

ANE CIBELE PALMA

**A BUSCA PELA MINIMIZAÇÃO DOS EFEITOS DA SUBJETIVIDADE
NA AVALIAÇÃO DA ESCRITA EM LÍNGUA ESTRANGEIRA**

CURITIBA

2005

ANE CIBELE PALMA

**A BUSCA PELA MINIMIZAÇÃO DOS EFEITOS DA SUBJETIVIDADE
NA AVALIAÇÃO DA ESCRITA EM LÍNGUA ESTRANGEIRA**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Estudos Lingüísticos, pelo Programa de Pós-Graduação em Letras, Setor de Ciências Humanas, da Universidade Federal do Paraná.

Orientador: Prof. José Erasmo Gruginski.

CURITIBA

2005

Ao meu filho, Joaquim, que acompanhou o processo de pesquisa e escrita desta dissertação desde a gestação.

AGRADECIMENTOS

Agradeço a todas as pessoas que participaram direta ou indiretamente deste trabalho, me apoiando com palavras de incentivo e com seu carinho.

A Deus, por me dar força e saúde para enfrentar os desafios dos estudos.

Aos meus pais, que sempre me incentivaram a trabalhar mais quando pensei em desistir.

Às minhas amigas e colegas de trabalho, Candida, Maria Rute e Mirian, que passaram boa parte do seu tempo corrigindo redações, respondendo questionários e discutindo suas avaliações.

Aos demais colegas de trabalho, do Inter Americano e da PUCPR, que sempre me incentivaram, apoiaram e me substituíram quando necessitei de tempo extra para trabalhar.

Aos meus queridos alunos, por autorizarem o uso de suas composições neste estudo.

Ao colegiado do Programa de Pós-Graduação em Letras da UFPR, pela oportunidade de aperfeiçoamento que me proporcionaram.

Ao Odair, por todo o seu apoio.

E, por fim, agradeço ao meu orientador, professor Erasmo, por me ouvir, responder os meus questionamentos, pela sua tranquilidade, paciência e por acreditar que este trabalho seria possível.

SUMÁRIO

AGRADECIMENTOS	iv
LISTA DE QUADROS	vi
LISTA DE TABELAS	vii
RESUMO	viii
ABSTRACT	ix
INTRODUÇÃO	1
1. REVISÃO DE LITERATURA	7
1.1 AVALIAÇÃO	7
1.1.1 Características Gerais	7
1.1.2 Princípios da Avaliação.....	11
1.1.3 Tipos de Testes e seus Objetivos.....	18
1.2 AVALIAÇÃO DA ESCRITA.....	23
1.3 AVALIAÇÃO DE DESEMPENHO	32
1.4 O PROCESSO DE AVALIAÇÃO E MENSURAÇÃO DE TESTES	38
1.4.1 As Fontes de Variabilidade na Avaliação da Escrita.....	41
1.4.2 A Mensuração Multifacetada	45
2. METODOLOGIA	54
3. ANÁLISE DOS RESULTADOS.....	59
3.1 CRITÉRIOS E OUTROS FATORES DE INFLUÊNCIA NAS NOTAS	62
3.2 DISCUSSÃO E CALIBRAÇÃO DOS CRITÉRIOS	64
3.3 EFEITO DA DIFICULDADE DO TEMA ESCOLHIDO	73
3.4 FACETAS	76
CONCLUSÃO	78
REFERÊNCIA BIBLIOGRÁFICA	81
APÊNDICES	85
ANEXOS	92

LISTA DE QUADROS

QUADRO 1 – RELAÇÃO DOS GÊNEROS DA ESCRITA	24
QUADRO 2 – REPRESENTAÇÃO ESQUEMÁTICA DA AVALIAÇÃO TRADICIONAL E DE DESEMPENHO	33
QUADRO 3 – HABILIDADE DE CANDIDATOS MEDIDA POR PADRÕES E CORRETORES DIFERENTES	39
QUADRO 4 – MAPA DAS FACETAS SOBRE OS DADOS DO “ACCESS: TEST”	48
QUADRO 5 – DISTRIBUIÇÃO DAS NOTAS ATRIBUÍDAS PELOS CORRETORES, SEGUNDO VALOR NUMÉRICO DA ESCALA DE NOTAS	60
QUADRO 6 – DEMONSTRATIVO DO NÚMERO DE VEZES QUE CADA CORRETOR ATRIBUIU AS NOTAS DE 1 A 6 E SUA REPRESENTAÇÃO EM PERCENTAGEM.....	61
QUADRO 7 – RESULTADOS DAS CORREÇÕES EFETUADAS PELOS CORRETORES A e B e AS ALTERAÇÕES PÓS-DISSCUSSÃO	66
QUADRO 8 – GRAU DE DIVERGÊNCIA DAS NOTAS ATRIBUÍDAS POR A e B.....	67
QUADRO 9 – RESULTADOS DAS CORREÇÕES DAS COMPOSIÇÕES NÚMERO 1 EFETUADAS PELOS CORRETORES A, B e C e A MÉDIA DE NOTAS DE CADA CORRETOR	71
QUADRO 10 – DEMONSTRATIVO DA MÉDIA DAS NOTAS ATRIBUÍDAS ÀS COMPOSIÇÕES 1 E 2 PELOS CORRETORES A,B e C.....	73
QUADRO 11 – DEMONSTRATIVO DA MÉDIA ENTRE A NOTAS ATRIBUÍDAS PELOS CORRETORES A, B e C PARA CADA ALUNO NAS COMPOSIÇÕES 1 E 2.....	74

LISTA DE TABELAS

TABELA 1 – MATRIZ DE DADOS.....	46
TABELA 2 – RELATÓRIO SOBRE AS MEDIDAS DAS HABILIDADES DOS CANDIDATOS NO “ACCESS: TEST”	50
TABELA 3 – RELATÓRIO SOBRE A DIFICULDADE ESTIMADA DOS ITENS DO “ACCESS: TEST”	51
TABELA 4 – RELATÓRIO SOBRE A RIGOROSIDADE ESTIMADA DOS CORRETORES DO “ACCESS: TEST”	52

RESUMO

Esta dissertação, fundamentada em pesquisa teórica e empírica, tem o objetivo de buscar alternativas que possam minimizar os efeitos da subjetividade na avaliação da escrita em língua estrangeira. Seu referencial teórico tem por base uma revisão da literatura que compreende a teoria da avaliação, suas características gerais, seus princípios, os tipos de testes e seus objetivos, cujo propósito é o de fundamentar os aspectos levantados na literatura sobre a avaliação da escrita, sua caracterização como avaliação de desempenho e a explanação do processo de mensuração multifacetada do modelo de Rasch. A análise empírica centra-se nas correções de trinta e nove composições efetuadas por três corretores distintos, com o objetivo de evidenciar as divergências dos resultados, considerando os diversos fatores que envolvem os corretores, a habilidade dos alunos e o grau de dificuldade do instrumento de avaliação. A partir do embasamento teórico da literatura pesquisada e do resultado das correções, este estudo demonstra algumas medidas que podem ser adotadas para reduzir a subjetividade na avaliação, como, por exemplo, a conferência entre corretores, o cálculo de médias entre as notas atribuídas e a análise das facetas que interferem nos resultados, por meio de um *software* denominado FACETS.

Palavras-chave: lingüística aplicada; avaliação da escrita; critérios de correção; efeitos da subjetividade.

ABSTRACT

This dissertation is based on theoretical and empirical research with the objective of searching for alternatives which may come to minimize the effects of subjectivity in the assessment of writing in a foreign language. Its theoretical references comprise a literature review on the theory of assessment, its general characteristics, its principles, the kinds of evaluation and their purposes, aiming at providing theoretical underpinnings for the issues raised about the evaluation of writing, its characteristics as performance assessment and the explanation about the process of the Rasch Model multi-faceted measurement. The empirical analysis focuses on the corrections of thirty-nine compositions by three different raters with the purpose of achieving divergent results, considering various factors which involve the raters, the ability of the writers and the level of difficulty of the evaluation instrument. From the theoretical foundations in the literature review and the results of the ratings, this research demonstrates some measures which may be adopted in order to reduce subjectivity in assessment, such as, for instance, conferencing between raters, the calculation of average scores amongst the grades attributed and the analysis of the facets which interfere in the results through a software program named FACETS.

Key-words: applied linguistics; writing assessment; evaluation criteria; effects of subjectivity.

INTRODUÇÃO

A avaliação em todos os setores do ensino tem mudado rapidamente. Testes de desempenho nos quais os alunos têm que demonstrar domínio prático dos conteúdos vêm substituindo instrumentos de avaliação mais tradicionais, como os testes objetivos. Na percepção de McNamara (1996), estas mudanças na avaliação estão ocorrendo em função das exigências do mercado de trabalho por profissionais mais habilidosos e flexíveis, que demonstrem no seu desempenho profissional resultados de uma aprendizagem prática e relevante, principalmente em relação às línguas estrangeiras, hoje uma habilidade determinante para os profissionais do mundo globalizado.

O ensino do inglês como língua estrangeira passou por diversas transformações ao longo dos anos e as abordagens didáticas foram modificadas para que melhor se adaptassem às necessidades dos alunos e dos professores. A pesquisa em busca de novas metodologias e melhores técnicas de ensino favoreceu a aprendizagem e a interação entre professores e alunos; entretanto, a quantificação do conhecimento dos alunos é um aspecto ainda insatisfatório para o professor, que busca constantemente formas válidas e confiáveis de avaliar o seu desempenho.

As técnicas de ensino e metodologias podem ser testadas a cada aula ou no contato com os alunos; todavia, a avaliação dos resultados e da eficácia da aprendizagem, independente da metodologia utilizada, só se torna possível com medidas mais sistemáticas por meio de instrumentos confiáveis. Para obter tais medidas e tirar conclusões sobre a competência dos aprendizes, a avaliação é o melhor caminho, seja esta realizada por meio de testes, seja mediante desempenhos demonstrados de forma processual.

Muitos professores e alunos confundem o termo “avaliação” com “teste”, que, segundo H. Douglas Brown (1994), são atividades complementares, embora diferentes. A avaliação é um processo que permite inferir sobre o conhecimento dos

alunos, e neste processo são utilizados vários instrumentos, entre eles os testes propriamente ditos. É possível avaliar o desempenho de um aluno pelo seu engajamento, pela sua participação em sala de aula, em atividades informais ou por meio de testes formais, que podem ter diversos formatos. De acordo com esse mesmo autor, que faz uso do termo *teste* no sentido global de um conjunto de atividades para que se infira sobre a competência de um aluno, ou seja, como *avaliação*, os testes (a avaliação) têm por objetivo a obtenção de medidas. Os testes informais, conforme Brown, são difíceis de quantificar devido ao julgamento intuitivo por parte dos professores. Em contrapartida, a testagem formal, que utiliza técnicas de avaliação planejadas, fornece resultados mais precisos e uma quantificação mais objetiva.

Ainda segundo H. Douglas Brown (1994), os testes medem a habilidade ou o conhecimento de uma pessoa, em uma determinada área ou conteúdo. É preciso identificar cuidadosamente o aluno e o seu nível de conhecimento em relação ao conteúdo ensinado e testado, para que seja possível inferir sobre sua competência demonstrada no desempenho em um teste ou durante um processo avaliativo. O cuidado com a avaliação deve respeitar dois princípios básicos para que se tenham resultados objetivos, a confiabilidade e a validade. Em um teste confiável e válido devem ser considerados diversos aspectos, tais como o momento do planejamento da avaliação, a sua aplicação, a correção, a análise dos resultados e o feedback aos alunos, conforme será discutido na revisão de literatura.

O histórico da avaliação da língua estrangeira, de acordo com H. Douglas Brown (1994), divide-se em três fases cronológicas. A primeira caracteriza-se por um período pré-científico anterior à década de 1950, no qual a avaliação seguia os mesmos princípios adotados nas áreas humanísticas ou de ciências sociais, devido à falta de pesquisas nesta área. O período seguinte, definido como “psicométrico-estruturalista”, teve início em aproximadamente 1950, e se estendeu até o final de 1960. Por se tratar de um período sob influência da análise contrastiva e uma época em que a lingüística estrutural e a psicologia do behaviorismo buscavam explicar o caráter científico do ensino de idiomas, os testes conseqüentemente enfocavam aspectos lingüísticos específicos, tais

como questões fonológicas, gramaticais, contrastes lexicais, entre outros, sempre em questões isoladas; essa abordagem é denominada “discrete point” ou “pontos discretos”. A terceira e última fase cronológica caracteriza-se por um período “sociológico-integrativo”, que data do final de 1960 à atualidade. As pesquisas provenientes da insatisfação com o estruturalismo e o behaviorismo levaram à competência comunicativa, aos estudos em torno do contexto de uma linguagem e à “aceitação de que um evento comunicativo vai muito além da soma dos seus elementos lingüísticos” (CLARK 1983:432). Esse período é conhecido por “integrado” (“integrative testing”) e sua ênfase recai na avaliação das competências e dos desempenhos que integram conhecimentos declarativos e procedimentais.

Bachman (1991) distingue a avaliação da habilidade lingüística comunicativa das outras formas históricas de avaliação por ter na sua construção quatro características determinantes: a) a lacuna de informações, que requer o uso de fontes variadas de insumo; b) a dependência entre as diversas questões do teste; c) a integração das questões e do conteúdo do teste com um determinado domínio discursivo; d) a tentativa de medir uma maior variedade de habilidades lingüísticas, como o conhecimento de coesão, funções, apropriação sociolingüística, entre outros.

Existem várias razões para que se desenvolva um entendimento crítico sobre a avaliação e seus princípios, especialmente para aqueles que trabalham com o ensino de uma língua estrangeira. McNamara (2000) agrega aos resultados dos testes o papel fundamental de classificar indivíduos em momentos importantes na sua educação, profissão, entre outras situações classificatórias, que interferem expressivamente em suas histórias. Ademais, profissionais que trabalham com o ensino e tomam decisões baseadas em avaliações e testes devem ter conhecimento suficiente do assunto para que possam fazer escolhas conscientes sobre o seu uso e reconhecer o seu impacto nos aprendizes.

O estudo desenvolvido nesta dissertação enfoca a habilidade escrita, precisamente a avaliação da escrita em língua estrangeira. É fato que a escrita é uma atividade pessoal e, ao escrever, segundo Hamp-Lyons (1990), o indivíduo utiliza suas idéias e as transforma em tópicos “auto-iniciados”, ou seja, o escritor recorre ao seu

conhecimento do mundo e a processos mentais complexos para desenvolver suas idéias. De acordo com Hillocks (1987), para escrever bem os alunos precisam recorrer a quatro tipos de conhecimento: a) de conteúdo; b) procedimental, para organizar o conteúdo; c) conhecimento das convenções da escrita e, por último, d) conhecimento procedimental para aplicar os três primeiros conhecimentos ao escrever. Esses quatro conhecimentos aplicam-se ao processo da escrita em geral, quer o produto seja escrito na língua materna, quer em uma língua estrangeira. Percebe-se, a partir das reflexões de Hamp-Lyons (1990) e Hillocks (1987), a complexidade da operacionalização da habilidade escrita e, conseqüentemente, a complexidade que deve ser considerada ao avaliar esta habilidade.

A avaliação de desempenho, tal qual a da habilidade escrita, envolve necessariamente julgamentos subjetivos em função da complexidade humana e das diferenças na sua capacidade produtiva, tanto para a produção de um texto quanto para sua avaliação. Em face dessa situação de variabilidade na avaliação, como garantir uma correção justa e que considere as diversas facetas que tornam a nota bruta pouco confiável? Aspectos como as características dos avaliadores, dos alunos, da tarefa em si, entre outros, devem ser considerados para que se analise todo o contexto de uma avaliação.

De acordo com McNamara (1996), uma das respostas a esses questionamentos podem ser obtidas por meio dos conceitos e procedimentos da mensuração multifacetada, uma das fórmulas da teoria de resposta ao item do modelo Rasch e método de mensuração que se adapta às avaliações de desempenho. Os avanços trazidos por esta teoria enfocam a natureza e a diversidade dos corretores, a possibilidade de fazer novas pesquisas na área de avaliação em testes de desempenho, a compensação dos efeitos na correção causados pelos corretores e outras variáveis que impeçam uma avaliação adequada da capacidade dos alunos.

Os aspectos que influenciam o contexto da avaliação podem ser considerados facetas, e as informações sobre o impacto destas facetas, ou uma combinação delas, podem ser obtidas pela mensuração multifacetada. Em outras palavras, uma vez que se

tenha informação sobre tais facetas e seu efeito na nota bruta de um texto, torna-se possível medir a habilidade do aluno a partir do ajuste da nota bruta com o resultado da influência das facetas sobre a mesma.

O interesse pelo tema deste estudo, a busca da minimização da subjetividade na avaliação da escrita, surgiu da minha percepção de que há uma falta de padronização nas notas auferidas às composições de alunos de inglês em contextos escolares diversificados. Minha experiência profissional com o ensino da língua inglesa divide-se em instituições com propósitos diferentes, fator que atuou positivamente para ampliar meu contato com a avaliação da escrita. Lecionei por vários anos em um centro binacional, onde tive turmas praticamente homogêneas, considerando que nessa instituição o curso regular divide os alunos em 12 módulos/níveis, do básico ao avançado. Concomitantemente, nos últimos cinco anos ministrei aulas de inglês como língua estrangeira para o curso de Letras da Pontifícia Universidade Católica do Paraná (PUCPR), onde os alunos ingressantes têm níveis heterogêneos, pois as turmas são divididas em apenas três níveis, básico, intermediário e avançado. Alguns desses alunos esperam aprender inglês na universidade, enquanto outros já são professores de inglês e buscam aperfeiçoamento no idioma e técnicas de metodologia e didática.

Em ambas as instituições, tanto no binacional quanto na universidade, um dos problemas enfrentados sempre foi – e continua sendo – a avaliação da produção escrita dos alunos. No binacional, os professores recebem treinamento sobre como avaliar a escrita e utilizar critérios de correção, fornecidos para todos os professores, que aplicam as mesmas avaliações escritas para os alunos do básico ao avançado. Portanto, existe padronização no planejamento das atividades escritas e nos seus critérios de correção e avaliação. Na referida universidade, trabalhamos com diversos tipos de escrita, mas cada professor define seus critérios de avaliação e o conteúdo que deve ser avaliado, e as notas atribuídas não são discutidas entre colegas que lecionam em outras turmas. Espera-se que, com o treinamento, com turmas homogêneas e com a uniformidade no processo de ensino e avaliação da escrita, como no binacional, quase

não se apresentem problemas relacionados às diversidades na avaliação. No entanto, não é isso que ocorre. Tanto no binacional quanto na universidade, a variabilidade nos resultados das avaliações revela problemas de inconsistência nas correções e conseqüentemente a atribuição de notas inadequadas e pouco confiáveis, tornando a avaliação da habilidade escrita uma tarefa complexa, e em certos contextos injusta. O aspecto subjetivo da avaliação da escrita e o seu impacto nos professores e alunos despertaram meu interesse em pesquisar sobre o assunto.

Em decorrência da variabilidade dos resultados atribuídos às avaliações escritas, proponho neste estudo analisar os resultados de correções efetuadas por três corretores distintos a trinta e nove composições, tendo como objetivo demonstrar a presença da subjetividade a partir dos resultados auferidos e demonstrar medidas que minimizem estas divergências.

As medidas adotadas nesta pesquisa como alternativas na redução das divergências nos resultados de avaliação têm como pressuposto a análise dos resultados de correções efetuadas por dois ou mais corretores, uma vez que as correções múltiplas, segundo Green (2004), diminuem a discrepância entre as notas atribuídas. A primeira análise será feita a partir dos resultados de dois corretores e uma discussão entre eles acerca das notas atribuídas, haverá uma segunda análise com o acréscimo dos resultados de um terceiro corretor e, por fim, uma descrição da análise baseada no modelo multifacetado de Rasch por meio do programa FACETS de Linacre e Wright (1992).

