para ele, “[...] por trás de cada texto está o sistema da linguagem. A esse sistema correspondem no texto tudo o que é repetido e reproduzido e tudo o que pode ser repetido e reproduzido, tudo o que pode ser dado fora de tal texto (o dado).” (p. 39). Esse sistema da linguagem é da esfera da linguística positivista, cujo método é racional, quantitativo. Para o positivismo só é real e material aquilo que pode ser medido, pesado, tocado, manipulado. Referindo-se à linguística positivista, diz Bakhtin (2006, p. 79):

Não há lugar, aqui, para quaisquer distinções ideológicas, de caráter apreciativo: é pior, é melhor, belo ou repugnante, etc. Na verdade só existe um critério lingüístico: está certo ou errado; além do mais, por correção lingüística deve-se entender apenas a conformidade a uma dada norma do sistema normativo da língua.

Ainda, para Bakhtin (2006, p. 17), nas palavras de Jakobson, “[...] todos os procedimentos de análise lingüística (fonéticos, morfológicos e sintáticos) são inadequados para dar conta da enunciação completa, seja ela uma palavra, uma frase ou uma sequência de frases.” Portanto, na ótica de Bakhtin, a avaliação somativa do texto, do enunciado completo, não é possível. Só é possível avaliar e pontuar, classificar os aspectos lingüísticos do enunciado. De fato, tomando, por exemplo, as competências do ENEM, veremos que só duas delas podem ser objetivamente avaliadas: I - Demonstrar domínio da norma culta da língua escrita; IV - Demonstrar conhecimento dos mecanismos lingüísticos necessários para a construção da argumentação. As demais apresentam alto índice de subjetividade.

Bakhtin, no entanto, relativiza seus pontos de vista quando afirma:

O critério de correção só se aplica à enunciação em situações anormais ou particulares (por exemplo, no estudo de uma língua estrangeira). Em condições normais, o critério de correção lingüística cede lugar ao critério puramente ideológico: importa-nos menos a correção da enunciação do que seu valor de verdade ou de mentira, seu caráter poético ou vulgar, etc. (BAKHTIN, 2006, p. 97).

Quanto a isso, não me parece que a avaliação de redações seja uma situação anormal.

Ainda, Bakhtin (2006, p. 156), referindo-se à linguagem judicial, afirma que ela “intrinsecamente assume uma discrepância nítida entre o subjetivismo verbal das partes num processo e a objetividade do julgamento. A retórica política é análoga.” Refletindo sobre a tarefa do avaliador de redações, vê-se que este

acumula os papéis de leitor-avaliador e de juiz, uma vez que suas decisões, à maneira das decisões dos juízes de direito, vão afetar a vida das pessoas: algumas serão condenadas, outras serão absolvidas. Na avaliação de redações não temos também uma discrepância nítida entre o subjetivismo verbal e a objetividade da avaliação?

Continuando, diz Bakhtin (2006, p. 156)

Quanto mais forte for o sentimento de eminência hierárquica na enunciação de outrem, mais claramente definidas serão as suas fronteiras, e menos acessível será ela à penetração por tendências exteriores de réplica e comentário.

Portanto, a palavra do juiz, que se encontra no mais alto grau de eminência na hierarquia judicial, é a palavra da autoridade, e, como tal, fechada a quaisquer réplicas de seus interlocutores. Considerando essas palavras de Bakhtin, podemos entender que a palavra do candidato-enunciador, que se encontra no mais baixo grau na hierarquia da avaliação, vai se deparar com a palavra da autoridade, o avaliador, ou melhor, com a nota do avaliador, fechada, de fato, para qualquer réplica, sem direito a defesa e ao contraditório.

Como se pôde ver nos trechos de Bakhtin, parece existir incompatibilidade entre suas ideias e a avaliação somativa de redações.

Em suma, a avaliação somativa de redações apresenta sérios problemas decorrentes da própria natureza da linguagem. É o tipo de avaliação na qual a impressão do avaliador determina a pontuação, na qual os critérios de sucesso não podem ser totalmente previstos.

9 OUTRAS SUBJETIVIDADES

Além da subjetividade intrínseca ao leitor-avaliador, decorrentes da própria natureza da linguagem, outros fatores também subjetivos interferem na confiabilidade da avaliação somativa, já apontados e discutidos por diversos especialistas. Há problemas na avaliação que se devem somente ao perfil do avaliador, à sua confiabilidade, à sua tolerância ao cansaço, à sua interação com outros avaliadores e a seus erros na avaliação.

Entende-se por confiabilidade do avaliador sua consistência na avaliação de textos de mesma qualidade, em qualquer ponto da escala de avaliação.

A avaliação de redações é um processo cansativo e psicologicamente debilitante. Nem todos suportam bem a situação. Segundo Wolcott (1998), em situação de cansaço, a atenção do avaliador começa a fugir, levando-o a

sobrepontuar ou penalizar um texto. A tendência, nesse caso, é sacrificar a precisão pela rapidez.

Há problemas também relativos à interação entre os avaliadores: estudos de Barrit et al (1986) e Freedman (1984) mostram que a avaliação pode ser afetada pelo fato de os avaliadores não partilharem preferências, treinamento e *background*; os textos variam muito, uns têm excelente conteúdo, mas apresentam muitos problemas estruturais, gramaticais e lexicais, outros têm poucos problemas gramaticais, mas são fracos quanto ao conteúdo. Isso pode levar a discrepância entre os avaliadores: um pode valorizar mais o conteúdo e penalizar menos os erros gramaticais ou o contrário; estudos de Erickson (2001) Charney (1984) e Stach (1987) mostram que as pontuações feitas por vários avaliadores não coincidem nem quanto a erros gramaticais, vocabulário, complexidade sintática ou mesmo aparência do texto; estudo de Huot (1993) aponta para o fato de que avaliadores não treinados se apoiam mais em aspectos estruturais do que em aspectos comunicativos.

Além desses problemas já observados, há também erros dos avaliadores, que podem ocorrer devido a fatores afetivos e cognitivos (WOLLCOTT, 1998) : erros halo – quando o avaliador lê um ensaio de qualidade média depois de ter lido um, ou mais de um, de excelente qualidade, a tendência é penalizar mais do que devido o texto mediano seguinte ou vice-versa; erros por leniência – quando o avaliador tende a ser muito clemente ou muito rigoroso; erros por restrição de calibragem – quando o avaliador sofre da tendência a evitar notas muito altas ou muito baixas. Há ainda erros por displicência; erros por pressa; erros por incompetência.

Como se pôde ver, além da subjetividade própria da natureza da linguagem, são inúmeros os fatores e variáveis que interferem na avaliação somativa de redações.

10 PESQUISA EMPÍRICA

Buscando evidências que confirmem a influência da subjetividade na correção de redações em processos somativos, desenvolvi uma pesquisa de cunho etnográfico que apresento resumidamente a seguir. A pesquisa teve como *corpus* um conjunto de 700 provas de redação feitas pelos alunos de um curso pré-vestibular, em um vestibular simulado realizado em julho de 2011. Cópias dessas redações me foram cedidas para a pesquisa.

A pesquisa constituiu-se da correção dessas provas, realizada por esta pesquisadora e um grupo de avaliadores convidados, docentes e alunos de

pós-graduação. Foi utilizado o modelo de avaliação do ENEM 2010, e buscou-se reproduzir, quando possível, a sistemática usada neste exame: as provas foram escaneadas, não continham identificação dos candidatos e foram corrigidas *on-line*. A correção foi feita às cegas, por duplas de avaliadores. Coube à coordenadora – esta pesquisadora – enviar, pela *internet*, aos avaliadores as provas a serem corrigidas e as receber de volta, estabelecendo prazos e demais instruções. A pesquisa foi desenvolvida em três fases:

PRIMEIRA FASE:

1ª tarefa: Reunião com a equipe para um treinamento e discussão dos critérios, usando o material utilizado no ENEM para esse fim.

2ª tarefa

Corpus: Uma amostra de 120 provas que foram corrigidas num período de 8 horas consecutivas por uma dupla de pesquisadores. Os membros da dupla não tiveram conhecimento de quem eram seus pares.