1. REVISÃO DE LITERATURA

1.1 AVALIAÇÃO

1.1.1 Características Gerais

A *avaliação* está inter-relacionada a outros dois conceitos básicos, que são os *testes* e o *ensino*. A dependência entre estes três aspectos na aquisição de uma língua estrangeira ocorre devido ao fato de a prática em sala de aula propiciar a aprendizagem de um idioma, que é constantemente avaliada pelo professor por meio de testes ou observações formais e/ou informais do desempenho dos alunos. A partir da análise dos resultados observados, os professores redirecionam a sua prática para que a aprendizagem flua da melhor forma possível ou redirecionam o acompanhamento da aprendizagem de alguns alunos em específico.

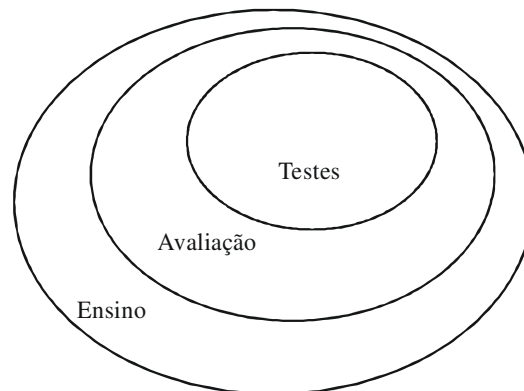
Os testes, de maneira geral, são métodos de mensuração da habilidade, do conhecimento ou do desempenho de um aluno em um determinado domínio ou conteúdo. Para atingir os objetivos de mensuração, o teste deve primeiramente ter o papel de instrumento, ou seja, um conjunto de técnicas, procedimentos e itens, que permita demonstrar o conhecimento de um aprendiz. O teste pode também medir desde competências específicas até conhecimentos gerais de um grupo, ou a habilidade e o desempenho de um aluno.

Um teste mede desempenho, mas o seu resultado geral demonstra a competência do aluno, ou seja, por meio da amostragem obtida nos testes de língua estrangeira, o avaliador pode inferir sobre a competência geral dele. A avaliação, no entanto, é um processo contínuo que compreende todo o acompanhamento que o professor faz do desempenho de seus alunos, por meio de atividades corriqueiras da

sala de aula, bem como de testes formais aplicados no decorrer das aulas. É importante diferenciar os conceitos de *teste* e *avaliação* para maior compreensão da teoria acerca de um processo avaliativo e as partes envolvidas neste processo, entre elas os testes propriamente ditos.

O diagrama de H. Douglas Brown (2004), apresentado a seguir, especifica adequadamente a relação entre testes, avaliação e ensino. Para o autor, os testes são parte integrante, porém não exaustiva, do processo de avaliação, que busca ilustrar a competência dos alunos por diversas formas, com o objetivo de validar ou redimensionar o processo de ensino.

FIGURA 1 – Testes, Avaliação e Ensino



FONTE: BROWN, H. D., 2004 (p.5)

Conforme mencionado, a avaliação pode ser formal ou informal. As avaliações formais são procedimentos desenvolvidos sistematicamente, com técnicas planejadas que proporcionem ao avaliador um retrato das conquistas e dos progressos do avaliando. As avaliações informais, embora também sejam planejadas e organizadas com seriedade pelo avaliador, estão geralmente embutidas nas tarefas de sala de aula, desenvolvidas para demonstrar o desempenho dos alunos, com o objetivo de verificar e qualificar a sua competência, no entanto sem a conotação tradicional da avaliação feita por meio de *provas*.

Outra distinção importante sobre as funções da avaliação, geralmente apontadas na literatura, são as noções de avaliação formativa e somativa. Na avaliação formativa, o professor avalia o aluno durante o processo contínuo de ensino, diariamente, mensalmente, anualmente, fornecendo o *feedback* de seu progresso constantemente de forma que este possa canalizar seus esforços para melhorar sua atuação. É também importante para que o professor possa reformular seu planejamento, caso haja necessidade. A avaliação formativa é composta de simples perguntas de verificação até exercícios escritos e testes formais, cujos resultados são somados para obter uma média final. No que diz respeito à avaliação somativa, conforme o nome sugere, são somadas todas as competências e habilidades desenvolvidas pelo aluno ao término de um período de estudo. O seu objetivo principal é medir ou resumir, por meio de tarefas, testes ou avaliações, o que o aprendiz atingiu em termos de conhecimento. Portanto, o conteúdo da avaliação somativa deve refletir ao máximo os conteúdos trabalhados ou adquiridos durante o percurso dos estudos.

Outra distinção importante para a compreensão da teoria e terminologia sobre a avaliação é a dicotomia entre testes referenciados em normas (*norm-referenced*) e testes referenciados em critérios (*criterion-referenced*). Os testes referenciados em normas têm por objetivo interpretar e classificar os resultados em relação a uma média, a uma nota intermediária, a um desvio padrão, ou um *ranking* percentual. O objetivo desses testes é classificar os candidatos em um *continuum* matemático. Por outro lado, os testes referenciados em critérios são desenvolvidos para prover aos candidatos *feedback*, geralmente por meio de notas e/ou, conceitos sobre os objetivos do curso, da aula ou de um conteúdo específico. Os objetivos dos testes referenciados em critérios são mais adequados ao contexto do estudo de uma língua estrangeira, cujo enfoque está no melhor desenvolvimento da habilidade lingüística dos alunos, sem a necessidade de classificá-los num *continuum* numérico.

As tendências na avaliação da linguagem, historicamente, vêm seguindo as metodologias de ensino; por exemplo, na década de 1950, a era do behaviorismo, cujo enfoque era a análise contrastiva, a testagem lingüística deu ênfase a elementos

fonológicos, gramaticais e a contrastes léxicos entre a língua materna e a língua estrangeira em estudo. Já entre 1970 e 1980, as teorias comunicativas da linguagem trouxeram uma visão mais integrada dos testes, e especialistas como Clark (1983, p. 432) discutiram que “o conjunto do evento comunicativo era consideravelmente mais importante do que a soma dos seus elementos lingüísticos”. Esta trajetória marcou a transição dos testes de pontos discretos, cujo objetivo era quebrar a linguagem em componentes que pudessem ser testados isoladamente, para testes integrados, cuja descendência da competência lingüística não permitia a separação de habilidades lingüísticas interativas e interdependentes, nem mesmo para o evento de testes.

A visão sociointeracionista do ensino das línguas estrangeiras teve um papel fundamental na sua testagem, pois, ao enfatizar a competência estratégica e a habilidade de sua aplicação para a compensação de falhas na linguagem, impôs à avaliação de uma língua estrangeira, testes que correspondessem a esta necessidade e demonstrassem o desempenho de candidatos no uso da língua alvo, em situações semelhantes às reais. Assim, mediante a avaliação tornou-se possível fazer inferências sobre o provável desempenho de um aluno em uma atividade real, pela observação da sua competência em uma situação similar, que imitasse o comportamento do mundo real. Contudo, as avaliações tendiam a ser artificiais e pouco espelhavam o uso da língua na vida real.

Os desafios no campo da avaliação fizeram com que pesquisadores voltassem seus olhares para os contextos de uso da língua estrangeira, com o objetivo de tornar as atividades dos testes mais autênticas e genuínas. De acordo com Weir (1990, p. 11):

“Para medir a proficiência lingüística... deve-se considerar: onde, quando, como, com quem, e porque a língua é usada, sobre quais tópicos e com qual efeito.”

Diante disso, o cenário da utilização da língua alvo torna-se fundamental e determinante para a obtenção de resultados mais concretos. Uma alternativa adequada às necessidades de demonstrar o conhecimento dos alunos por meio de situações reais

ou de atividades que simulem a realidade, é a utilização de avaliações baseadas em desempenhos. Uma característica importante desta avaliação, que envolve a produção linguística por parte dos alunos, é a presença de tarefas interativas e significativas. Em tais situações, a avaliação exige dos alunos um comportamento real dentro da habilidade que se queira medir, ou seja, os candidatos podem ser avaliados durante um ato de fala, uma combinação entre compreensão aural e conversação, ou durante uma atividade que envolva compreensão de texto e escrita, por exemplo.

Diferente da avaliação de componentes isolados de uma língua, como a gramática, o vocabulário, ou dos itens de compreensão e recepção, como a compreensão de texto. A avaliação de habilidades de produção, como a escrita e a fala, é mais complexa e depende de um ato comunicativo, escrito ou falado. Por se tratar de uma habilidade produtiva, a escrita é dependente do contexto, do tema e de fatores diversos que venham a interferir na sua produção, e que interferem direta ou indiretamente no desempenho do aluno. Devido a esses aspectos complexos que envolvem a habilidade escrita, as transformações acima mencionadas no campo da avaliação são essenciais.

1.1.2 Princípios da Avaliação

Como saber se um teste é eficaz? A resposta a esta pergunta pode ser obtida por meio de questionamentos sobre aspectos relacionados às limitações administrativas da aplicação do teste, se este é confiável, se ele mede precisamente o que se espera medir. Essas indagações entre outras, trazem à discussão critérios fundamentais para avaliar um teste: praticidade, confiabilidade, validade e autenticidade.

A praticidade de um teste está relacionada a diversos fatores, entre eles custos, pois o teste não pode ser muito dispendioso para uma instituição ou um professor. De acordo com H. D. Brown (2004), o fator tempo é também importante para a praticidade, uma vez que o teste não deve ser muito demorado, tanto para aplicação quanto para a correção. O teste deve ser relativamente fácil de administrar,

assim como ter um procedimento de avaliação/correção específico, eficaz e produtivo; portanto, ele deve ser fácil e eficiente para a correção, sem tomar muito tempo e desgastar o avaliador.

Quanto à confiabilidade, um teste deve ser consistente e fidedigno. Ao ser aplicado a um ou mais alunos, em ocasiões diferentes, o mesmo teste deve produzir resultados similares. De acordo com Mousavi (2002, p. 804), a questão da confiabilidade de um teste está relacionada a um número de fatores que podem interferir ou contribuir para a sua fidedignidade, como alterações no comportamento do aluno, na correção, na administração do teste e a aspectos direcionados ao teste em si.

Existem diferentes aspectos relevantes relacionados à confiabilidade, que dizem respeito ao aluno, ao avaliador, à administração do teste e do teste propriamente dito, que precisam ser analisados para que se torne possível obter uma avaliação mais fidedigna. A confiabilidade relacionada ao aluno pode ser ameaçada em situações como doenças temporárias, fadiga, um dia difícil, ansiedade e outros fatores físicos ou psicológicos que podem desviar a nota “observada” no dia do teste da nota “real” do aluno, comprometendo a qualidade do seu desempenho na avaliação. Entende-se por nota real, o resultado que demonstre a verdadeira habilidade de um aluno, enquanto a nota “observada” pode trazer um resultado alterado por fatores alheios ao conhecimento do aluno. São considerados também neste quesito de confiabilidade, fatores como facilidade para fazer provas ou estratégias para responder uma prova com eficiência. Por exemplo, alunos que estudam regularmente e fazem provas com frequência, podem estar acostumados aos testes e desenvolverem estratégias próprias para responderem às avaliações, com maior facilidade.

A confiabilidade relacionada ao corretor é um aspecto importante da avaliação, pois a falha humana, a subjetividade e os julgamentos pré-concebidos podem interferir no processo da correção de duas formas diferentes, confiabilidade intercorretor e intracoretor. A confiabilidade intercorretor pode ser abalada quando dois ou mais corretores atribuem notas diferentes a um mesmo teste, por razões diversas, tais como a falta de atenção aos critérios, pouca experiência, falta de atenção

no momento da avaliação ou até pré-concepções sobre o desempenho dos alunos. Já a confiabilidade intracorretor envolve a inconsistência do próprio professor/corretor ao atribuir notas entre seus alunos. Geralmente, ocorre entre uma turma, ou entre turmas diferentes, por razões como falta de clareza nos critérios de correção, fadiga, pré-concepções em relação ao “bom” ou “mau” aluno, ou simples falta de cuidado. O número elevado de turmas, aulas e alunos, acarreta em muitas correções e notas para atribuir, agravando o cansaço e o estresse do corretor, fator que pode alterar os resultados atribuídos ao teste. Em relação à escrita, uma possibilidade de minimizar este problema é corrigir mais de uma vez cada composição e iniciar do começo para o fim e vice-versa, em momentos diferentes, eliminando os efeitos do cansaço. Segundo J. D. Brown (1991), “a especificação criteriosa de um instrumento de correção analítico pode aumentar a confiabilidade do corretor”, ou seja, quanto maior o entrosamento do corretor com os critérios de correção, maior a possibilidade de alcançar avaliações justas e padronizadas. A questão da preferência do autor por um critério analítico não é relevante para este momento da discussão.

No que se refere à falta de confiabilidade na administração do teste, é necessário verificar toda a estrutura do local da testagem, as condições de temperatura, poluição sonora interna e externa, sonorização em caso de testes com compreensão aural, qualidade das cópias e da digitação dos materiais, aspectos relacionados à segurança da prova, para que essa não seja de conhecimento prévio de alguns alunos, e medidas de segurança para evitar que os alunos troquem informações sobre o conteúdo durante a aplicação do teste. A iluminação da sala também é importante na qualidade dos resultados, tanto quanto a condição e o conforto das carteiras e cadeiras escolares. É de extrema importância que todos os aspectos administrativos sejam considerados anteriormente à aplicação de um teste, para que não interfiram na qualidade do desempenho dos alunos.

Por último relatamos a interferência da confiabilidade do teste em si, ou seja, aspectos relacionados à elaboração do teste, tais como a ambigüidade em algumas questões, o número de itens testados, testes muito curtos não dão oportunidades suficientes de demonstrar conhecimento sobre o conteúdo, o tempo que se tem para

responder ao teste pode ser insuficiente ou simplesmente causar ansiedade àqueles alunos que não conseguem pensar contra o relógio. A escolha de questões com nível de dificuldade muito elevado também impede que o aluno possa se sentir um pouco menos pressionado, e não permite que ele se familiarize com a situação de avaliação.

Além de considerações com a praticidade e as diversas fontes de confiabilidade de um teste, outro fator de importância em situações de avaliação é a validade, que é o critério mais complexo de um teste eficaz. Para Gronlund (1998, p. 226), validade é “a extensão do quanto inferências feitas a partir dos resultados de uma avaliação são apropriadas, significativas, e úteis em relação ao propósito dessa avaliação”. Isto significa dizer que um teste de compreensão de texto para ser válido deve testar a habilidade de leitura de um aluno e não o seu conhecimento prévio sobre o assunto, sua capacidade de dedução lógica a partir de um texto, ou outra variável de relevância questionável.

A validade, segundo H. D. Brown, pode ser estabelecida a partir de diferentes tipos de evidência, por exemplo, verificar se o desempenho solicitado no teste está adequado à proposta do curso ou ao conteúdo sendo avaliado; verificar também se o teste determina a extensão da compreensão do aluno relacionada ao nível de competência exigido pelo curso. Outra forma de averiguar a validade é mediante a correlação estatística, aliada a outras formas de mensuração de notas. Outros fatores relacionados à validade podem focar as conseqüências de um determinado teste ou até mesmo a percepção de validade de um aluno e/ou candidato em relação ao teste. Para analisar criteriosamente cada aspecto da validade, é necessário ponderar algumas variáveis, tais como as relacionadas: a) ao conteúdo, b) ao critério, c) ao construto, d) à validade conseqüencial e e) à validade da face.

Validade relacionada ao conteúdo

Quando um teste fornece uma amostragem real do conteúdo testado, no sentido de obter do aluno um desempenho da habilidade em avaliação, pode-se dizer que este

teste evidencia validade relacionada ao conteúdo. Em outros termos, se o objetivo do teste é mensurar a competência do aluno ao redigir um texto com criatividade, ele deve ter a oportunidade na prova de criar um texto e demonstrar sua criatividade e competência.

Outra perspectiva relacionada à validade de conteúdo é a diferenciação entre testagem direta e indireta. Na testagem direta o candidato demonstra seu conhecimento por meio de tarefas que forneçam uma amostragem da habilidade sob avaliação. No entanto, na testagem indireta, outras habilidades são empregadas na avaliação para que se permita, por meio delas, inferir sobre o provável desempenho de um aluno na habilidade em questão. Por exemplo, para avaliar a produção de parágrafos bem estruturados o aluno se submete a um teste que envolva a organização de idéias apresentadas em um exercício de compreensão aural. Por questões de praticidade nem sempre é possível atender aos aspectos da validade de conteúdo em um contexto de sala de aula real, onde o elevado número de alunos, a falta de materiais de apoio, o custo, o tempo ou até mesmo o nível de preparo do professor comprometem reflexões maiores sobre o assunto.

Validade relacionada ao critério

A validade relacionada ao critério tem duas categorias, concomitante e de predição. A validade concomitante existirá se os resultados de um teste aplicado a um grupo possam ser comparados a outras fontes de desempenhos dos alunos, como outros testes ou instrumentos de avaliação, desde que os resultados sejam semelhantes. Já a validade de predição diz respeito a testes de classificação, testes de admissão ou testes de aptidão lingüística, que possam ser utilizados para mensurar a habilidade de um aluno/candidato em um dado momento, como no início de um curso de língua inglesa ou para a sua classificação dentro de um currículo escolar, e que o seu acompanhamento após o início dos estudos possa ser verificado, e constatada a sua probabilidade de sucesso identificada no início, pelo instrumento de avaliação aplicado.

Validade relacionada ao construto

Um construto é qualquer teoria, hipótese ou modelo que procure explicar fenômenos observados em nossos universos de percepções. A verificação de construtos geralmente requer dados inferenciais, uma vez que não podem ser mensurados diretamente. A proficiência e a competência comunicativa são exemplos de construtos lingüísticos que, assim como outros assuntos na aprendizagem de uma língua estrangeira, são embasados em construtos teóricos.

De acordo com Davidson, Hudson e Lynch (1985), “testes são, por assim dizer, definições operacionais de construtos no sentido de operacionalizarem a entidade que está sendo mensurada”. Em outras palavras, por meio dos testes é possível verificar se o construto teórico que embasa uma prática de ensino-aprendizagem específica está funcionando de forma adequada, pois os resultados podem ser utilizados para balizar a atividade prática de ensino.

Na maioria das vezes, os professores ficam temerosos com a possibilidade de formalizar os procedimentos de validação de construto dos seus testes, por acreditarem ser uma tarefa difícil, e acabam por simplesmente verificar o conteúdo do teste com o conteúdo estudado.

Validade conseqüencial

A validade conseqüencial compreende todas as conseqüências de um teste no universo acadêmico, inclusive considerações referentes à sua precisão em mensurar os critérios propostos, o seu impacto no preparo dos candidatos ou alunos, o seu efeito no aprendiz e as conseqüências sociais voluntárias ou involuntárias da interpretação e uso de um teste.

Nas últimas duas décadas, o aumento pela procura de testes de proficiência ou avaliações de pré-requisito para estudos no exterior, como TOEFL-Test of English as a Foreign Language, GRE-Graduate Record Examination,

GMAT-Graduate Management Admission Test, TOEIC-Test of English International Communication, entre outros, teve sua parcela de validade consequencial no desenvolvimento de cursos preparatórios e manuais com dicas para obter bons resultados. Para McNamara (2000, p. 54), é preciso ter cautela, pois os resultados desses testes podem refletir a condição socioeconômica dos candidatos, que por meio de aulas preparatórias específicas ou de auxílio que possam receber de terceiros, por disporem de situações econômicas que favoreçam os estudos, conseguem maior índice de aprovação que os demais.

Outro aspecto importante da validade consequencial dos testes é que os resultados podem influenciar a motivação dos alunos, o seu desempenho no curso, a aprendizagem independente, seus hábitos de estudo e até mesmo a sua atitude em relação aos trabalhos escolares.

Validade de face

“A validade de face de um teste refere-se à aparência do mesmo, no sentido de parecer adequado, passar a percepção de mensurar o conhecimento ou a habilidade que se propõe a avaliar, com base no julgamento subjetivo dos examinandos, do pessoal administrativo que opta por usá-lo, ou em outros observadores psicometricamente pouco sofisticados.” De acordo com esta citação de Mousavi (2002, p. 244), é possível perceber que a validade da face de um teste é baseada na percepção que se tenha da sua aparência, no seu impacto na percepção intuitiva daqueles que façam ou apliquem o teste. Embora um tanto quanto pobre na sua fundamentação, pois a percepção da validade da face é meramente intuitiva, esse aspecto da validade compromete a avaliação uma vez que reflete diretamente na percepção daqueles envolvidos com a testagem e, portanto, na imagem que possam desenvolver do processo de avaliação.

Não é possível testar empiricamente a validade da face de um teste, mas a

validade será alta se o examinando perceber características como: um formato de tarefas familiar e bem elaborado; um teste que seja possível de fazer dentro do espaço de tempo existente; itens claros e simples; enunciados claros e objetivos; tarefas que exijam o conteúdo ensinado, e um nível de dificuldade compatível com o conhecimento.

Segundo H. D. Brown (2004), o impacto de um teste desencadeia vários princípios básicos da aquisição de uma língua, como a motivação intrínseca, a autonomia do aprendiz, a autoconfiança, o ego lingüístico, a interlíngua, o investimento em estratégias de aprendizagem, entre outros, e tem uma influência que pode ser positiva ou negativa, dependendo do enfoque dado aos resultados dos testes. Por essas razões é muito importante que o avaliador e/ou professor tenha bastante cautela, no sentido de auxiliar o seu aluno a refletir sobre os seus resultados, sem permitir que o *feedback* negativo tome uma dimensão exacerbada.

Outro princípio importante na testagem da linguagem é a autenticidade. Bachman e Palmer (1996, p. 23) definem a autenticidade como o “grau de correspondência entre as características das tarefas de um teste, com os aspectos reais dos usos do idioma”, em outras palavras, é necessário que o teste apresente itens que imitem ou reproduzam o uso real da língua em estudo. A autenticidade pode ser percebida de maneira variada, tais como: o quanto a linguagem do teste é natural; se a apresentação de itens é contextualizada; a presença de tópicos que sejam significativos para o aluno; uma organização temática ou episódica dos itens; o quanto as tarefas representam ou se aproximam da realidade.

1.1.3 Tipos de testes e seus objetivos

A compreensão e melhor aplicação e planejamento da avaliação e suas características gerais implica o conhecimento dos tipos de testes existentes e os objetivos de cada um, para que sejam feitas escolhas adequadas. Cada avaliação programada tem um propósito específico em relação aos seus avaliandos. Por exemplo, ao iniciar um curso de inglês, nos quais existem diversos níveis de

competência lingüística, o aluno precisa passar por uma testagem que o enquadre em um dos níveis. Para este aluno, é necessário que se aplique um teste de classificação, cujo objetivo final seja o seu nivelamento dentro de uma grade de níveis diferentes. Existem outras opções de testes, para objetivos variados, conforme será demonstrado a seguir.

O objetivo dos testes de proficiência é medir a competência global em um idioma, que não está relacionada a um currículo ou a um curso específico e sim ao nível de conhecimento desejado, para que um aluno possa ser considerado proficiente. Um dos problemas que se enfrenta com os testes de proficiência é o que se entende por nível de proficiência, qual habilidade deve ser considerada, como se mede ou se define a proficiência. Um exemplo de teste de proficiência é o ECPE (*Examination for the Certificate of Proficiency in English*), da Universidade de Michigan, EUA, que mede a competência lingüística de um candidato por meio de exercícios que englobam habilidades de compreensão de texto e aural, vocabulário, estrutura gramatical, uma composição e uma entrevista.

Os *testes diagnósticos* são desenvolvidos com o objetivo de diagnosticar aspectos específicos de uma língua estrangeira, para verificar o nível de conhecimento dos alunos dentro do assunto testado, e os *testes de classificação* para classificá-los dentro da divisão de um curso, ou em relação a outros alunos ou candidatos. Alguns testes de proficiência podem ser utilizados para classificar ou diagnosticar conhecimentos diversos.

Testes de domínio de conteúdo (achievement) são geralmente utilizados para mensurar o desempenho dos alunos relacionados aos conteúdos trabalhados nos materiais didáticos ou nos conteúdos propostos nos currículos. O seu objetivo é detectar o quanto os alunos apreenderam do universo de assuntos abordados e, para tanto, se limitam aos livros ou conteúdos estudados.

Existem também os *testes de aptidão*, que têm por objetivo prever o desempenho de um aluno em uma língua estrangeira, sem que ele já tenha sido exposto a ensinamentos, ou seja, o teste mede a capacidade ou a habilidade de um

indivíduo para aprender um idioma. Para Skehan (1989), testes de aptidão lingüística refletem apenas inteligência geral ou habilidades acadêmicas do aluno. Para H. D. Brown (1994), outro aspecto desfavorável do teste de aptidão é que, devido a sua origem mal fundamentada teoricamente, sua interpretação é ambígua e pode atribuir a quem o faz sucessos ou fracassos pouco embasados teoricamente ou confiáveis.

Os testes podem ser **diretos** e **indiretos**. Segundo Nicholls (2001), os primeiros caracterizam-se pelas perguntas abertas que permitem respostas individuais, obviamente diferentes. Apesar de serem fáceis de elaborar, o seu julgamento se torna difícil, principalmente tratando-se de uma língua estrangeira, em que fatores constituintes do discurso, como a pronúncia, a fluência e a ortografia, sobrepõem-se aos fatores pragmáticos, de sintaxe e de escolha lexical, que compõem o item testado. Os testes indiretos, entre eles o de múltipla escolha, o de verdadeiro e falso, o de associação, entre outros, embora mais difíceis de elaborar, apresentam vantagens, tais como a facilidade de correção, maior confiabilidade e validade, considerando-se a existência de apenas uma resposta correta, conforme a elaboração do teste.

A amostragem do conhecimento do aluno pode ser obtida por meio de diferentes formatos de atividades nos testes. Para decidir entre as opções existentes, é necessário que o avaliador considere aspectos como a praticidade, confiabilidade e a validade de cada formato e então escolha o modelo mais adequado à sua realidade. Segundo Penny Ur (1996), existem doze modelos de atividades, entre as mais comuns, de obter o conhecimento do aluno, conforme descritas abaixo:

- 1) **Associação.** Os exercícios dessa técnica são fáceis de elaborar e corrigir; essas questões tornam-se às vezes ambíguas, possibilitando respostas sobrepostas. Nos exercícios de associação são dadas palavras semelhantes para que os alunos as relacionem com as outras opções dadas nos exercícios. Verificam essencialmente conhecimentos lexicais.
- 2) **Cloze.** Nos exercícios de cloze, palavras são omitidas de um texto com intervalos regulares, por exemplo, a cada sétima palavra. É popular entre os professores para testar leitura, ortografia, conhecimento

gramatical e lexical. Sua elaboração é fácil, contudo, sua correção envolve decisões sobre a adequação dos itens que possam ser apresentados pelo aluno.

- 3) **Ditado.** O objetivo dessa atividade é avaliar a ortografia e a pontuação. Contudo, indiretamente, avalia a compreensão aural, uma vez que o aluno só consegue escrever o que compreendeu. Apesar de ser fácil de elaborar e de executar, sua correção envolve decisões subjetivas sobre quantos pontos atribuir a diferentes tipos de erros.
- 4) **Dramatização.** Essa técnica evidencia a competência comunicativa dos alunos. Por meio de tarefas realizadas em pares ou grupos, observam-se a fluência e a adequação lingüística; é fácil de executar, mas difícil de avaliar devido aos diferentes aspectos a serem considerados.
- 5) **Ensaio.** O aluno recebe um tópico sobre o qual deve discorrer para avaliar suas habilidades discursivas. É fácil de planejar, contudo sua avaliação é extremamente subjetiva e difícil, pois itens como ortografia, pronúncia, fluência, gramática, conteúdo, entre outros, exigem pontuação diferenciada, dificultando a elaboração e a aplicação de critérios de avaliação.
- 6) **Múltipla Escolha.** Trata-se de uma forma muito eficiente de testar, uma vez que oferece um número de opções para uma pergunta onde apenas uma é a opção correta. Apesar de ser mais fácil de corrigir, sua elaboração é difícil.
- 7) **Perguntas e Respostas.** Esse tipo de questão é útil para testar praticamente qualquer conteúdo. São elaboradas perguntas cujas respostas são formuladas pelo aluno livremente. Ademais, é fácil de elaborar; no entanto, a sua correção tende a ser muito subjetiva e, pelo fato de permitir liberdade de resposta, é considerada uma testagem aberta.
- 8) **Preenchimento de lacunas ou completamento.** Geralmente usado para verificar conhecimentos gramaticais e vocabulário, esse tipo de

questão é fácil de elaborar assim como de atribuir notas, tratam-se de frases ou textos em que algumas palavras são omitidas para que o aluno complete a lacuna; no entanto, as respostas podem ser múltiplas, tornando a questão ambígua e comprometendo a sua validade.

- 9) **Reescrita.** Dada uma sentença, o aluno deverá reescrevê-la, alterando sua estrutura sem alterar o seu sentido inicial. É similar às questões de transformação, porém, mais válida, uma vez que o aluno deve parafrasear o sentido da sentença, demonstrando maior conhecimento do item examinado.
- 10) **Tradução.** Um dos pontos controversos da tradução é que o aluno cria o hábito de constantemente buscar apoio na língua materna, levando-o a tradução literal para a língua alvo e vice-versa. Contudo, é uma forma rápida e confiável de verificar o conhecimento do aluno, principalmente quando envolve expressões idiomáticas.
- 11) **Transformação.** Nesse tipo de exercício o aluno deve transformar uma sentença específica de acordo com instruções. É fácil de elaborar, executar e corrigir; sua validade é, porém, suspeita, uma vez que testar a habilidade de transformar estruturas gramaticais não significa testar a competência comunicativa. É possível que o aluno consiga fazer transformações nas sentenças, sem entender o sentido ou o uso que se faz delas.
- 12) **Verdadeiro ou Falso.** Questões de V ou F avaliam apenas a compreensão e assimilação de informações. Podem ser usadas para verificar aspectos gramaticais e vocabulário. São relativamente fáceis de elaborar, aplicar e corrigir. Nesta forma de avaliação são dadas sentenças com informações verdadeiras ou falsas para que o aluno aponte as respostas V ou F, de acordo com um texto ou com o seu conhecimento do conteúdo.

1.2 AVALIAÇÃO DA ESCRITA

A habilidade escrita há cerca de uns vinte anos era vista como uma convenção primária para registrar a fala e para reforçar aspectos gramaticais e lexicais de uma língua. No entanto, segundo Smith (1994), para uma maior compreensão da complexidade da escrita e conseqüentemente da sua avaliação, é necessário desconectar o processo de escrita do processo da fala.

De acordo ainda com Smith, o ritmo no qual as palavras nos vem à mente é de 200 a 300 unidades por minuto. Ao transcrevermos manualmente as palavras para o papel, escrevemos aproximadamente 25 unidades por minuto, ou 60 palavras por minuto em um computador. Essa “lentidão” da transcrição escrita em relação ao processamento mental das idéias é um fator complicador da produção escrita, pois muitas vezes o autor não consegue diminuir o ritmo do seu pensamento, chegando até mesmo a esquecer suas idéias. Ainda segundo Smith, existe um limite na quantidade de informações que o cérebro humano pode processar conscientemente, e quando o pensamento se ocupa com preocupações relacionadas às convenções da escrita, dificilmente é possível refletir sobre aspectos relacionados ao conteúdo. Para resolver conflitos relacionados à habilidade escrita no sentido de produzir textos bem escritos e de bom conteúdo, Smith sugere a re-escrita e que se trabalhe esta habilidade como um processo.

Atualmente a importância da escrita como uma habilidade lingüística, com aspectos e convenções próprias, é inquestionável. Compreende-se também a dificuldade em aprender a escrever em qualquer língua, mesmo na língua materna. Todas as pessoas educadas aprendem no mínimo os rudimentos da escrita em sua língua materna e, no entanto, poucos aprendem a se expressar claramente, de forma lógica e bem organizada. Mesmo assim, é esperado dos aprendizes de uma língua estrangeira que produzam textos coerentes, bem elaborados, e que façam bom uso de aspectos do discurso.

A avaliação da escrita não poderia seguir uma linha diferente no que diz respeito

às expectativas dos avaliadores em relação aos alunos. Ao se considerar a análise da escrita, é necessário ter clareza sobre os objetivos e critérios da avaliação, e diversos aspectos devem ser revistos, tais como o contexto e o propósito de uma composição, os diferentes estágios do desenvolvimento da habilidade escrita e os objetivos a serem trabalhados. Para tanto, é preciso antes rever os diferentes gêneros da linguagem escrita, os tipos de escrita e as micro e macro-habilidades da escrita.

Os gêneros da linguagem escrita necessários para o contexto de ensino-aprendizagem de uma língua estrangeira são vários. Nesta pesquisa, porém, serão listados os mais comuns, aqueles que um escritor deve aprender a produzir dentro dos limites de um currículo escolar. H. Douglas Brown (2004, p. 219) divide os gêneros textuais principais em três partes, a redação acadêmica, a redação profissional e a redação pessoal, cada um delas com seus gêneros específicos, conforme quadro a seguir.

QUADRO 1 – RELAÇÃO DOS GÊNEROS DA ESCRITA

<p>GÊNEROS DA ESCRITA</p> <p>1. Redação Acadêmica</p> <p>Trabalhos e relatórios gerais</p> <p>Ensaio, composições</p> <p>Jornais com enfoque acadêmico</p> <p>Respostas curtas em testes</p> <p>Relatórios técnicos</p> <p>Teses, dissertações</p> <p>2. Redação Profissional</p> <p>Recados</p> <p>Cartas, e-mails</p> <p>Memorandos</p> <p>Relatórios</p> <p>Cronogramas, rótulos e sinais</p> <p>Propagandas, anúncios</p> <p>Manuais</p>
--

3. Redação Pessoal

Cartas, e-mails, cartões, convites

Recados, bilhetes

Anotações em calendários, listas de compras, lembretes

Documentos financeiros

Formulários, questionários, relatórios médicos, documentos de imigração

Diários, jornais pessoais

Ficção

FONTE: BROWN, H. D. 2004, (p. 219).

A existência de formatos diferentes de apresentação da escrita, para responder ao objetivo do texto, determina um cuidado por parte do escritor com o gênero em questão e do avaliador no momento de planejar o instrumento da avaliação e de fazer a avaliação propriamente dita. Para O'Malley e Pierce (1996), os gêneros, na sua maioria, dividem-se em informativo, narrativo/descritivo ou persuasivo, e o desempenho de um aluno pode variar consideravelmente em razão do gênero escolhido. Portanto, é aconselhável que se avalie a escrita por meio de gêneros diferentes para que seja possível fazer inferência sobre o nível de competência escrita de um aluno. Segundo H. D. Brown (2004), os tipos de desempenho escrito que englobam uma diversidade de produções escritas são divididos em quatro categorias; a imitação, a intensiva ou controlada, a responsiva e a extensiva. Por considerar a explicação do autor H. D. Brown bastante detalhada ao se referir às habilidades exigidas dentro de cada categoria, sua exemplificação sobre o enfoque e as tarefas das categorias elucidadoras para o contexto de escrita na língua estrangeira, explicarei com mais detalhes as quatro categorias de desempenho escrito.

Na categoria da *imitação* a produção escrita limita-se às habilidades básicas, como a escrita de palavras isoladas, a mecânica da pontuação e frases curtas. Neste nível os alunos aprendem sobre a mecânica da escrita, e o enfoque é dado para a forma, a estrutura, enquanto os aspectos relacionados ao contexto e ao significado são secundários. As tarefas trabalhadas nesta categoria limitam-se à escrita de palavras, a pontuação, tarefas de soletrar e identificar correspondência entre fonemas e grafemas.

Já a categoria *intensiva ou controlada* exige dos aprendizes que tenham

habilidade de selecionar vocabulário apropriado a contextos específicos, que façam uso de colocações e expressões idiomáticas, assim como escolhas de estruturas gramaticais apropriadas, no nível da sentença. O significado e o contexto são importantes para determinar o acerto e a adequação estrutural; no entanto, as atividades de avaliação enfocam a forma e são controladas na elaboração dos testes. Alguns exemplos de tarefas encontradas nos testes desta categoria são: ditado, tarefas de transformação gramatical, tarefas baseadas em figuras, exercícios para avaliar vocabulário, atividades de colocar palavras em ordem e tarefas de respostas curtas ou de completar sentenças.

Na categoria *responsiva* a avaliação requer do aluno um desempenho limitado no nível do discurso, a elaboração de parágrafos e uma seqüência lógica entre eles. As tarefas devem respeitar direcionamentos pedagógicos, listas de critérios, “*outlines*” e outras formas de direcionamento. Entre os gêneros trabalhados nesta categoria, encontram-se narrativas e descrições, relatórios curtos, relatórios laboratoriais, resumos, respostas breves, bem como a interpretação de gráficos ou tabelas. De acordo com H. D. Brown, o escritor dá início a experiências escritas com maior liberdade de escolha entre as formas de expressão de idéias alternativas; neste estágio já conhece os fundamentos da estrutura gramatical e está mais preocupado com as convenções do discurso para que atinja os objetivos da sua composição. A ênfase recai no contexto e significado, e a atenção à forma é direcionada apenas no nível do discurso.

A categoria *extensiva* da escrita implica o domínio complexo de todos os processos e estratégias da escrita, para propósitos variados, tais como ensaios, trabalhos, relatórios de pesquisas formais, dissertações e teses. O enfoque está em atingir os objetivos e propósitos de escrita, organizar e desenvolver idéias de maneira lógica, adicionar detalhes para apoiar ou ilustrar idéias, demonstrar variedade sintática e lexical, e, em diversas situações, engajar-se no processo da escrita de vários esboços até atingir um produto final. O enfoque na estrutura gramatical limita-se a correções ocasionais. Na categoria *responsiva* tanto quanto na *extensiva*, as atividades de escrita são mais complexas e abertas, como, por exemplo, a paráfrase, perguntas e resposta

direcionadas, tarefas que envolvam estratégias para uma boa redação.

Dentro das categorias *responsiva e extensiva*, os escritores envolvem-se na arte da composição, ou na ciência da escrita propriamente dita, e estas categorias são o alvo de discussões relacionadas ao ensino e à aprendizagem da escrita, à avaliação da escrita e a outros aspectos que não são percebidos na produção escrita de nível básico.

Outro fator importante a ser considerado na avaliação da escrita é a taxonomia que divide as micro e macro-habilidades que, segundo H. D. Brown (2004), é útil na definição de critérios de avaliação, como veremos na seqüência. As micro-habilidades estão mais relacionadas às categorias de imitação e intensiva, devido às suas limitações ao nível de sentença, com aspectos relacionados ao uso aceitável das regras gramaticais, utilização de formas gramaticais diferentes para expressar um significado específico e o uso de mecanismos de coesão no discurso escrito. No entanto, as macro-habilidades são adequadas às categorias responsiva e extensiva, pois trabalham a redação no nível do discurso. Procuram, assim, atingir as funções comunicativas do texto adequadamente à sua forma e ao propósito, demonstrar ligação e conexão entre eventos, demonstrar referências culturais corretamente e fazer uso de estratégias da escrita, como avaliar a interpretação da audiência, parafrasear, usar sinônimos, usar *feedback* de outros para editar o texto.