Objetivos:

1º Objetivo: Buscar evidências da influência do cansaço como fator que gera maior discrepância entre as notas atribuídas pelas duplas. O primeiro arquivo de 40 provas foi enviado às duplas no início da manhã, o segundo por volta de onze horas e o terceiro por volta de 15 horas.

Método: Análise quantitativa do corpus

Achados:

1º arquivo: **40%** de discrepâncias*

2º arquivo: **10%** de discrepâncias

3º arquivo: **50%** de discrepâncias

***Discrepância:** qualquer diferença entre as duas notas atribuídas pela dupla de avaliação.

Comentário: Verifiquei que o arquivo corrigido no fim da tarde foi onde se deu maior número de discrepâncias. O arquivo enviado no meio do dia apresentou o menor número de discrepâncias. No primeiro arquivo, corrigido na parte da manhã também houve muita discrepância. Esses dados evidenciam que o cansaço afeta a correção. Entretanto, o melhor resultado não foi o da parte da manhã, em que se esperava que os avaliadores ainda não estivessem cansados. Atribuo isso ao fato de que os avaliadores demoram algum tempo para entrar no clima da correção, o que ocorre depois de algumas horas de trabalho.

2º Objetivo: Buscar evidências sobre a ocorrência do efeito halo.

Método: Análise qualitativa do corpus

Achados: 12 ocorrências (10%) do total de 120 provas

Comentário: Nas 120 provas, encontrei evidências que atribuo ao efeito halo em 10% delas: penalização excessiva de provas medianas corrigidas imediatamente depois de uma prova excelente e o contrário também. Isso evidencia que, ao longo da correção, os parâmetros vão sendo ajustados ao contexto imediato.

3ª Tarefa

Experimento

Corpus: 10 das 120 provas foram reavaliadas pelo mesmo avaliador, no mesmo dia. As duplas não tinham conhecimento de que estavam reavaliando as mesmas provas.

Objetivos:

1º Objetivo: Buscar evidências de que o mesmo avaliador, no mesmo contexto de avaliação, numa segunda avaliação da mesma prova, pode atribuir notas diferentes.

Achados:

Em 8 das 10 provas (**80%**) reavaliadas houve discrepância entre as notas atribuídas na primeira e na segunda avaliação.

2º Objetivo: Dimensionar as discrepâncias encontradas.

Método: Análise quantitativa

Achados:

Diferença de 1 a 10 pontos : 1 prova (**12,5%**)

Diferença de 11 a 30 pontos: 5 provas (**62,5%**)

Diferença de mais de 30 pontos: 2 provas (**25%**)

Comentário: Esse experimento evidenciou que o mesmo avaliador, no mesmo contexto de correção, no mesmo dia, atribui notas às vezes muito diferentes às mesmas provas. Se há discrepâncias nas avaliações realizadas por um mesmo avaliador, pode-se supor que o problema com avaliadores diferentes é ainda mais grave.

SEGUNDA FASE: Desenvolvida cerca de dois meses após o encerramento do processo de avaliação.

Corpus: amostra de 200 redações do vestibular simulado, colhida aleatoriamente de um universo de cerca de 700 provas.

Método: Análise qualitativa /quantitativa do corpus

Objetivos:

1º Objetivo: Verificar se as avaliações se pautaram adequadamente nos critérios estabelecidos pelo ENEM e se houve erros grosseiros na avaliação.

Método: Análise qualitativa

Achados:

As avaliações, pelo menos formalmente, se pautaram nos critérios estabelecidos. Não houve evidência de erros grosseiros na avaliação.

2º Objetivo: Verificar o percentual das provas em que ocorreram discrepâncias entre as notas atribuídas pelos membros da dupla.

Método: Análise quantitativa do corpus

Achados:

82% das provas (164) apresentaram discrepâncias.

Comentário: Como se pode ver, é muito alta a porcentagem de discrepâncias. Isso leva a crer que a regra é a presença de discrepância e a exceção é a atribuição da mesma nota, o que confirma a presença da subjetividade na avaliação. .

3º Objetivo: Dimensionar as discrepâncias nas notas atribuídas pelas duplas, identificadas na amostra.