A complexidade das exigências no desempenho dos alunos ao escrever dentro das categorias responsiva e extensiva, não poderia deixar de existir no que diz respeito à avaliação. Dentre as ressalvas que devem ser observadas no momento de preparar ou aplicar um teste de escrita, ou uma avaliação escrita, aspectos como a autenticidade da tarefa, o tempo necessário para desempenhá-la e o critério adotado para a análise e atribuição da nota, são de extrema relevância nos resultados. Considerando-se que o enfoque deste estudo está na avaliação da escrita e na redução da subjetividade, na avaliação deve ser observado cada ponto da descrição dos critérios.

Ao programar uma avaliação escrita, o critério de análise e correção do teste deve ser desenvolvido juntamente com a avaliação, pois os objetivos do teste devem

estar adequados aos critérios de avaliação para que haja confiabilidade e validade nos resultados. McNamara (2000), em seu quadro “Test and Criterion” (“Teste e Critérios de Referência”), classifica “critérios” como características observáveis do desempenho de um aluno em situações reais, subseqüentes ao teste. Ou seja, de acordo com o que se quer avaliar sobre o provável desempenho de um aluno em situações reais, como na produção de um texto escrito, elabora-se um “teste” com situações e exigências semelhantes às reais para que se torne possível prever a sua competência. Na avaliação de desempenho o aluno demonstra domínio prático das habilidades aprendidas para que seja possível por meio dessa amostra inferir sobre o seu provável desempenho em situações reais.

No que tange à avaliação da produção escrita dos alunos de língua estrangeira, Kathleen M. Bailey (1988) aponta duas abordagens como as mais tradicionais, avaliação holística e analítica. Cada uma delas é definida pelo critério de avaliação utilizado. Segundo Wesche (1983), entende-se por critério de avaliação que as questões de um teste devem indicar não somente se um aluno domina ou não um conteúdo específico, e sim o quanto ele domina e quanto mais precisa saber, ou seja, o critério demonstra conhecimento de uma determinada habilidade, mas deve demonstrar também como tal conhecimento é aplicado na prática. De acordo com Bailey (1988), a característica de um critério demonstra a função qualitativa que uma avaliação pode ter. Ainda segundo a mesma autora, o que distingue a avaliação holística da analítica é a maneira de atribuir notas aos textos produzidos à luz destes critérios de correção.

Na abordagem holística, o corretor reage à totalidade do texto escrito, atribui uma nota final única. Geralmente os critérios são formados por escalas numéricas, e para cada número da escala existem descrições dos requisitos para que se obtenha a referida nota. Para Arthur Hughes (1989), a vantagem da avaliação holística é a rapidez na aplicação.

A segunda abordagem mais utilizada, segundo Bailey (1988), é a analítica. Neste sistema os corretores avaliam o desempenho dos alunos em vários quesitos. Para Jacobs, Zingraf, Wormuth, Hartfiel e Hughey (1981), entre outros autores, aspectos

como conteúdo, organização, vocabulário, estrutura e “mecanismos da escrita” devem ser avaliados separadamente, e recebem pesos diferentes no somatório final da correção. De acordo com Bailey (1988) e Hughes (1989), a desvantagem do sistema analítico de avaliação é desviar a atenção do corretor do texto enquanto um produto completo e inteiro, para focar os diferentes aspectos a serem registrados na avaliação, de acordo com os critérios da correção.

Diversos autores e pesquisadores desse campo do saber, como Herman, Aschbacher e Winters (1992), e H. Douglas Brown (2004), colocam três tipos de critérios de correção como os mais utilizados na avaliação da escrita, são eles os modelos analítico, de traços primários e holístico. Bailey (1988) coloca ainda, além desses três modelos, o critério objetivo. Conforme o objetivo deste estudo, a avaliação da escrita e sua subjetividade é importante para a pesquisa e o detalhamento dos critérios de avaliação, como apresentado na seqüência.

A. Modelo de Traços Primários

O enfoque desse critério de correção está na qualidade da composição do aluno dentro de um discurso limitado, ou seja, a nota é atribuída com base na eficácia do texto em atingir o seu objetivo. Por exemplo, se o objetivo da composição for persuadir o leitor a fazer algo, a nota será dada considerando essa função da composição. Aspectos como a organização, detalhamento do assunto, fluência, variedade sintática, entre outros, são avaliados no processo de atribuição da nota; entretanto, a vantagem desse método é permitir que, tanto o escritor quanto o avaliador dêem enfoque à função do texto.

B. Modelo Objetivo

O modelo de correção objetivo, conforme o nome sugere, baseia-se em métodos quantitativos para avaliar uma composição. Apesar de a nomenclatura “objetivo” parecer bastante positiva, existem aspectos negativos e de subjetividade

nesse modelo. As composições avaliadas pelo modelo objetivo devem considerar um número máximo, limitado, de palavras – 250. Para tanto, é necessário contar as primeiras 250 palavras e ignorar o restante, porém, se o aluno não tiver concluído seu texto até esse limite de palavras e tiver escrito uma conclusão excelente, esta não será considerada. Feito isso, cada erro deve ser sublinhado e receber um valor numérico de 1 a 3, sendo 1 para erros leves, 2 para erros moderados e 3 para erros graves. O segundo aspecto negativo e que gera subjetividade está no entendimento do que deve ser considerado um erro leve, moderado ou grave. Ademais, nesse modelo o enfoque está nos erros e não nos acertos. No final da correção, o número de erros é calculado e dividido pelo número de palavras, fornecendo assim uma nota com base no número de palavras produzidas e na quantidade de erros até esse limite de produção.

$$\text{Fórmula} \rightarrow \frac{(\text{n}^\circ \text{ de palavras consideradas})}{(\text{somatório dos erros})} = \textit{score} \text{ de acertos}$$

Apesar dos aspectos negativos já apontados, esse modelo de correção, ao meu ver, tem um ponto positivo a seu favor: o número de palavras consideradas, dividido pelo número de erros, elimina as diferenças de notas entre alunos que escrevem muito e outros que escrevem pouco.

C. Modelo Analítico

O critério analítico de avaliação verifica o desempenho do aluno sob vários aspectos de um texto, alguns deles são estrutura gramatical, expressão idiomática, organização, relevância, coerência, entre outros. Cada uma das categorias adotadas recebe uma pontuação, por exemplo, 20 pontos, que numa escala são somadas atribuindo uma nota de até 100 pontos. As escalas de correção analítica podem ser adaptadas de acordo com o nível de proficiência dos alunos / escritores.

Ao adotar um critério de correção analítico, a atenção do avaliador estará concentrada em um número de aspectos específicos, avaliados separadamente, que

atribuirá um resultado numérico preciso e informativo, possibilitando um diagnóstico adequado da proficiência escrita do aluno. Outro aspecto favorável desse modelo é que o *feedback* dado ao aluno sobre a sua composição é detalhado e rico, possibilitando um impacto positivo na sua capacidade de redigir. Contudo, o modelo analítico é bastante específico com relação aos seus critérios e esse fator torna a correção mais demorada, perdendo em praticidade; além disso exige maior confiabilidade no entendimento do corretor ao utilizar esse critério.

D. Modelo Holístico

Nesse modelo é utilizada uma escala de correção única, que descreve diferentes níveis de produção e desempenho escrito. O avaliador usando o modelo holístico deve reagir à totalidade do texto escrito, ou seja, avaliá-lo como um produto único e atribuir uma nota.

Ao efetuar a avaliação holística da proficiência escrita, os avaliadores devem ponderar aspectos além da precisão lingüística dos alunos. Fatores como as exigências da tarefa, o nível de proficiência lingüística, a escolha de vocabulário e a complexidade da sentença, são tão importantes numa escala holística quanto aspectos da exatidão da linguagem.

O sistema de avaliação holístico pode ser aplicado a uma infinidade de tópicos e tipos de redação, ademais, esse critério enfatiza os pontos fortes e positivos de uma composição. A visão global de uma redação permite que o corretor, após o treinamento e a compreensão dos critérios e suas especificidades, atribua um resultado com agilidade e rapidez, tornando o critério holístico prático em relação ao fator tempo. Devido a essa característica, diversos testes como o TOEFL e o FCE (First Certificate in English) da Universidade de Cambridge, que tem um número elevado de candidatos para serem avaliados e limites de tempo, adotam critérios de correção holísticos (ver modelos nos anexos 1 e 2).

A desvantagem do modelo holístico de correção está no fato de não fornecer aos alunos um *feedback* diagnóstico. No caso de testes como o TOEFL e o FCE, o candidato não recebe *feedback* do seu desempenho, mas no contexto da sala de aula é

necessário que o professor adicione comentários e sugestões que redirecionem a escrita de seus alunos.

Apesar de os sistemas de avaliação, holísticos ou analíticos, fornecerem subsídios para que os corretores avaliem criteriosamente a produção escrita em língua estrangeira, existem vários outros fatores que podem influenciar no processo de avaliação, conforme aspectos apontados na revisão de literatura sobre validade e confiabilidade dos testes. Estes fatores podem impedir a obtenção de resultados precisos, ou ainda levar um mesmo texto a resultados diferentes.

1.3 AVALIAÇÃO DE DESEMPENHO

Um fator determinante dos testes de desempenho é que os candidatos demonstram seu conhecimento por meio de amostragem de habilidades reais. McNamara (2000) caracteriza a avaliação de desempenho com duas funções principais, uma amostragem profissional e outra cognitiva. A primeira delas tem uma abordagem que se baseia na amostragem de simulações de contextos profissionais; a segunda tem um caráter mais cognitivo e uma abordagem lingüística diferenciada das avaliações tradicionais, pois o enfoque não está em uma tarefa específica, e sim na qualidade do desempenho do aluno ao executá-la. Ou seja, a evidência do conhecimento do aluno sobre um determinado sistema lingüístico durante o processo da avaliação é o que interessa ao avaliador, que mediante os dados coletados poderá inferir sobre o comportamento lingüístico dos alunos a partir da amostragem das avaliações.

Ainda segundo McNamara (1996), existem duas tradições em avaliação de desempenho, uma aplicação simples de técnicas de avaliação em contextos profissionalizantes, portanto não lingüísticos, e outra tradição que classifica o desempenho em uma língua estrangeira como uma conquista cognitiva complexa e que envolve a integração de vários processos psicolingüísticos.

A avaliação de desempenho de uma língua estrangeira enfoca dois aspectos principais, a qualidade da execução da atividade e o conhecimento subjacente da língua, que é revelado durante o desenvolvimento da atividade.

O diferencial da avaliação de desempenho em relação à avaliação tradicional, segundo Kenyon (1992), em sua representação esquemática das características dos testes, é o processo do desempenho, que envolve um número maior de etapas até a obtenção de um resultado final, conforme Quadro a seguir.

QUADRO 2 – REPRESENTAÇÃO ESQUEMÁTICA DA AVALIAÇÃO TRADICIONAL E DE DESEMPENHO

Avaliação tradicional com resposta fixa	Avaliação baseada em desempenho
Instrumento → Nota ↑ Candidato	Corretor ↓ → Resultado (Nota) Critérios ↓ Desempenho ↑ Instrumento ↑ Candidato

FONTE: Kenyon, 1992.

Conforme o esquema de Kenyon (1992), é possível perceber que o resultado de uma avaliação de desempenho é mais eficaz quando o objetivo da avaliação for inferir sobre o comportamento de um candidato ao usar o idioma em uma situação real. Assim, mediante uma amostragem do seu desempenho, o avaliador poderá prever o seu comportamento em um contexto em que o uso do idioma seja necessário. Pela avaliação tradicional, será atribuída apenas uma nota ao candidato, baseada em seu conhecimento do conteúdo no momento da avaliação.

Existem diversas descrições para os testes de desempenho, no entanto todas revelam o objetivo de fornecer um retrato de uma situação. Fitzpatrick e Morrison (1971 e 1991) definem testes de desempenho como aqueles em que “situações ideais e de referência” são simuladas com muito mais detalhamento e realismo do que as situações representadas nos testes escritos tradicionais. Os testes de desempenho avaliam tanto processos, tais como os testes de conversação numa língua estrangeira,

quanto produtos, como os testes de avaliação de escrita. Já Cronbach (1971), em sua definição de testes, acredita que, em geral, se trata de:

“A systematic procedure for observing a person’s behavior and describing it with the aid of a numerical scale or category system”.¹

Slater (1980) complementa a definição de Cronbach (1971) defendendo que o diferencial dos testes de desempenho está na forma como o termo “comportamento” é operacionalizado, ou seja, que tipos de estímulo provocam o comportamento a ser observado, que tipo de resposta é esperada do aluno e sob quais condições o comportamento é analisado. Para Slater (1980), estes três aspectos são os diferenciais dos testes de desempenho.

Em relação à operacionalização da testagem, existe a preocupação com a autenticidade e o realismo do comportamento avaliado, em comparação a situações ideais. Considerando que o objetivo de um teste de desempenho é inferir sobre o conhecimento do aluno quando confrontado com uma situação real, a qualidade do comportamento observado durante o teste deve ser abrangente e fiel à realidade, no sentido de fornecer uma representação adequada dos aspectos esperados em situações ideais.

De modo geral, uma das preocupações com a avaliação está relacionada à sua autenticidade e realismo. Em se tratando do ensino de uma língua estrangeira, em que o contexto não é de uso cotidiano do idioma em estudo, aspectos sobre o grau de realismo das questões ensinadas e avaliadas tornam-se relativamente delicados, uma vez que a sala de aula de língua estrangeira, onde os alunos dividem um idioma comum, não é um contexto ideal e real.

Fitzpatrick e Morrison (1991) declaram que não existe uma distinção absoluta

¹ “Um procedimento sistemático para observar o comportamento de uma pessoa, e a descrição de tal comportamento com o auxílio de uma escala numérica, ou um sistema de classificação”.

entre os testes de desempenho e outros tipos de teste, além do fato de os testes de desempenho serem relativamente reais devido à tendência de envolverem problemas específicos e demandarem decisões e procedimentos não presentes nos testes tradicionais.

Haertel (1992) concorda com a colocação de Fitzpatrick e Morrison (1991), pois classifica como testes de desempenho qualquer avaliação, cujo estímulo ou a resposta solicitada reproduza aspectos do contexto real, e afirma ainda que algumas mensurações de desempenho são baseadas em produtos obtidos apenas com o objetivo de avaliar. Para o autor, as avaliações de escrita têm a finalidade única de demonstrar a proficiência dos alunos, e em atividades cujo objetivo principal é avaliar, a reprodução das características de autenticidade dos contextos reais torna-se uma tarefa difícil, desnecessária e dispendiosa; ademais, as tarefas selecionadas para a avaliação de desempenho são bem mais simplificadas do que a realidade.

A autenticidade dos testes de desempenho e sua fidelidade às atividades da vida real é questão um tanto complexa; diversos autores dividem opiniões semelhantes sobre o que significa uma avaliação autêntica. Wiggins (1989) esclarece este assunto ao classificar desempenhos “escolares” como idênticos aos desempenhos da “vida real”. Segundo o autor, as avaliações autênticas de escrita, mesmo no contexto da segunda língua, constituem uma réplica dos desafios enfrentados por escritores, empresários, cientistas, líderes da comunidade, designers ou historiadores, que incluem tarefas como redigir ensaios ou relatórios, conduzir pesquisas individuais ou em grupos, desenvolver propostas ou simulações, elaborar *portfolios*, entre outras.

Existe uma sobreposição das técnicas utilizadas nas avaliações de desempenho na educação geral, com as técnicas aplicadas em áreas específicas de treinamento ocupacional e seleção de pessoal, cuja influência sobre os testes de segunda língua para contextos ocupacionais, desde 1970, foi direta e explícita. Diversos autores citam a prática da avaliação de desempenho em contextos profissionais, entre eles Wesche (1985, 1987, 1992) e Jones (1977, 1979, 1985). Segundo Jones, a avaliação da qualificação de pessoal divide-se entre testes de

desempenho e conhecimento, ou testes tradicionais, úteis em situações em que a proficiência não pode ser medida por testes escritos, como, por exemplo, em avaliações como a digitação ou tocar um instrumento musical. Ainda de acordo com Jones (1979:51), a avaliação de desempenho tornou-se mais presente no contexto lingüístico, a partir do crescimento do inglês para propósitos específicos (ESP), quando diz que:

“There is an increasing demand for language tests to provide predictive information on how well a candidate will perform in a given situation... it is impossible for a language test to predict task-oriented proficiency unless it includes or approximates actual samples of the tasks.”²

Dessa forma, na avaliação de desempenho tornou-se imprescindível a presença de amostras ou exemplificações do nível de conhecimento e produção dos alunos em relação ao conhecimento testado. No estudo em questão, o conhecimento da língua estrangeira, mais especificamente, da habilidade escrita.

Assim como Slater (1980), Jones (1985) também distingue três tipos principais de testes de desempenho: a avaliação direta, na qual a observação é feita diretamente no local de trabalho; métodos de amostragem de trabalho, cuja avaliação é feita *in loco*, embora as tarefas sejam controladas para que se obtenha um padrão de avaliação, e técnicas de simulação diferenciadas pelo fato de envolverem um grau de abstração da realidade no contexto de trabalho. Para o autor, as técnicas de simulação são as mais viáveis na testagem lingüística.

O estabelecimento da validade dos testes de desempenho deve ser analisado criteriosamente durante o estágio de desenvolvimento dos testes, considerando aspectos como o conteúdo da amostragem e sua veracidade em relação ao contexto real, assim como sua aplicabilidade no desenvolvimento do critério de correção.

Na abordagem de testes baseados em amostragem ou simulações de tarefas do contexto real, é necessário que se considere o desempenho como o alvo da

² “Existe uma demanda crescente por testes de idiomas que forneçam informações sobre o desempenho de um candidato ao usar o idioma em estudo em uma determinada situação, que se torna impossível por meio de um teste fazer inferências sobre a proficiência desse candidato ao executar tarefas, a não ser que na avaliação sejam incluídas amostras reais do seu desempenho.”

avaliação. Para Messick (1994), esta abordagem traz à tona aspectos como a réplica e generalização, ou seja, como fazer generalizações sobre o desempenho dos alunos baseadas em tarefas e simulações do contexto real, uma vez que muitos aspectos não são incluídos nos testes. Na realidade, o conteúdo selecionado para a avaliação deve ser representativo para tornar possível fazer inferências sobre o desempenho do aluno/candidato.

Conforme Ryans e Frederiksen (1951), os pontos principais a serem considerados ao produzir um teste de simulação são: a) fazer uma análise da tarefa, b) selecionar atividades que representem o seu contexto, e c) desenvolver um critério de avaliação adequado.

Henning (1988) afirma que a preocupação com a validade dos testes de desempenho não deve estar somente relacionada a aspectos *a priori*, e sim considerar a validade empírica dos testes, baseando-se em dados operacionais após a sua aplicação.