Método: Análise quantitativa do corpus

Achados: (em 164 provas):

Diferença de 1 a 10 pontos : 33 provas (**20%**)

Diferença de 11 a 30 pontos: 82 provas (**50%**)

Diferença de mais de 30 pontos: 49 provas (**30%**)

Comentário: Verifiquei que as maiores diferenças entre as notas dos dois avaliadores da dupla ficam na faixa de 20 a 50 pontos, com 50% das ocorrências, e as menores diferenças são de 10 pontos, com 20% das ocorrências. Como se pode ver, além de haver muitas discrepâncias, elas são também quantitativamente muito significativas.

4º Objetivo: Verificar em que faixa de pontuação ocorreram mais discrepâncias.

Método: Análise quantitativa do corpus

Achados:

Nota de 0 a 16 pontos : 33 provas (**20%**)

Nota de 17 a 32 pontos: 106 provas (**65%**)

Nota de 33 a 50 pontos: 25 provas (**15%**)

Comentário: Verifiquei que 65% das discrepâncias se situam na faixa de pontuação de 30 a 80 pontos. De fato, é mais fácil discriminar as redações muito boas e as muito fracas. As da faixa intermediária são mais problemáticas.

5º Objetivo: Identificar as discrepâncias por competências e verificar em qual(is) competência(s) houve mais discrepâncias e em que proporção.

Método: Análise quantitativa e qualitativa do corpus

Achados:

Competências do ENEM

I Demonstrar domínio da norma culta da língua escrita - 10% do total de discrepâncias verificadas na pontuação das competências

II Compreender a proposta de redação e aplicar conceitos das várias áreas do conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo - 20% do total de discrepâncias verificadas na pontuação das competências

III Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista - 35% do total de discrepâncias verificadas na pontuação das competências.

IV Demonstrar conhecimento dos mecanismos lingüísticos necessários para a construção da argumentação - 10% do total de discrepâncias verificadas na pontuação das competências

V Elaborar proposta de intervenção para o problema abordado, demonstrando respeito aos direitos humanos - 25% do total de discrepâncias verificadas na pontuação das competências

Comentário: Como era de se esperar, houve maior número de discrepância nas competências que avaliam aspectos relativos à enunciação e menor número nas competências que avaliam aspectos lingüísticos e estruturais do texto, que se prestam mais a uma avaliação positivista. Esse resultado corrobora, como se pode verificar, a posição de Bakhtin.

6º Objetivo: Verificar se há evidências de avaliadores mais tolerantes em relação a erros gramaticais.

Método: Análise quantitativa e qualitativa do corpus

Achados:

91% dos avaliadores desconsideraram alguns problemas gramaticais.

Comentário: Verifiquei que é alta a porcentagem (91%) de omissão de correção de erros quanto ao uso da norma culta. Isso também era de se esperar. Grande maioria dos profissionais da área de Letras na universidade não valoriza a correção gramatical.

TERCEIRA FASE: Desenvolvida três meses após o encerramento do processo de avaliação do Vestibular simulado.

Experimento

Sujeitos: três duplas de avaliação que participaram da pesquisa

Perfil das duplas:

1 dupla de avaliadores inexperientes

1 dupla de avaliadores experientes e titulados

1 dupla de avaliadores experientes e não titulados

Tarefa 1:

Reavaliação de uma amostra aleatória de 10 redações avaliadas por eles próprios na correção original.

Objetivos:

1º Objetivo: Verificar se ocorre mais discrepâncias em avaliações feitas por avaliadores inexperientes e se o fato de o avaliador ser titulado garante avaliação mais confiável.

Achados:

Total de discrepâncias entre as avaliações da dupla: 10 ocorrências (num total de 10 redações – 100%)

Discrepâncias por duplas

dupla de avaliadores inexperientes: 20% do total de discrepâncias registradas

dupla de avaliadores experientes e titulados: 50% do total de discrepâncias registradas

dupla de avaliadores experientes e não titulados: 30% do total de discrepâncias registradas

Comentário: Como se pode ver a experiência é importante. O pior resultado foi de titulados e inexperientes, o que atribuo ao fato de estes serem mais avessos a seguir os critérios estabelecidos.