Os testes de desempenho tornaram-se mais representativos no contexto e na avaliação da segunda língua com o advento da abordagem comunicativa na década de 1970. Nos últimos quarenta anos os testes foram desenvolvidos para atender a dois aspectos principais, segundo McNamara (1996, p.25):

“(1) the practical need to develop selection procedures involving, for example, the selection of foreign students for study at English-medium universities, or personnel selection for particular occupations requiring communicative skill in a second language; and

(2) the need to bring testing into line with developments in language teaching which had resulted from the advent of theories of communicative competence.”³

As duas vertentes dos testes de desempenho têm objetivos diferentes: a primeira enfoca a “amostragem profissional” e é, portanto, mais pragmática e atórica, apesar da sua conexão com o contexto sociolinguístico do uso da língua; e a segunda tem um enfoque cognitivo e psicolinguístico, no qual a fidelidade dos desempenhos à

³ “(1) a necessidade prática de desenvolver procedimentos seletivos; por exemplo, a seleção de alunos para ingressar em universidades de língua inglesa, ou seleção de pessoal para ocupações que exijam habilidade comunicativa em uma segunda língua; e

(2) a necessidade de tornar a testagem mais adequada aos desenvolvimentos na área de ensino de línguas, cuja origem são as teorias da competência comunicativa.”

realidade e à habilidade subjacente, aspectos revelados pelo desempenho, são menos importantes do que o conhecimento. A origem do teste de desempenho como um veículo de avaliação que busca inferir o conhecimento do aluno vem da testagem pré-comunicativa, nas pesquisas de Lado (1961) sobre testes tradicionais, que buscavam especificar claramente, dentro de sua perspectiva estruturalista, o conhecimento lingüístico subjacente.

A difusão da abordagem comunicativa evidenciou a necessidade da avaliação de desempenho e, aliada aos resultados dos estudos recentes na área de mensuração, aumentou a possibilidade de se conduzir pesquisas empíricas sobre essa nova formatação das avaliações, tipicamente envolvendo julgamentos sobre qualidade classificados em uma escala de avaliação “pronta”. Alguns problemas surgem ao se pretender mensurar de forma objetiva e numérica desempenhos pouco objetivos e complexos. Esta pesquisa busca exemplificar métodos que possam minimizar a subjetividade na avaliação dos desempenhos escritos.

1.4 O PROCESSO DE AVALIAÇÃO E MENSURAÇÃO DE TESTES

A avaliação de desempenho, conforme estudos apresentados na revisão bibliográfica, envolve julgamentos subjetivos pelo fato de avaliar desempenhos humanos complexos. Tais desempenhos não podem ser avaliados de forma automática, exigindo a participação de um corretor, que precisa interpretar o desempenho em questão para que possa auferir um resultado; sendo assim, o processo de avaliação torna-se subjetivo.

O aspecto subjetivo da avaliação de desempenho é reconhecido há tempos e alguns métodos foram propostos para que fosse possível identificar e quantificar a extensão de subjetividade entre corretores e para minimizar os seus efeitos em níveis aceitáveis. Geralmente, essas metodologias sugerem o treinamento dos avaliadores e o monitoramento dos resultados.

As diferenças na atribuição de notas entre corretores são reduzidas por meio de treinamento, mas ainda assim persistem, e as notas brutas, que desconsideram

fatores que levam a resultados numéricos pouco confiáveis, não são fontes seguras para se medir a habilidade de um candidato. Segundo o exemplo dado por McNamara, 1996 (p.118), consideremos um contexto de avaliação de escrita na língua estrangeira, no qual dois alunos A e B terão sua habilidade escrita mensurada por dois corretores X e Y. O aluno A escolheu entre os temas propostos o mais complexo, enquanto o aluno B escolheu o mais simples. Os seus desempenhos tenderão a ser diferentes se considerarmos o grau de dificuldade apresentado nas tarefas, no entanto, os corretores atribuirão notas baseadas em uma mesma escala numérica, por exemplo, de 0 a 6. Outro aspecto que interfere nos resultados brutos é a rigidez do corretor; por exemplo, o aluno A teve seu teste corrigido pelo avaliador X que é mais rígido na correção, e lhe atribuiu uma nota 4, enquanto B teve a sorte de ser avaliado por Y, um avaliador menos exigente, que lhe atribuiu nota 5. Portanto, diante da variabilidade entre a complexidade das tarefas escolhidas e do nível de exigência dos corretores, não é possível afirmar a partir da nota bruta que o aluno A é menos hábil na escrita que o aluno B (quadro 3).

QUADRO 3 – HABILIDADE DE CANDIDATOS MEDIDA POR PADRÕES E CORRETORES DIFERENTES

PADRÃO / CORRETORES	CANDIDATO	
	A	B
Corretor	X (+ rigoroso)	Y (+ benevolente)
Tema	+ difícil	+ fácil
Resultados	4	5
Habilidade	inferior	superior

O Quadro 3 foi adaptado a partir dos dados citados no Quadro 5.1 de McNamara, 1996 (p.118).

Na realidade, existem diversas fontes de variabilidade no contexto da avaliação de desempenho que influenciam no resultado dos candidatos, às vezes trazendo conseqüências sérias. Estudos recentes apresentam uma abordagem diferenciada para minimizar os efeitos da subjetividade na avaliação de desempenho. A partir dos conceitos e procedimentos da medida multifacetada – uma nova teoria e

metodologia de mensuração relevante para a avaliação de desempenho –, torna-se possível especificar a natureza e a extensão da influência do corretor e de outros aspectos no contexto da avaliação; é possível aprofundar as pesquisas nesta área; e, em contextos administrativos adequados, é possível compensar os resultados do candidato pelos efeitos da influência do corretor e de outras variáveis.

A importância do avaliador em um processo de avaliação é tão fundamental que, para McNamara (1996), é possível dividir os procedimentos de avaliação em duas partes, a que envolve corretores, e a que não; o procedimento que envolve o corretor é a avaliação de desempenho. Alguns autores consideram o papel do corretor tão fundamental na avaliação que o colocam como eixo principal na definição da avaliação de desempenho, como Stiggins (1987, p. 33), que classifica como base o papel do avaliador e não a atividade do candidato sendo avaliado:

“Achievement can be, and often is, measured by observation and professional judgement. This form of measurement is called performance assessment.”⁴

O contraste entre a avaliação tradicional e a avaliação de desempenho, detalhado por Kenyon (1992), pode ser observado no Quadro 2, apresentado anteriormente.

Na avaliação tradicional com resposta fixa, a nota é obtida por meio de um instrumento (ou teste), que oferece para cada item um número de opções em que apenas uma delas é a correta. Nesta modalidade de avaliação, a tarefa de corrigir limita-se somente a contabilizar as respostas corretas e chegar a um resultado numérico, uma vez que o resultado baseia-se no conhecimento demonstrado nas respostas corretas. O foco da análise dos dados do teste é composto de dois estágios, cada um baseado na interação entre o candidato e o instrumento de avaliação ou teste, ou seja, na compreensão do teste pelo candidato, que é demonstrada a partir dos seus acertos. As respostas do candidato podem servir como evidência para analisar as suas

⁴ “A aprendizagem dos alunos, seus resultados podem ser, e geralmente são mensurados a partir da observação e do julgamento profissional. A esta modalidade de mensuração chamamos de avaliação de desempenho.”

habilidades e a qualidade do instrumento. A análise das respostas pode também levar a conclusões sobre a adequação dos itens do teste, ou seja, a facilidade de cada item e o seu poder discriminatório, e do instrumento de avaliação, sua confiabilidade de modo geral.

No que diz respeito à avaliação baseada em desempenho, o instrumento elucida um comportamento, que é julgado e avaliado por meio de um critério. Existe então uma outra forma de interação entre o corretor e os critérios que interfere no resultado: a interpretação que o corretor faz dos critérios.

1.4.1 As Fontes de Variabilidade na Avaliação da Escrita

A. A Natureza da Variabilidade entre Corretores

As diferenças entre os corretores, na teoria tradicional, foram conceituadas por conta da rigorosidade ou benevolência do corretor e características aleatórias ou sistemáticas na correção de erros. Segundo Linacre (1989), o grau de dificuldade dos corretores no momento da avaliação e os erros cometidos por eles são tão importantes quanto a questão da rigorosidade dos corretores e mais problemática, pois os erros aleatórios dos corretores são mais difíceis de serem compensados na nota final. As divergências das correções podem ser analisados e compõem diferentes fontes de variância revelando subpadrões no comportamento dos corretores, que podem ser sistemáticos ou aleatórios, portanto mais difíceis de serem analisados e padronizados.

Segundo McNamara (1996, p. 123-124), as formas de variabilidade mais significativas entre os corretores detalhadas a seguir podem elucidar para os objetivos desta pesquisa alguns aspectos importantes na subjetividade dos resultados:

- A. Corretores podem variar simplesmente na sua rigorosidade ou benevolência.
- B. Corretores podem revelar padrões de severidade ou benevolência em relação a um grupo de candidatos específicos, ou em relação a algumas

questões e não a outras. Um corretor pode também apresentar inconsistências na correção de determinados aspectos de um teste, deixando de lado outros, como, por exemplo, avaliar com maior rigorosidade a estrutura gramatical em um teste de escrita, desconsiderando erros na parte de coesão textual. Frequentemente corretores avaliam mais aspectos relacionados aos recursos formais da linguagem, como estruturas gramaticais, do que outros aspectos do desempenho. Em conclusão, existem inconsistências na correção relacionadas à interação entre corretor-candidato e entre corretor-item.

- C. Corretores podem diferir na maneira como interpretam os critérios de correção adotados. O problema aparece, pois critérios de correção usualmente apresentam categorias distintas de notas, como, por exemplo, de 1-6. Se um candidato em particular apresentar maior habilidade que os candidatos de nota “2”, mas não for tão hábil quanto os candidatos de nota “3”, que nota ele deverá receber? Alguns corretores atribuirão “2”, enquanto outros atribuirão “3”; no entanto, a dúvida é qual deverá ser a nota ou habilidade real do candidato.

Uma escala numérica dos critérios de correção deveria ser visualizada com intervalos iguais, da seguinte forma:

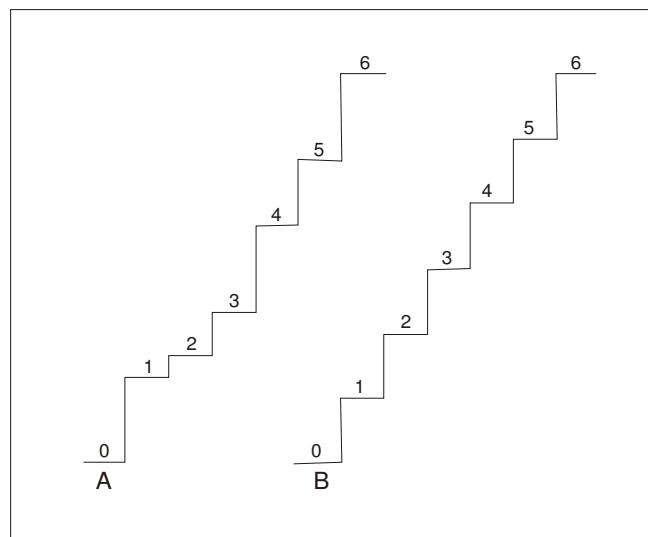
1 2 3 4 5 6

No caso das notas atribuídas que não se encaixam nem em categorias “2” ou “3”, a interpretação visual desta escala fica comprometida, conforme pode ser visualizada a seguir:

1 2 3 4 5 6
1 2 3 4 5 6

Na realidade, as escalas numéricas raramente têm intervalos iguais de acordo com a interpretação dos corretores. Do ponto de vista dos candidatos, a chance de obter resultados melhores depende de aperfeiçoamentos diferentes no nível de habilidade exigido. Para uma maior clareza e compreensão da diferença do peso de cada número em uma escala numérica, McNamara (1996) representa essa problemática em uma escadaria com a escala numérica de dificuldades a seguir (figura 2); a escada à esquerda (A) apresenta degraus de tamanhos variados e, portanto, pesos variados, quando uma escala numérica intervalar deveria ser representada por degraus simétricos, conforme os degraus da escada ao lado direito (B).

FIGURA 2 – ESTRUTURA DE DEGRAUS REAL E APARENTES DE UMA ESCALA DE CORREÇÕES DE 0 A 6



ESQUERDA: (A) Representação real da estrutura de degraus conforme notas atribuídas por um corretor específico (degraus de tamanhos desiguais)

DIREITA: (B) Estrutura aparente dos degraus de um critério de correção (degraus de tamanhos iguais)

Fonte: McNamara, 1996 (p. 125).

Os corretores também podem divergir na interpretação de um critério de correção no que tange à variedade numérica da escala, por exemplo, alguns evitam os números das extremidades da escala, característica conhecida como *tendência*

central, enquanto outros evitam números do centro da escala, não atribuindo a nota 3, por exemplo.

- D. Finalmente, corretores podem também diferir em termos de sua consistência (ou inconsistência) em relação aos outros corretores. Os padrões de notas atribuídas por um corretor para um mesmo candidato podem não conferir com os demais corretores por questões aleatórias. Nesse caso, o corretor deve ser retreinado ou até mesmo excluído do processo de correção.

Por meio do treinamento dos corretores sobre como interpretar os critérios de correção, a prática da aplicação dos critérios às composições reais e a troca de experiências são algumas medidas na busca da minimização das diferenças. Recentemente, a natureza e os efeitos do processo de treinamento dos corretores foram revistos. McIntyre (1993), Weigle (1994) e Shohamy (1992) examinaram os efeitos do treinamento e demonstraram que a consistência dos corretores geralmente é otimizada, podendo reduzir os erros aleatórios nas correções, e também que o treinamento reduz a variabilidade em relação ao grau de rigorosidade do corretor, sem, no entanto, eliminar o problema. De acordo com Lunz e Stahl (1990), ao avaliar a escrita, os avaliadores empregam percepções únicas e expectativas pessoais, que não são facilmente alteradas pelo treinamento.

O objetivo do treinamento dos corretores é torná-los internamente consistentes para que suas características possam ser demonstradas pela estatística. Mas, além disso, deve-se aceitar que existem variabilidades estáveis nas características dos corretores que devem ser compensadas de alguma maneira, seja por meio de avaliações múltiplas, reavaliando-se a partir daí os resultados, seja pela redistribuição da nota, seja utilizando uma técnica de análise multifacetada mais sofisticada, como o programa de software FACETS de Linacre e Wright (1992).

B. Outras Facetas da Avaliação

Outros aspectos importantes podem ocasionar a variabilidade nos resultados de

testes de desempenho. Em avaliações que pressupõem interação direta com o corretor, vários fatores aparecem como reações diferenciadas com alguns candidatos ou a interferência do sexo do candidato ou do avaliador, mesmo que inconscientemente. O horário em que a avaliação ocorre pode ser um fator determinante; por exemplo, no período da manhã, tanto candidatos quanto corretores podem estar mais descansados e este fator influenciará diretamente nos resultados. O local onde a avaliação ocorre pode também influenciar; por exemplo, um ambiente com iluminação adequada, silencioso, espaçoso, arejado, organizado e limpo, pode determinar o resultado de uma avaliação, uma vez que interfere na concentração, no conforto e no bem-estar tanto do candidato quanto do avaliador. O sistema nervoso do candidato, por se tratar de um momento de avaliação, pode igualmente ser um aspecto de efeito negativo sobre os resultados.

Cada um dos aspectos do contexto da testagem, entre outros, é denominado de *faceta*, e é possível pesquisar sobre os seus efeitos, ou uma combinação deles, uma vez que essa informação esteja disponível, já que ela pode ser incorporada automaticamente às estimativas da habilidade do candidato. Em outras palavras, a *medida* da habilidade resulta de um ajuste automático da nota bruta de um candidato e a influência das facetas nesta nota. O tipo de ajuste a ser feito e como este deve ser feito são problemas tratados na teoria de mensuração multifacetada. Há um programa de computador denominado FACETS (Linacre e Wright, 1992) que possibilita a análise desta informação, e os ajustes adequados são feitos de forma automática.

1.4.2 A Mensuração Multifacetada

A mensuração multifacetada de Rasch fornece informações precisas, baseadas em dados estatísticos, sobre a influência do corretor e de outras características da avaliação na habilidade estimada em contextos de avaliação de desempenho. McNamara usa os dados do “Access:test” para exemplificar o modelo de análise multifacetada. Os testes de habilidade oral, chamados de “Access: test”, fazem parte de um projeto de testagem de inglês como segunda língua, aplicado pelo governo australiano em 1993

para candidatas a imigração, cujos dados foram retirados de McNamara e Lynch (1993).

Nesse país, o “Access: test” é utilizado para fornecer informações sobre a proficiência lingüística de candidatas à imigração e envolve a avaliação de quatro macro-habilidades, fluência, recursos gramaticais, adequação e vocabulário. Existem dois formatos de testes, que são aplicados em diversos locais no mundo, formato A, uma entrevista gravada com um entrevistador treinado, e formato B, respostas gravadas para estímulos gravados em uma fita. Em ambos os formatos, os candidatos são gravados e o seu desempenho é avaliado na Austrália por avaliadores treinados.

Participaram da análise desse projeto 83 candidatas que responderam a 23 itens, cada um avaliado por quatro corretores diferentes. Em síntese, foram investigadas três facetas do contexto de avaliação do “Access: test”, candidatas, itens e corretores. Considerando-se o número de itens (23), existem diversas fontes de desempenho dos candidatos; cada item foi também avaliado por quatro corretores e respondido por 83 candidatas, fornecendo inúmeras fontes de análise; e finalmente para cada avaliador, 83 candidatas com respostas dadas a 23 itens, ou seja, existem números suficientes para que se considere o estudo significativo.

Os dados coletados para análise são colocados em uma matriz, conforme a Tabela apresentada por McNamara (1996).

TABELA 1 – MATRIZ DE DADOS

CANDIDATO	ITEM																
	1				2				3				...	23			
	AVALIADOR				AVALIADOR				AVALIADOR				...	AVALIADOR			
	A	B	C	D	A	B	C	D	A	B	C	D	...	A	B	C	D
1	2	3	3	3	3	3	2	4	4	3	3	4	...	3	1	1	2
2	4	5	5	5	5	5	4	4	6	5	6	6	...	5	4	3	4
3	1	2	2	1	2	3	2	3	3	2	2	3	...	2	1	3	2
.																	
.																	
.																	
83	6	4	4	5	6	5	5	5	6	5	4	4	...	4	6	6	5

FONTE: MCNAMARA, 1996 (p. 133)

Analisando a matriz de dados é possível ver cada correção como uma função da interação das três facetas ou fatores: a *habilidade* do *candidato*, a *dificuldade* do *item* e as *características* do *corretor*. Percebe-se também que cada célula da matriz não está isolada, mas sim em combinação com o restante da matriz, ou seja, para cada combinação de candidato, item e corretor, existem informações na matriz relevantes à probabilidade de qualquer nota individual. Por exemplo, é possível saber sobre as características do corretor a partir de todas as notas por ele atribuídas que constam na matriz de dados. Estas mesmas informações indicam a habilidade dos candidatos, pois todas as suas respostas estão quantificadas, assim como os dados sobre a dificuldade dos itens, a partir de todas as respostas dos candidatos. Portanto, pela matriz de dados do programa de mensuração multifacetada, é possível prever o provável resultado das combinações do candidato, item e corretor, bem como avaliar a precisão desta previsão.

Ainda quanto aos dados da matriz, é possível, no modelo de Rasch, cruzar os dados disponíveis em uma relação única, expressa em termos do efeito que os resultados podem ter nas probabilidades de sucesso de um candidato, ou seja, é possível precisar que tipo de desafio o candidato encontrará a cada item, com cada corretor, assim como interpretar os resultados obtidos. Da mesma forma, é possível expressar a dificuldade dos itens de acordo com a probabilidade de um candidato com habilidade X de conseguir um resultado Y, ou mais alto, com um corretor mais severo naquele item em específico. Finalmente, pode-se também, mediante o programa FACETS, expressar a rigorosidade dos corretores em termos das chances do corretor atribuir uma nota X (ou maior) em certo item a um candidato com habilidade Y. As facetas podem ser todas unificadas em uma moldura de referência única, a relação entre elas pode ser mapeada, como o Quadro 4 (McNamara, 1996), que mostra um mapa das facetas dos dados do “Access: test.”

QUADRO 4 – MAPA DAS FACETAS SOBRE OS DADOS DO “ACCESS: TEST”.