2º Objetivo: Verificar se em uma situação não tensa, com volume de trabalho reduzido, 10 provas, as discrepâncias são em menor número, comparadas com as discrepâncias registradas na correção original.

Achados:

8 dos 10 avaliadores (**80%**) não confirmaram as notas atribuídas por eles na avaliação original.

3º Objetivo: Dimensionar as discrepâncias encontradas.

Achados:

As discrepâncias entre as notas originais e as notas deste experimento tiveram uma diferença que variou de 10 a 25 pontos.

4º Objetivo: Verificar se houve discrepâncias entre as notas atribuídas pelos membros das duplas na correção deste experimento.

Achados:

Houve 5 (50%) de discrepâncias entre as notas atribuídas pelos membros das 5 duplas na avaliação feita neste experimento.

Comentário: Como se pode ver a situação não tensa, com uma tarefa pequena, não elimina as discrepâncias entre os membros das duplas.

11 CONSIDERAÇÕES SOBRE A PESQUISA EMPÍRICA

Apesar das limitações e incompletude desta pesquisa, alguns achados foram importantes, pois confirmam a forte presença de subjetividade nesse tipo de avaliação, corroborando os pontos de vista teóricos de Bahktin e de outros especialistas em avaliação. xzzPesquisas mais elaboradas podem e devem ser realizadas, e divulgadas, mas não creio que se possa provar que tal tipo de avaliação seja isenta de subjetividade.

12 CONSIDERAÇÕES FINAIS

Com base nas considerações, argumentos e evidências apresentados, concluo que a avaliação somativa da prova de redação em vestibulares e concursos não garante justiça a todos os candidatos e que a injustiça é potencializada nos concursos de grande porte, que precisam contar com grande número de avaliadores.

Sou, portanto, levada a me posicionar a favor da eliminação desse tipo de prova, por estar convicta de que não vale a pena investir em tentativas de torná-la objetiva, já que eventuais alterações nos modelos não serão capazes de eliminar a subjetividade que lhe é inerente. No meu entender, considerando a inegável importância do saber redigir, sou favorável a um maior investimento na melhoria do ensino da redação, em todos os níveis. Este saber deve ser efetivamente valorizado por todos, não pelo fato de ser cobrado nos vestibulares e concursos, mas pelo que representa para a formação de todos.

REFERÊNCIAS

BAKHTIN, Mikhail M. **Problemas da poética de Dostoiévski**. Trad. Paulo Bezerra. Rio de Janeiro: Forense Universitária, 1981.

_____. **Marxismo e filosofia da linguagem**. São Paulo: Hucitec, 2006.

_____. **Estética da criação verbal**. São Paulo: Martins Fontes, 2010.

BARRIT, Louis et al. Researching practice evaluating assessment essays. **College Composition and Communication**, Illinois, n. 37, P. 315-327, 1986.

CHARNEY, Davida. The validity of using holistic scoring to evaluate writing. a critical overview. **Research on the Teaching of English**, Urbana, Il, n. 18, p. 65-81, 1984,

DAVIS, Barbara Gross et al. **The evaluation of composition instruction**. New York; London: Teachers College Press, 1987.

DUBOIS, Philip H. **A history of psychological testing**. Boston: Allyn and Bacon, 1970.

ERICKSON, H. L. **Stirring the head, heart, and soul: Redefining curriculum and instruction**. 2. ed. Thousand Oaks, CA: Corwin Press, 2001.

FREEDMAN, Sarah W. Influences on evaluators of expository essays: beyond the text. **Research in the Teaching of English**, Urbana, Il, n. 15, p. 245-55, 1984.

HUOT, Brian. **Validating holistic scoring for writing assessment: theoretical and empirical foundations**. Cresskill, NJ: Hampton, 1993.

RANK, Otto. **The myth of the birth of the hero and other writings**. New York: Random House (Vintage books), 1974.

STACH, Charles L. **The component parts of general impressions: predicting holistic Scores in College level essays**. Dissertação (Mestrado) - University microfilms no. DA8722706, 1987.

WOLLCOTT, Willa. **An overview of writing assessment**. Urbana, Il.: National Council of Teachers of English, 1998.