Measure (Logits)	Candidate	Rater	Item
6	24 40 94 90		
5	28 37 77		
4	1 53 9		
3	27 46 42 61 93 14 78 86		
2	20 62 7 73 82 89 23 68 31 65 17 32 4 56 6 83		
1	11 36 39 45 15 29 85 12 30 22 34 79 81 19 5 55 64 67 69 13 33 49 8 44 63 75 3 54 92 66 80	B A	9 11 19 14 5 6 17 2 12 22 3
0	2 21 52 72 84 87 91 26 59 10 41	D C	8 20 15 10 21 7 13 16 14 18 23
-1	16 51		
-2	25 57 88		
-3	18		
-4	4 7		
Measure (Logits)	Candidate	Rater	Item

FONTE: MCNAMARA, 1996 (p. 135)

O Quadro 4 mostra as habilidades relativas dos candidatos, a rigorosidade relativa dos corretores e a dificuldade relativa dos itens. A informação do quadro é dada numa escala das probabilidades de sucesso dos candidatos e do grau de desafio apresentado por certos corretores e certos itens. Existem quatro colunas no Quadro 4, uma para a escala de mensuração utilizada e uma para cada faceta, *candidato*, *corretor* e *item*. Candidatos, corretores e itens são alocados em pontos específicos da escala na coluna 1, que funciona como uma “régua” que permite localizar a habilidade de cada

candidato, a rigorosidade de cada avaliador e a dificuldade de cada item. A escala de mensuração é utilizada para relatar as probabilidades estimadas de respostas dos candidatos sob várias condições de mensuração, ou cada faceta que foi colocada para análise. As probabilidades são expressas em unidades chamadas “logits” por meio de uma transformação simples como a de porcentagem feita automaticamente pelo programa com o objetivo de trabalhar com números reduzidos, mais significativos e conclusivos. A média de dificuldade dos itens nesta escala de “logits” está no número zero, que representa 50% ou a média; portanto os números negativos são mais fáceis do que a média e os positivos, mais difíceis. Já para os candidatos e avaliadores, quanto mais alto na escala, mais hábeis e rigorosos, respectivamente. Portanto, percebe-se pelos números do Quadro 4 que os candidatos mais habilitados são os números 24, 40 e 94, conseqüentemente mais altos na escala e que o candidato número 47 é o menos habilitado. A grande maioria dos candidatos está entre a média de acertos e habilidade ou um pouco acima entre 0 e 3, que se localiza no centro da escala. No que se diz respeito aos corretores, percebe-se que o corretor B é mais rigoroso e que C é o menos rigoroso, enquanto A demonstra ser o corretor mais neutro; no entanto, é visível na escala a diferença no grau de exigência entre eles. Em relação aos itens da prova, nota-se que a questão 9 é a mais difícil e a questão 23 a mais fácil, porém todos os itens localizam-se em torno do centro da escala, fator que demonstra padronização entre o nível de dificuldade das questões.

A análise multifacetada executada pelo programa FACETS fornece também informações detalhadas em quadros separados sobre cada uma das facetas em estudo, ou seja, habilidade do candidato, dificuldade do item e rigorosidade do corretor. A Tabela 2, McNamara (1996), demonstra o tipo de informações fornecidas sobre os candidatos. A habilidade estimada de cada um é representada em “logits”, e candidatos com 50% de probabilidade de sucesso no item em questão estão no número zero, candidatos acima de zero consideram a questão mais fácil e os abaixo de zero a consideram, mais difícil. Por se tratarem de extrapolações a partir dos dados disponíveis, as estimativas de habilidade estão sujeitas a erros, conforme dados

fornecidos na coluna 3; a margem de erro dependerá da quantidade de informação da matriz de dados, em geral, e quanto maior o número de itens e de corretores, menor a margem de erro. Na coluna 4 existem dados sobre até que ponto o modelo serve como base para estimar as notas observadas de cada candidato em todos os itens do teste. Esta informação é expressa em graus de correlação ou *fit* (adequação) entre as expectativas do modelo e os dados reais de cada candidato, em cada item.

TABELA 2 – RELATÓRIO SOBRE AS MEDIDAS DAS HABILIDADES DOS CANDIDATOS NO “ACCESS: TEST”

Candidate	Ability Estimate	Error	Fit	
			(Infit mean square)	(Standardized infit)
1	4.18	0.18	1.2	1
2	-0.06	0.16	0.8	-1
3	0.43	0.16	0.6	-2
4	2.24	0.17	0.9	0
5	0.99	0.16	1.1	0
6	2.12	0.17	1.6*	3*
7	2.70	0.17	0.8	-1
8	0.72	0.16	0.7	-2
9	3.93	0.18	1.3	2
10	-0.32	0.16	0.9	0
11	2.07	0.17	1.2	1
12	1.45	0.17	0.9	0
13	0.72	0.16	1.2	0
14	3.41	0.17	1.0	0
15	1.64	0.17	0.9	0
.
.
85	1.56	0.17	1.0	0
86	3.23	0.17	0.9	0
87	-0.09	0.16	2.2*	5*
88	-2.69	0.16	0.9	0
89	2.79	0.17	1.0	0
90	5.41	0.20	0.9	0
91	-0.09	0.16	0.9	0
92	0.43	0.16	0.9	0
93	3.65	0.17	1.0	0
94	5.62	0.21	0.9	0
Candidate	Ability Estimate	Error	Fit	
Mean (n = 83)	1.57	0.17	1.0	-0.1
SD	1.96	0.01	0.3	1.7

Reliability (reliable difference in ability of candidates) 0.99

* = unsatisfactory fit (misfit)

FONTE: MCNAMARA,1996 (p. 137)

Existem também resumos das várias estimativas fornecidas, a média e o desvio padrão das estimativas de habilidade, erro e adequação, assim como o quanto o

teste define diferentes níveis de habilidade, ou seja, a capacidade do teste de fazer distinção e discriminar os candidatos. Na Tabela 3, de McNamara (1996), são apresentadas as estimativas de dificuldade de cada item, da mesma maneira que na Tabela 2 foram apresentadas as estimativas dos candidatos.

Conforme a Tabela 3 a seguir, percebe-se a dificuldade estimada dos itens. No entanto, na tabela original de McNamara, 1996 (p. 38), houve um equívoco do autor, que trocou as palavras “Item” por “Candidate” e “Difficulty” por “Ability”. A discussão no livro, posterior à tabela, na página 139, demonstra a troca das palavras e ilustra claramente o conteúdo da tabela.

TABELA 3 – RELATÓRIO SOBRE A DIFICULDADE ESTIMADA DOS ITENS DO “ACCESS: TEST”

Item	Difficulty Estimate	Error	Fit	
			(Infit mean square)	(Standardized infit)
1	-0.81	0.09	1.0	0
2	0.46	0.09	0.8	-2
3	0.15	0.09	1.0	0
4	-0.79	0.09	0.8	-2
5	0.67	0.09	0.8	-2
6	0.52	0.09	0.8	-2
7	-0.45	0.09	1.0	0
8	-0.09	0.09	1.0	0
9	1.20	0.09	0.8	-2
10	-0.10	0.09	1.1	0
11	0.85	0.09	0.9	-1
12	0.42	0.09	1.9*	9*
13	-0.59	0.09	0.8	-2
14	0.56	0.09	1.0	0
15	-0.09	0.09	1.4	5*
16	-0.67	0.09	0.7	-3
17	0.63	0.09	0.7	-4
18	-0.91	0.09	0.9	-1
19	0.56	0.09	0.9	-1
20	-0.00	0.09	0.9	0
21	-0.38	0.09	1.3	3
22	0.11	0.09	0.9	0
23	-1.26	0.09	1.6*	6*

Item	Difficulty Estimate	Error	Fit	
			(Infit mean square)	(Standardized infit)
Mean (<i>n</i> = 23)	0.00	0.09	1.0	-0.1
SD	1.96	0.01	0.3	1.7

Reliability (reliable difference in difficulty of items) 0.98

* = unsatisfactory fit (misfit)

FONTE: MCNAMARA, 1996 (p. 138)

Por fim, são fornecidas pelo programa FACETS informações sobre as características de rigorosidade ou benevolência dos corretores. Na Tabela 4, de McNamara (1996), os corretores são identificados na coluna 1 e uma estimativa da sua rigorosidade é apresentada na coluna 2, novamente em termos da probabilidade de atingir um certo resultado com um corretor em específico. Os desvios dessas estimativas são apresentados na coluna 3 e as estatísticas de adequação na coluna 4, que neste caso indica a consistência relativa na avaliação dos corretores. A falta de consistência é indicada por valores altos, e obviamente apresenta problemas para os resultados e desafios à análise.

TABELA 4 – RELATÓRIO SOBRE A RIGOROSIDADE ESTIMADA DOS CORRETORES DO “ACCESS: TEST”

Rater	Severity Estimate	Error	Fit	
			(Infit mean square)	(Standardized infit)
A	0.41	0.04	1.0	0
B	1.40	0.04	1.2	5
C	-1.03	0.04	1.0	1
D	-0.77	0.04	0.8	-7
Mean ($n = 23$)	0.00	0.04	1.0	-0.2
SD	0.97	0.00	0.1	4.8
Reliability (reliable difference in severity of raters) 0.99				

FONTE: MCNAMARA, 1996 (p. 140)

As condições operacionais da análise multifacetada devem respeitar ou prever alguns requisitos para garantir bons resultados, conforme descrição a seguir. De acordo com McNamara, o primeiro deles aponta que, ao desenvolver o plano de coleta de dados, é importante prever correções múltiplas de cada elemento da faceta em análise. Segundo, os modelos que envolvem muitas facetas na análise precisarão de conjuntos de dados maiores para garantir estabilidade na estimativa das facetas. O terceiro requisito consiste na necessidade de análises recentes a cada administração de testes, pois não se pode presumir que as características dos corretores permaneçam

constantes a cada nova avaliação. O quarto aspecto diz respeito à possibilidade de aumentar os conjuntos de dados mediante a combinação de dados antigos com dados de análises recentes, contanto que as matrizes apresentem semelhanças que permitam comparações entre as duas análises. Outro aspecto importante está relacionado ao tempo, ou seja, deve-se calcular o tempo necessário para proceder às avaliações, efetuar as correções e permitir as análises e transformações dos dados para calcular as medidas. De modo geral, o uso rotineiro da análise multifacetada é restrito a testes de avaliações de alto risco, tanto para os candidatos, quanto para os administradores, e para tanto é preciso que existam recursos básicos e conhecimento suficiente por parte dos aplicadores para tornar o processo ágil. A desvantagem da aplicação da medida multifacetada é que a análise exige um número relativamente grande de conjuntos de dados.

O embasamento teórico apresentado neste capítulo contribuiu para a reflexão sobre os aspectos relevantes das fontes de subjetividade na correção da escrita, assim como apontou medidas que podem ser tomadas no sentido de minimizar os seus efeitos nos resultados. A partir das medidas apresentadas nesta revisão, serão analisados os dados coletados para o estudo empírico desta pesquisa.

2. METODOLOGIA

O estudo foi composto por duas fases diferenciadas. A primeira consiste em uma revisão bibliográfica, cuja principal finalidade foi levantar na literatura existente os problemas discutidos sobre a avaliação da escrita em língua estrangeira, assim como propostas correntes e mais utilizadas. Neste estudo foram relatadas questões sobre a avaliação, a avaliação de desempenho – no caso a escrita e também aspectos relacionadas às fontes de variabilidade na correção, ou seja, as diversas facetas que surgem no momento da avaliação.

Em sua segunda fase, este estudo recebeu um tratamento empírico no qual foram apresentadas as avaliações das composições redigidas por alunos do curso de Letras Português-Inglês da Pontifícia Universidade Católica do Paraná. Os sujeitos desse estudo foram 28 alunos brasileiros de inglês como língua estrangeira, cursando, na época da coleta de dados, o primeiro e terceiros períodos do programa de Letras, cuja faixa etária variou entre 17 e 32 anos. São alunos com nível de língua inglesa pós-intermediário, e que foram submetidos a teste de nivelamento no início do curso. No entanto, a diferença de nivelamento dos alunos explicitada nos textos escritos é devida ao fato de que alguns estudaram inglês antes de ingressarem no curso universitário, enquanto outros classificados para o mesmo grupo já lecionam a língua inglesa.

Todos os sujeitos escreveram um texto de gênero descritivo a partir de um modelo estudado em sala de aula e entregaram esse produto como avaliação escrita do 1.º semestre de 2004 (ver anexo 4). Alguns desses 28 alunos, especificamente 11 deles do 3.º período, escreveram sobre dois temas diferentes (ver anexo 4) para possibilitar um estudo da variabilidade relacionada à dificuldade do item, conforme descrito posteriormente. Esclareço que obtive permissão expressa por escrito para utilizar as composições deste estudo. A avaliação destes textos foi conduzida por três corretores seguindo um critério holístico, que, conforme as características já apresentadas no decorrer deste estudo, se mostrou mais adequado.

Os alunos receberam exercícios diversos e técnicas sobre como redigir parágrafos, material que fazia parte do conteúdo para as aulas de escrita do 1º semestre de 2004, data em que as composições foram coletadas. A intenção de trabalhar bem com os alunos a estruturação de um parágrafo teve o objetivo de auxiliá-los a organizar suas idéias, uma vez que os sujeitos deste estudo apresentavam bom conhecimento de linguagem para redigir parágrafos. Os exercícios dados aos alunos como preparação para a escrita foram adaptados de dois livros de inglês, um com enfoque em redação acadêmica e outro cujo enfoque era o ensino das quatro habilidades. O gênero trabalhado neste período foi descritivo e os exemplos analisados foram um parágrafo sobre “Meu Colega de Apartamento” – “My Roommate Rita” e outra composição sobre a “Descrição de um lugar” – “Description of a Place”. Durante as aulas em que os textos modelos foram analisados, os alunos reviram a estruturação de parágrafos, assim como a estrutura de um texto do gênero descritivo. A partir desses exercícios os alunos foram solicitados a redigir textos semelhantes, em momentos diferentes, sobre os dois temas dados. As atividades foram feitas em sala de aula, os alunos dispunham de dicionários para auxiliá-los durante o processo da escrita e tinham conhecimento de que o modelo deveria ser utilizado apenas como apoio, mas que não deveriam copiá-lo.

O critério selecionado para avaliar as composições deste estudo é holístico (ver Apêndice I), pois conforme explicitado na sessão 2.2, enfoca o aspecto positivo da produção escrita dos alunos, propõe uma visão global do texto, oferece maior praticidade para o professor ao corrigir e, portanto, viabiliza a aplicação de testes escritos para professores, mesmo que esses tenham muitas turmas e muitos alunos. O referido critério holístico foi adaptado a partir de um critério de avaliação elaborado por professores de inglês como segunda língua de uma escola pública em Fairfax County, Virginia, EUA (ver anexo 3). Trata-se de uma escala numérica de seis níveis, variando de Elementar a Fluente, compreendendo macro-habilidades relacionadas ao desenvolvimento de idéias, a organização, a fluência escrita, a utilização adequada de estruturas sintáticas, a escolha e variedade de vocabulário no decorrer do texto e aspectos gerais sobre o conteúdo.

Seguindo o critério de correção, o corretor seleciona notas de 1 a 6 de acordo

com a melhor descrição o desempenho do aluno. A cada valor numérico do critério de correção, o nível de exigência em relação ao desempenho do aluno aumenta. O produto apresentado não precisa obedecer a todas as condições nas quatro dimensões uma vez que seu resultado é estipulado a partir de uma adequação geral da composição aos critérios descritos em um dos valores numéricos. O fato de a nota atribuída variar entre 1 e 6 contribui para uma avaliação numérica mais objetiva das composições pelo fato da escala numérica conter menos números e possibilitar uma atribuição de notas baseada em critérios mais objetivos. Dois exemplos de critérios com valores numéricos reduzidos que reforçaram a opção por uma escala numérica menor nesta pesquisa, são os testes FCE, com variação de 0 a 5, e TOEFL, cuja variação é de 0 a 6 (critérios nos anexos 1 e 2 para ilustração).

Os corretores selecionados são profissionais experientes no ensino da língua inglesa, todos lecionam em cursos livres, uma delas na universidade e alguns atuam na área de treinamento de professores, seleção de materiais e elaboração de testes. Portanto, o nível de conhecimento e experiência entre os corretores é semelhante, um fator que deve contribuir para maior confiabilidade na sua atuação ao corrigir as redações. Outro fator importante a ser mencionado é que os corretores não conhecem os alunos e, sendo assim, não tiveram julgamentos pré-concebidos sobre os seus desempenhos.

As composições foram corrigidas por três corretores distintos para fornecer resultados mais confiáveis e tornar possível a análise dos resultados divergentes. Os corretores receberam treinamento sobre como usar o critério de correção preestabelecido e responderam a um questionário (ver apêndice 2) sobre seu histórico profissional, sua experiência na avaliação da escrita, seus sentimentos e pensamentos em relação ao momento da correção, a estrutura da correção e os critérios. Por meio deste questionário será possível detectar algumas diferenças entre as experiências e expectativas dos corretores que venham a ser significativas na explicação de algumas divergências nos resultados.

A análise dos resultados foi feita em etapas distintas e o objetivo de efetuar a análise em etapas diz respeito ao tema da dissertação, ou seja, a busca por

alternativas que pudessem minimizar os efeitos da subjetividade na avaliação da escrita. Na primeira etapa foram analisadas as características dos corretores A, B e C, a partir dos questionários respondidos e dos resultados das avaliações. Em seguida foram analisadas as divergências entre os resultados dos corretores nas avaliações de A e B às 28 composições de número 1 e analisada a transcrição da sua discussão acerca das correções, em busca de uma padronização. Em um terceiro momento foram comparadas as correções das composições 1 e divergências intercorretores A, B e C, analisadas as características que interferiram nos resultados e calculada a partir das três correções uma média entre os resultados para que fosse possível quantificar a divergência que cada corretor apresentou do resultado médio. Por fim, os resultados das avaliações das composições 1 e 2 pelos corretores A, B e C foram comparados e houve uma reflexão sobre dois aspectos: quanto o fator de escolha de um tema pode interferir nos desempenhos dos alunos e na interpretação e avaliação dos corretores, e como o modelo multifacetado de Rasch, explicado no capítulo anterior, poderia ser utilizado como alternativa de minimizar a subjetividade. Não se pretende com este estudo demonstrar a utilização do programa de software FACETS (Linacre e Wright, 1992), apenas mostrar a teoria sobre a sua utilização e demonstrar por meio de avaliações diversas às composições, a existência da subjetividade e algumas medidas que possam melhorar os resultados auferidos.

O presente estudo apresenta algumas limitações, que foram apontadas no desenvolvimento da pesquisa, mas que se referiram a) ao número reduzido de textos compostos pelos alunos e que receberam o tratamento empírico, impossibilitando resultados concretos, e sim demonstrando a aplicabilidade do modelo na mensuração de notas; b) ao critério de avaliação utilizado não é exaustivo, e certamente deixou alguns aspectos da avaliação da escrita desatendidos, uma vez que proporciona aos alunos uma resposta global da sua composição, enquanto um critério analítico lhes proporcionaria um maior retorno em termos de estruturação gramatical, no entanto, exigindo maior tempo e dedicação dos avaliadores, tornando o estudo menos prático; c) à diversidade na

competência lingüística de alguns alunos, produzindo resultados discrepantes em alguns casos, impossibilitando generalizações em torno dos resultados.

3. ANÁLISE DOS RESULTADOS

A análise empírica desta pesquisa envolve 28 composições de número 1, sobre o tema “My Roommate Rita”, e 11 composições de número 2, sobre o tema “Description of a Place”, portanto 11 alunos escreveram duas composições diferentes. No decorrer da coleta de dados para a pesquisa empírica não houve tempo hábil para coletar 2 composições de todos os alunos, uma vez que alguns deles já estavam em outras turmas. As redações foram corrigidas e avaliadas por três corretores diferentes, A, B e C, e os resultados estão analisados em etapas distintas, conforme as características dos corretores, a interpretação dos critérios, o grau de dificuldade das composições e a relação entre os resultados e os apontamentos teóricos da revisão de literatura.

Ao desenvolver um estudo sobre a avaliação, diversos aspectos devem ser ponderados para que se torne possível verificar os fatores que podem influenciar nos resultados dos alunos. Entre os fatores que interferem bastante nos resultados estão os corretores, suas expectativas em relação ao instrumento da avaliação, sua compreensão dos critérios de correção, assim como fatores externos que possam alterar as notas auferidas.

Nesta pesquisa, com o objetivo de conhecer melhor as características de cada corretor, suas experiências com a avaliação de modo geral, e outros dados sobre a correção das composições deste estudo, os corretores A, B e C responderam a um questionário, cujas respostas serviram como base para traçar um perfil de cada corretor e ressaltar aspectos importantes das suas correções (anexo 6).

O Quadro 5 a seguir apresenta os resultados das avaliações efetuadas pelos corretores A, B e C, com a distribuição de notas atribuídas para as 28 composições de número 1.

QUADRO 5 – DISTRIBUIÇÃO DAS NOTAS ATRIBUÍDAS PELOS CORRETORES, SEGUNDO VALOR NUMÉRICO DA ESCALA DE NOTAS

Valor numérico da escala de notas	Número de vezes que cada nota foi atribuída pelos corretores		
	Corretor A	Corretor B	Corretor C
Nota 1	0	0	0
Nota 2	12	0	3
Nota 3	8	7	7
Nota 4	5	7	6
Nota 5	1	7	9
Nota 6	2	7	3

Apesar de o quadro apresentado não permitir verificar a quais composições foram atribuídas as referidas notas, ou a diferença entre as notas dadas pelos corretores A, B e C para cada composição, o quadro é representativo das características e do grau de rigorosidade ou benevolência de cada corretor. Percebe-se que o corretor A é mais rigoroso e atribui um maior número de notas 2 e 3 e quase nenhuma nota entre 5 e 6, enquanto o corretor B não atribui nenhuma nota 2 e conferiu sete notas 6, demonstrando pela distribuição de suas notas, todas em números de 7, entre os números 3, 4, 5 e 6, estar em menor sintonia com o grau de rigorosidade do corretor A. Já o corretor C distribuiu suas notas entre 2 e 6, com três composições de valor 2 e 6 cada e distribuiu o restante de suas notas entre 3, 4 e 5. Não foi atribuída nenhuma nota 1 para nenhuma composição, por nenhum corretor, e conferindo a qualidade das composições no anexo 4, verifica-se que as produções são superiores ao nível descrito na escala para a nota 1, o que demonstra um consenso entre os corretores pelo fato de não atribuírem notas baixas.

As notas atribuídas pelos corretores A, B e C para as 28 composições de número 1 dentro da escala de 1 a 6, demonstram as características de rigorosidade ou benevolência de cada corretor. A partir do número de vezes que cada corretor atribui uma nota 3, por exemplo, é possível calcular o percentual de cada nota atribuída, permitindo a comparação entre os corretores e suas correções.

O Quadro 6 a seguir quantifica em valores percentuais, quantas vezes os corretores A, B e C auferiram cada nota da escala de 1 a 6.

QUADRO 6– DEMONSTRATIVO DO NÚMERO DE VEZES QUE CADA CORRETOR ATRIBUIU AS NOTAS DE 1 A 6 E SUA REPRESENTAÇÃO EM PORCENTAGEM

Corretor	Nota 1	%	Nota 2	%	Nota 3	%	Nota 4	%	Nota 5	%	Nota 6	%
A	0	0	12	42,85%	8	28,57%	5	17,85%	1	3,57%	2	7,14%
B	0	0	0	0	7	25%	7	25%	7	25%	7	25%
C	0	0	3	10,71%	7	25%	6	21,42%	9	32,14%	3	10,71%

Percebe-se que o corretor A é mais rigoroso do que os corretores B e C, uma vez que atribuiu 42,85% das suas notas 2 na escala, enquanto B não atribuiu notas 2 e C, apenas 10,71%. No que diz respeito a notas altas, A atribuiu apenas 7,14% de notas 6, enquanto B atribuiu 25% de notas 6 e C 10,71%. O corretor A distribuiu suas notas na sua maioria entre 2 com 42,85% e 3 com 28,57%, restando poucas composições entre 4 com 17,85%, 5 com 3,57% e 6 com 7,14%, aspecto que revela rigorosidade e não tendência central, pois seus resultados não se concentram nas notas intermediárias.

As correções efetuadas pelo corretor B apontam com maior objetividade sua benevolência uma vez que os resultados das correções estão distribuídos em percentuais iguais, 25%, entre as notas 3, 4, 5 e 6, ou seja, o corretor B não atribuiu notas baixas e atribuiu várias notas máximas, 6.

Nos resultados do corretor C, as notas estão distribuídas entre 2 e 6 na escala, no entanto, é possível perceber um acúmulo dos resultados em torno das notas 3, 4 e 5, com um percentual de 25%, 21,42% e 32,14% para cada nota, respectivamente. Os demais resultados ficam classificados entre as notas 2, com 10,71% e 6, com 10,71% também, demonstrando no corretor C uma tendência de auferir notas em torno de uma pontuação média na escala. De acordo com McNamara (2000), esta tendência é denominada central.

3.1 CRITÉRIOS E OUTROS FATORES DE INFLUÊNCIA NAS NOTAS

A partir das respostas ao questionário, identifica-se que o corretor A considerou com maior ênfase o conteúdo das composições, aspectos relacionados à coesão e coerência, conforme suas colocações nas perguntas A-3, D-3-b-c e D-7, ue dizem respeito ao seu interesse pelo conteúdo das redações, sobre como as pessoas se relacionam com as outras, demonstrando envolvimento com o texto e curiosidade. Outro aspecto de relevância para o corretor A é o uso e o emprego correto das palavras para “enriquecer” o conteúdo, conforme respostas D-4.

O corretor A sentiu-se à vontade em utilizar o critério de correção holístico e demonstrou preferência por ele em relação aos critérios analíticos, conforme respostas número C-3, 4, 5. As composições foram corrigidas a partir de duas leituras, no período da manhã, sem interrupções nas duas correções, e este aspecto permitiu ao corretor A coerência e segurança na atribuição das notas. Percebe-se sua característica pessoal mais rigorosa nos seus resultados e em nenhum momento o corretor demonstra dúvida em relação às notas dadas. O fato de não conhecer os nomes dos alunos para o corretor A foi positivo, pois não gerou pré-concepções sobre a habilidade dos alunos ao redigir ou interferências de ordem pessoal, demonstrando a imparcialidade do corretor.

O corretor B é aparentemente mais benevolente em suas correções, considerando-se que não atribuiu nenhuma nota 2 e auferiu todas as suas notas entre 3 e 6, coincidentemente de forma balanceada. Segundo as respostas do seu questionário, o próprio corretor revela-se “permissivo”, pois se sente impotente ao avaliar composições cujos escritores não receberão *feedback* sobre o seu desempenho. Este aspecto da preocupação do corretor B em ser justo em suas avaliações revela-se em sua insegurança ao atribuir notas e acredita-se que esta insegurança seja um dos fatores causador da sua benevolência, conforme verificado nas respostas B-3, C-2, D-3-B e C.

Em relação aos critérios de correção, o corretor B demonstrou entendimento e aceitação, mas, de acordo com suas características pessoais, preferiu os conceitos

(proficiente, fluente etc.) aos números, demonstrando sua dificuldade em atribuir notas numéricas que não lhe representam avaliações significativas da habilidade de um escritor. Observa-se também pela resposta C-1 que o corretor já atribuiu notas numéricas sem critérios preestabelecidos, a partir de sua compreensão do que determina uma habilidade em uma escala numérica de 1 a 10. Este aspecto demonstra que, além de suas dificuldades relacionadas à falta de *feedback*, o corretor não acredita numa avaliação por meio de uma escala de 1 a 6, na medida em que se sente confortável com notas de 0 a 10 e em que faz comparações sobre que valores teriam as notas 4 e 6 em uma escala numérica de 0 a 10, conforme respostas D-8 e 9. Portanto, as escalas alfabéticas, numéricas, ou de conceitos, entre outras, podem ter grau de utilização maior ou menor, dependendo da adaptação pessoal de cada corretor ou até mesmo da sua experiência com notas como aluno, professor e avaliador.

O corretor B corrigiu suas composições após o almoço, ou seja, já havia trabalhado um pouco e poderia estar cansado devido ao momento do dia, assim como efetuou as correções em momentos diferentes, com pausas e retomadas que provavelmente interferiram na sua capacidade de obter uma visão global de todas as correções. O corretor demonstrou curiosidade pelo conteúdo das composições, conforme resposta D-7, assim como sua correção demonstra maior preocupação com o efeito dos resultados nos escritores e não com o processo de avaliação como atribuição de notas. Provavelmente o corretor B trabalhe a escrita de forma mais adequada como processo, no qual o aluno tem a oportunidade de rever e refletir sobre seus erros.

O corretor C, diferentemente de A e B, atribuiu suas notas de forma mais centralizada, ou seja, em sua maioria as composições foram classificadas entre as notas 3, 4 e 5, fator que evidencia uma tendência por parte do corretor a atribuir notas aproximadas da média, não sendo rigoroso ou benevolente em demasia. A tendência a atribuir notas em torno de uma média pode demonstrar certo desvio dos critérios, a essa característica McNamara (1996) denomina tendência central.

Em seu questionário o corretor declara-se mais exigente em relação ao conteúdo do que à estrutura e a mecânica, demonstrando compreensão do critério de

correção, mas respeitando sua visão sobre a qualidade de uma composição, conforme suas respostas às perguntas B-3 e C-2. O corretor acredita na praticidade do critério holístico e acha positivo que a correção leve menos tempo; no entanto julga que cada elemento da escrita deveria ser avaliado separadamente, seguindo critérios analíticos, e considera pouco produtivo não poder traçar uma média que faça uma mistura entre as características de cada nível dos critérios de correção, conforme respostas C-3 e 4 e 5.

Em relação à escala numérica de 1 a 6 e não de 0 a 10, o corretor C declara que sentiu dificuldades para estabelecer diferenças entre os alunos com uma escala menor, ou seja, tanto o corretor B quanto o C transferem sua experiência e preferência por escalas de 0 a 10 na atribuição das notas, ignorando que a descrição dos critérios é que deve referenciar a escolha numérica. Portanto, o corretor C também, embora em menor grau, sentiu-se inseguro com as notas atribuídas e teve sensações de não estar discriminando as habilidades dos alunos e conseqüentemente de estar sendo injusto. O corretor C deu uma pausa entre as correções, contudo, efetuou todas as avaliações em um mesmo dia, nos períodos da tarde e noite, facilitando a coerência entre as notas dadas, pois pôde lembrar-se com clareza dos critérios pessoais que teve em mente no momento das correções.

A partir dos apontamentos dos horários das correções de A, B e C e das pausas efetuadas por eles, observa-se que, tal como aponta McNamara (1996), o período do dia é um fator que pode influenciar nas avaliações.

3.2 DISCUSSÃO E CALIBRAÇÃO DOS CRITÉRIOS

Tendo em mente o objetivo de minimizar a subjetividade na correção da escrita, uma das alternativas que surgem é promover uma discussão entre os corretores na busca de um denominador comum que respeite as avaliações individuais de cada corretor, assim como os critérios de correção utilizados. Embora nem sempre seja viável, por questões de praticidade, promover uma discussão entre os avaliadores, neste estudo pode-se perceber que os resultados de

uma discussão geram alterações nas notas, portanto mais justiça na correção, além de uma reflexão por parte dos corretores, fator que enriquece suas atuações nas avaliações. Este aspecto é ressaltado por Lunz e Stahl (1990), quando dizem que no momento da correção os avaliadores empregam percepções únicas, portanto, a partir de uma discussão podem rever suas concepções no momento da avaliação e atribuir notas com maior coerência nos resultados.

As correções efetuadas pelos corretores A e B nas 28 composições de número 1 apresentaram um índice alto de variabilidade, mantendo as características de maior rigorosidade do corretor A e maior benevolência do corretor B, conforme análise das características individuais de cada corretor. Os resultados apresentados a seguir e análise da discussão entre os corretores, transcrita no anexo 7, servem como ponto de partida para a reflexão sobre as causas destes fatores de variabilidade.

O quadro 7 a seguir expressa os resultados das notas atribuídas na primeira correção e as alterações efetuadas após a discussão.

QUADRO 7 – RESULTADOS DAS CORREÇÕES EFETUADAS PELOS CORRETORES A e B e AS ALTERAÇÕES PÓS-DISCUSSÃO

Composição n.º	1º Nota do Corretor A	Alteração nas notas pelo Corretor A	1º Nota do Corretor B	Alteração nas notas pelo corretor B	Notas que atingiram denominador comum
1	2	–	4	3	–
2	2	–	4	3	–
3	3	–	5	–	–
4	5	–	5	–	–
5	4	–	5	–	–
6	6	–	6	–	–
7	2	–	4	3	–
8	4	–	6	–	–
9	2	–	3	2	2
10	4	–	6	–	–
11	2	–	3	2	2
12	4	–	6	5	–
13	4	5	5	–	5
14	2	3	4	–	–
15	3	4	5	–	–
16	3	–	5	–	–
17	3	–	5	4	–
18	2	–	3	2	2
19	2	–	4	3	–
20	2	–	3	–	–
21	3	–	6	5	–
22	2	–	3	2	2
23	2	3	4	3	3
24	3	4	6	5	–
25	3	–	4	3	3
26	6	–	6	–	–
27	2	–	3	2	2
28	3	–	4	–	–

Ao analisar o quadro de correções, é possível perceber que os resultados da primeira correção, anterior à discussão, foram os mesmos em apenas 03 das 28 composições, números 4, 6 e 26, havendo concordância em apenas 10,7% das notas dadas. Coincidentemente as notas destas composições foram 5 e 6, ou seja, não houve divergências na avaliação das composições fluentes ou proficientes, demonstrando concepção semelhante sobre as características de uma composição de qualidade. Após a discussão entre os corretores, o número de notas em comum aumentou para 11 entre as 28 composições, elevando o percentual de notas iguais entre os corretores A e B de 10,7% para 39,2%. Este fator por si só demonstra a relevância e importância da discussão e os seus efeitos na diminuição da variabilidade.

No início da correção houve divergências de até 3 pontos na escala, ou seja, foram atribuídas notas 3 e 6 para uma mesma composição. O Quadro 8 a seguir quantifica o grau de divergência na escala anterior e posterior à discussão.

QUADRO 8 – GRAU DE DIVERGÊNCIA DAS NOTAS ATRIBUÍDAS POR A e B

Diferença Numérica das notas	Divergência anterior à discussão	Percentual de divergência	Divergência posterior à discussão	Percentual de divergência
Notas iguais / Concordância	3	10,7 %	11	39,2 %
Diferença de 1 ponto na escala	10	35,7 %	12	42,8 %
Diferença de 2 pontos na escala	13	46,4 %	5	17,8 %
Diferença de 3 pontos na escala	2	7,1 %	0	0
Diferença de 4 pontos na escala	0	0	0	0
Total de divergência	25 notas	89,2 %	17 notas	60,6 %

A partir do quadro apresentado percebe-se que as notas tiveram o grau da divergência consideravelmente reduzido. Foram eliminadas as divergências de 3 pontos na escala, cujo percentual era de 7,1% e baixou para 0. As divergências de 2 pontos baixaram de 13 para 5 composições, de 46,4, % para 17,8%, e o maior percentual de variabilidade entre os resultados dos corretores A e B, ficou em divergências de apenas 1 ponto na escala, aumentando de 10 para 12 composições, ou seja, de 35,7% para 42,8%. Os números apresentados neste quadro, apesar de ainda refletirem divergências nas notas,

elucidam os benefícios da discussão entre os corretores, uma vez que o percentual total de divergência foi reduzido de 89,2% para 60,6%. McIntyre (1993), Weigle (1994) e Shohamy et al. (1992) confirmam este aspecto positivo da discussão entre corretores, ao afirmarem que os efeitos do treinamento sobre como avaliar otimizam a consistência entre os corretores e a discussão sobre as notas atribuídas é uma das etapas de um treinamento de corretores.

A transcrição da discussão revela razões concretas para explicar a variabilidade entre os corretores. Nas respostas dos questionários (ver anexo 6) ficou visível a preocupação do corretor A com o conteúdo e a riqueza do vocabulário para atrair o leitor, assim como a importância dada pelo corretor B em alcançar o objetivo de comunicar uma idéia por meio de um texto escrito, além do seu desconforto com a nota numérica sem o acompanhamento de *feedback*. Tais características avaliadas pelos corretores foram mais fortes na atribuição das notas do que as características descritas nos critérios de correção, ou seja, reforçam a colocação de que os corretores não conseguem se desvincular totalmente das suas expectativas pessoais na avaliação.

A transcrição revela também o aspecto da insegurança do corretor B em relação às notas dadas, pois suas notas sofrem 15 alterações, ou seja, 53,8 % das suas notas são modificadas após a discussão, ao passo que o corretor A altera apenas cinco de suas notas, tendo um percentual de 17,8% de alteração, 36% menor que o percentual de alterações do corretor B. As características de rigorosidade do corretor A e benevolência do corretor B também são confirmadas na discussão pela análise das notas alteradas; o corretor A aumenta suas cinco notas por considerar que atribuiu notas abaixo do nível do aluno, enquanto o corretor B abaixa todas as 15 notas, evidenciando sua aceitação em demasia aos aspectos negativos das composições.

Durante a discussão dos corretores A e B ficou aparente que suas preocupações estão relacionadas ao conteúdo das redações, à riqueza de vocabulário, utilização de vocabulário de transição, boa organização e coerência do texto, muito mais evidente do que preocupações com estrutura gramatical, mecânica, ortografia, entre outros erros relacionados à linguagem. No entanto, nas composições 16 e 28, o

corretor A transpõe a preocupação até então imparcial com o conteúdo, colocando-se de maneira a concordar ou discordar com o escritor, ou seja, julgando de formas positiva e negativa, respectivamente, as colocações feitas pelos escritores. Nestas duas composições o corretor A deixa claro que sua visão pessoal do conteúdo do texto interferiu nas notas, demonstrando parcialidade e incoerência com suas avaliações das demais composições e reforçando que os corretores se deixam influenciar por fatores alheios aos critérios ou aos textos propriamente ditos em suas avaliações.

A legibilidade das composições foi um outro aspecto importante; interferiu nos resultados das composições 2, 9, 27 e na conclusão do corretor B, que alega que um texto bem organizado e legível faz diferença no momento da avaliação. De acordo com Hughes, Keeling e Tuck (1983), corretores com letra legível e bem apresentável diminuem significativamente as notas ao avaliarem textos desorganizados e ilegíveis. Charney (1984) acrescenta que, diante da necessidade de ler e avaliar muitos ensaios com rapidez, os corretores acabam recorrendo a características fáceis de se perceber, como a letra, mas que são irrelevantes no que diz respeito à habilidade de redigir. Portanto, os corretores precisam refletir constantemente sobre suas avaliações para não discriminar alunos a partir das suas letras.

Por fim, uma observação feita pelo corretor A em relação à composição 6 reforça a manutenção da cultura da escala de 0 a 10, pois o corretor A diz que a redação foi “extremamente desenvolvida, ele fechou nota 10”, ou seja, ainda existe a percepção de que se um texto é excelente, ele é nota 10, mesmo utilizando uma escala de avaliação de 1 a 6.

Na maioria dos contextos escolares as divergências nos resultados dos alunos devido à não compreensão dos critérios de correção, à falta de oportunidade de discutir os resultados com outros corretores, a fatores relacionados às características de rigorosidade ou benevolência dos corretores, ou à legibilidade das composições, são minimizadas pela sua baixa representatividade numérica no histórico escolar de um aluno. No entanto, em avaliações de alto risco, como os vestibulares, em testes admissionais como o TOEFL (Test of English as a Foreign Language) e o ECPE (Examination for the Certificate of Proficiency in English), nos quais a vida de um

candidato pode ser afetada por poucos décimos de pontuação, a avaliação deve ser efetuada com muito rigor e atenção aos critérios e aos aspectos que possam interferir nos resultados.

Outra proposta para reduzir a subjetividade na avaliação de composições consiste em acrescentar às correções efetuadas por dois corretores (A e B) nas composições de número 1 uma terceira correção (corretor C). Analisam-se as notas dos três corretores e a partir delas calcula-se uma média entre o resultado médio das notas atribuídas pelos corretores e, por último, analisa-se o desvio da média de cada corretor para fazer os ajustes necessários nos seus resultados. Acredita-se que dessa forma seja possível reduzir a variabilidade dos resultados.

A revisão de literatura sobre a confiabilidade da avaliação mostra que a interpretação das características apresentadas em cada escala dos critérios de correção não pode ser presumida, a não ser que os pontos da escala definam claramente os diferentes níveis. De acordo com Bachman (1990), a interpretação exata dos critérios por públicos diferentes provavelmente apresentará variações e Brindley (1998) relaciona tais variações com a experiência prévia dos avaliadores, expectativas inconscientes e preferências subjetivas quando à importância relativa dos diferentes critérios de comunicação. Alguns avaliadores atribuíram notas mais altas às mesmas composições que receberam notas baixas de outro corretor, apesar de seguirem os mesmos critérios de correção. Esta diferença na interpretação dos critérios é que deve instigar a dupla ou tripla correção para alcançar uma nota mais unificada e justa.

O Quadro 9 a seguir traz as avaliações feitas pelos corretores A, B e C e a média de cada corretor calculada a partir do somatório das notas dadas às 28 composições e da divisão por 28. O valor médio dos resultados atribuídos pelo corretor A é 3.03, pelo corretor B é 4.5 e pelo corretor C é 4.07. Apesar de os números não representarem um valor demasiadamente alto, há que se considerar uma média de 3.86 entre os resultados dos corretores A, B e C e que a divergência da média por cada corretor é de 0.83, 0.64 e 0.21, respectivamente. Percebe-se nos resultados dos desvios da média dos corretores A (0.83), B (0.64) e C (0.21), que o corretor A difere em maior grau da média do que B e C e que o retreinamento desses corretores é primordial à segurança dos resultados.

QUADRO 9 – RESULTADOS DAS CORREÇÕES DAS COMPOSIÇÕES NÚMERO 1 EFETUADAS PELOS CORRETORES A, B E C E A MÉDIA DE NOTAS DE CADA CORRETOR AVALIAÇÕES CONSIDERADAS – ANTERIORES À DISCUSSÃO

Composição Nº	Nota corretor A	Nota corretor B	Nota corretor C	
1.	2	4	4	
2.	2	3	5	
3.	3	5	5	
4.	5	5	6	
5.	4	5	5	
6.	6	6	5	
7.	2	4	4	
8.	4	6	6	
9.	2	3	2	
10.	4	6	5	
11.	2	3	3	
12.	4	6	4	
13.	4	5	6	
14.	2	4	5	
15.	3	5	5	
16.	3	5	4	
17.	3	5	4	
18.	2	3	3	
19.	2	4	3	
20.	2	3	3	
21.	3	6	2	
22.	2	3	3	
23.	2	4	3	
24.	3	6	5	
25.	3	4	4	
26.	6	6	5	
27.	2	3	2	
28.	3	4	3	
Valor médio das notas atribuídas pelos corretores A, B e C	3,03	4,5	4,07	Média calculada entre A, B e C 3,86

De modo geral, as escalas de avaliação tentam equacionar o desempenho dos alunos a descrições verbais específicas, conforme UPSHUR e TURNER (1995). Percebe-se neste quadro que as diferenças das notas atribuídas refletem as decisões dos corretores ao tentarem equacionar o desempenho dos alunos nas 28 composições. O corretor A é mais rigoroso na sua avaliação, enquanto o corretor B é mais benevolente, aspecto que não é novidade e que confirma a análise feita na transcrição da discussão entre eles. No entanto, o corretor C não apresenta um nível de exigência definido, em algumas correções ele é mais severo que o corretor A, como na composição de número 6, e em outras, mais benevolente que o corretor B, como na composição de número 13. Segundo Weigle (2002), corretores “novos” ou corretores menos experientes, ao trabalharem com um critério de correção diferente ou novo, tendem a ser mais rigorosos; portanto, acredita-se que tanto a rigorosidade quanto a benevolência do corretor C em algumas composições estejam relacionadas à sua pouca intimidade com o critério de correções, uma vez que o corretor C não participou da discussão sobre as notas, nem do primeiro treinamento entre os corretores A e B.

Considerando-se que o objetivo de uma avaliação é a aferição de valores justos e coerentes para cada instrumento de testagem, o desvio da média entre os corretores é um aspecto negativo e que interfere nos resultados, independente da importância do teste. Contudo, em avaliações corriqueiras de um curso de línguas, por exemplo, o desvio da média por um ou mais avaliadores não assume caráter definitivo no histórico de um aluno uma vez que não interferirá em seu currículo escolar – embora isso não sirva de justificativa para correções imprecisas, além de poder constituir-se em fator de desmotivação do aluno. Já, em avaliações classificatórias como o vestibular, o desvio de 0,1 ponto da média na correção de composições pode significar a reprovação de um candidato mais qualificado que outro, em um concurso que muda o futuro da vida estudantil de um aluno. Considerando-se a avaliação da língua estrangeira, foco desta pesquisa, em um teste como o TOEFL a atribuição de notas a uma composição pode resultar em um candidato ser ou não aceito para estudar em uma universidade norte-americana, novamente interferindo no seu futuro escolar ou profissional.

3.3 EFEITO DA DIFICULDADE DO TEMA ESCOLHIDO

Durante a análise dos resultados até este momento da dissertação foram consideradas apenas as 28 composições de número 1, corrigidas pelos corretores A, B e C. Nesta parte da análise foram acrescentados os resultados das 11 composições de número 2, avaliadas pelos corretores A, B e C. A intenção ao acrescentar estes outros resultados foi a de proporcionar mais uma composição para alguns dos alunos, para que se torne possível analisar o aspecto da dificuldade do item e seu impacto nos resultados, a partir de correções efetuadas sobre composições de dois temas diferentes.

O Quadro 10 a seguir demonstra a média dos resultados das correções por corretor, e as diferenças nos resultados atribuídos.

QUADRO 10 – DEMONSTRATIVO DA MÉDIA DAS NOTAS ATRIBUÍDAS ÀS COMPOSIÇÕES 1 E 2 PELOS CORRETORES A, B E C

Corretor	Médias das notas dadas à composição 1	Médias das notas dadas à composição 2	Diferença de pontuação entre a composição 1 e 2	Média entre as composições 1 e 2
A	3.27	4.09	0.82	3.68
B	4.9	4.09	0.81	4.49
C	4.18	3.9	0.28	4.04

O objetivo de comparar os resultados auferidos às composições 1 e 2 é analisar a consistência entre os corretores ao avaliarem composições de temas diferentes, comparando as correções de um mesmo corretor e entre os três.

Percebe-se entre os resultados das composições 1 e 2 que o corretor A foi consideravelmente mais benevolente na avaliação da composição número 2, apresentando uma diferença de pontuação de 0.82. O corretor A atribuiu uma nota média entre as composições 1 e 2 de 3.68 pontos, mantendo ainda assim sua característica de rigorosidade.

Nas avaliações do corretor B ocorre o contrário do corretor A, a pontuação atribuída por ele foi mais alta na composição 1 do que na composição 2, demonstrando inconsistência com a avaliação do corretor A e também entre as suas próprias

correções. A diferença da pontuação entre as composições 1 e 2 do corretor B é de 0.81 pontos, fator que demonstra que tanto o corretor A quanto B corrigiram as composições com expectativas diferentes, demonstrando preferência por um tema e que não aplicaram os critérios de correção da mesma maneira.

O corretor C apresentou em suas correções um valor reduzido da diferença de pontuação entre as composições 1 e 2, divergindo em apenas 0.28 pontos. O resultado desta pontuação demonstra maior consistência ao avaliar as composições. No entanto, por ter uma tendência central, conforme análise anterior, é previsível constatar que as notas do corretor C não seriam muito divergentes entre si, pois os resultados tendem a centralizar-se no centro da escala.

No Quadro 11 estão demonstradas as médias das notas atribuídas para as composições 1 e 2 para cada aluno, com o objetivo de possibilitar a análise da habilidade deles nas duas composições.

QUADRO 11 – DEMONSTRATIVO DA MÉDIA ENTRE AS NOTAS ATRIBUÍDAS PELOS CORRETORES A, B E C PARA CADA ALUNO NAS COMPOSIÇÕES 1 E 2.

Número do aluno	Média da composição 1 entre os resultados dos corretores A, B e C	Média da composição 2 entre os resultados dos corretores A, B e C	Média geral do aluno nas composições 1 e 2
2	3.3	3	3.15
5	4.6	5	4.8
8	5.3	5.3	5.3
12	4.6	2.3	3.45
15	4.3	4.3	4.3
21	3.6	3.3	3.45
22	2.6	4	3.3
23	3	2.6	2.8
24	4.6	5.3	4.95
26	5.6	5.3	5.45
28	3.3	3.6	3.45

A escolha de temas em uma avaliação é um aspecto importante no estudo da confiabilidade de um teste, considerando-se que o aluno pode ter a impressão de estar escolhendo o melhor tema e ter sua nota reduzida, pois o aluno nem sempre compreende a sua habilidade a ponto de escolher um tipo de composição em detrimento a outro. Conforme demonstra o Quadro 11, os corretores também têm suas expectativas em relação aos temas e podem auferir notas divergentes, prejudicando assim os alunos.

Outro aspecto importante sobre a escolha de temas em uma avaliação de escrita é que, provavelmente, segundo resultados apresentados no quadro 11, os alunos com maior nível de dificuldade podem ser os mais prejudicados.

Os alunos de número 12, 22 e 24 tiveram diferenças de 2.3, 1.4 e 0.7 em suas médias, respectivamente, e certamente, tendo apenas um tema avaliado, conseguiriam resultados diferentes.

Os dados apresentados não representam uma diferença significativa sobre o grau de dificuldade das composições 1 e 2. Apesar de a composição 1 ser de apenas um parágrafo, a organização mais detalhada, porém concisa, que foi proposta, geralmente apresenta desafios para os alunos em estudo, que costumam escrever muito e com menos atenção ao conteúdo. No entanto, a composição de número 2 requer uma estrutura mais extensa, com desafios maiores no que diz respeito ao número de palavras, às estruturas gramaticais e às idéias em relação à composição 1. Contudo, ao se avaliar composições de temas diferentes, a habilidade de cada um dos alunos passa a ser analisada a partir de dois textos, possibilitando uma avaliação mais aprofundada. Segundo Jones e Shaw (2003), estabelecer tarefas de tipos diferentes, com diferentes graus de desafio, dificulta a comparação do desempenho dos alunos, porém certamente fornece dados mais ricos para a avaliação.

Com base nos resultados apresentados nos quadros 10 e 11, percebe-se que a escolha do tema é uma variável que influencia na habilidade demonstrada dos alunos e na avaliação dos corretores, e que, portanto, precisa ser considerada ao avaliar a confiabilidade dos resultados de um teste. Em exames de alto risco, como testes

classificatórios para concursos, vestibulares, testes de proficiência, nos quais muitas composições são avaliadas e cujos resultados interferem na carreira estudantil ou profissional dos candidatos, oferecer a escolha do tema não configura a opção mais adequada, em minha análise.

3.4 FACETAS

A partir dos dados relatados acima, é possível fazer uma análise das facetas que interferem nos resultados das avaliações. As diferenças relacionadas aos corretores e suas características na correção, aos níveis de dificuldade das composições 1 e 2 e as habilidades dos alunos podem ser verificadas mediante a análise multifacetada de Rasch. Linacre e Wright (1992) elaboraram um software denominado FACETS que implementa a abordagem multifacetada. O modelo de Rasch descreve a probabilidade de sucesso em uma tarefa como uma função da diferença entre a habilidade de um aluno e a dificuldade de uma tarefa; por exemplo, se um aluno tem habilidade alta e a atividade for fácil, existe uma probabilidade alta de sucesso. O modelo de Rasch permite também decompor a dificuldade da tarefa em um número de facetas, por exemplo, neste estudo seria possível analisar pelo modelo a dificuldade das composições 1 e 2, a rigorosidade do corretor e a habilidade dos alunos. A nota seria o cálculo entre estas facetas:

Nota = habilidade do aluno – dificuldade das composições – rigorosidade do corretor.

As diferentes facetas interferem na probabilidade de o aluno conseguir uma nota mais alta. O quadro 4 da página 51 deste estudo demonstra os resultados de uma análise com o emprego do *software* FACETS. A primeira coluna representa a escala de medida em logaritmos, enquanto a segunda coluna representa a habilidade do aluno, a terceira a rigorosidade do corretor e a quarta coluna a dificuldade do item. Os elementos das colunas 2, 3 e 4 são alocados nas medidas de escala da primeira coluna dependendo dos resultados da análise, ou seja, os alunos mais habilidosos ficam

localizados no topo da coluna e os menos habilidosos mais abaixo. As outras facetas relacionadas aos corretores e aos itens seguem a mesma linha, sendo que os corretores mais rigorosos ficam no topo da coluna, assim como o item mais difícil.

A análise feita com o software FACETS resulta bastante informativa, pois os problemas levantados nela revelam aspectos importantes da avaliação. Conforme exemplificado no capítulo 2, a partir da transformação das notas brutas em escores baseados em uma escala e dos resultados da análise do software, que geram um quadro de dados semelhante ao quadro 4, torna-se possível efetuar outras análises, com maior detalhamento de cada faceta individualmente.

A análise multifacetada por meio do FACETS é efetuada em exames de alto risco, como os exames da Universidade de Cambridge e o TOEFL (entre outros exames), que avaliam um número bastante grande de candidatos, cujo impacto dos testes reflete de maneira importante em suas vidas. Uma avaliação como a das composições deste estudo, apesar de sua importância, não é passível de tal análise em função do reduzido número de sujeitos, corretores e itens envolvidos no processo, da falta de acesso ao software e do pouco tempo no contexto da avaliação para aplicar o modelo, impossibilitaram a sua ocorrência. No entanto, a intenção dessa pesquisa é apresentar a aplicação de um software que faça a análise das facetas, como uma das possibilidades existentes na busca pela diminuição da subjetividade na avaliação da escrita.

CONCLUSÃO

No decorrer desta pesquisa tornaram-se evidentes as diferenças entre a *avaliação indireta*, por meio de testes diretos, como, por exemplo, os testes para medir estruturas gramaticais com questões de múltipla escolha, e a avaliação de desempenho, em que são utilizados instrumentos semelhantes às situações reais para que se obtenham amostras mais verossímeis da atuação dos avaliandos em contextos reais.

Por envolverem demonstrações comportamentais dos avaliandos, as avaliações de desempenho são passíveis de várias alterações nos resultados, pois requerem um julgamento mais complexo por parte dos avaliadores, que, mesmo com bastante treinamento, acabam trazendo seu conhecimento de mundo e as suas expectativas sobre o contexto de avaliação no momento de auferir as notas.

Quanto às notas, elas podem sofrer também alterações nos resultados relacionadas à qualidade do instrumento de testagem desenvolvido para demonstrar o desempenho dos alunos, assim como alterações provenientes do comportamento dos avaliandos no momento do teste ou relacionadas as suas competências.

Para Pollitt (1991), as abordagens da avaliação indireta e de desempenho são caracterizadas de forma distintas como contagem e julgamento. O autor explica a contagem com metáforas esportivas como o salto em alturas, levantamento de peso ou a corrida. Jones e Shaw (2003) exemplificam o evento do julgamento com a metáfora esportiva de patinar no gelo, caracterizada pelos autores como um processo basicamente subjetivo, e no contexto da avaliação da língua estrangeira, com a avaliação de um teste de desempenho. Spolsky (1995) faz uso das mesmas analogias ao perguntar se a proficiência lingüística poderia ser medida por uma dimensão concreta, como o número de gols em um jogo ou o tempo de uma corrida, ou se deveria ser julgada mediante critérios subjetivos, como no caso do desempenho de um mergulhador, ginasta ou patinador. As colocações desses autores refletem a

preocupação com a avaliação de desempenho e com o tipo de tratamento que devem receber durante a avaliação.

Neste estudo foram analisadas 39 composições, as quais foram avaliadas por corretores diferentes com o propósito de evidenciar as diferenças entre os resultados, explicar dentro do possível as características e razões das variâncias, assim como apresentar opções que buscassem reduzir as divergências nas notas. Por tratar-se de uma tarefa complexa, tal qual a produção de texto, cujo objetivo de estudo é a produção escrita individual é única, analisada e avaliada por um corretor humano, que tem experiência de mundo e expectativas individuais, as divergências nos resultados apresentaram-se marcantes entre os corretores e houveram notas bem diversificadas para um mesmo candidato.

Essa diversidade nas notas pode ser atribuída a vários fatores, entre eles a utilização de critérios de correções que deve ser cuidadosamente planejada, pois os critérios são manipulados por alunos, professores e principalmente corretores, que são pessoas diferentes e que, portanto, podem divergir na sua compreensão dos desempenhos descritos em cada grau da escala numérica, assim como divergir na sua avaliação pelo fato de agregarem aos critérios suas expectativas pessoais em relação aos textos.

As diferenças de interpretação dos critérios de avaliação, sobre o entendimento da qualidade do conteúdo de um texto, ou até mesmo sobre os objetivos de um instrumento de avaliação, não são passíveis de eliminação; entretanto, o esforço deste estudo está em minimizar a subjetividade na avaliação, uma vez que os resultados obtidos demonstraram medidas que influenciam nas notas, embora os dados da pesquisa não permitam generalizações.

De modo geral, os resultados evidenciam a necessidade de treinar os corretores com frequência em relação ao seu entendimento dos critérios de correção adotados, de promover discussões entre eles, efetuar correções por dois ou até três corretores, utilizando desta forma medidas de correção que propiciem aos alunos uma minimização dos efeitos das facetas de divergências nos resultados das suas notas.

Existe também a opção de aplicar programas de estatística, como o FACETS, nas notas brutas com o objetivo de levantar dados sobre todas as facetas que possivelmente levam a uma diferenciação nas notas dos alunos, impossibilitando a inferência sobre as suas prováveis habilidades. Quaisquer que sejam, as medidas tomadas no sentido de reduzir a subjetividade na correção da escrita serão fundamentais, pois nesta pesquisa torna-se evidente a necessidade de ajustes no processo de avaliação da produção de textos em língua estrangeira para possibilitar uma demonstração segura da habilidade real dos alunos.

A dissertação apresentada evidenciou, ainda que não de forma exaustiva, algumas opções para minimizar os efeitos da subjetividade na avaliação; além disso, procurou demonstrar a necessidade de pesquisar sobre o assunto, refletir sobre os efeitos da avaliação e promover por meio dos resultados das correções analisadas uma amostragem das medidas que podem ser tomadas, como forma de contribuição a este campo de conhecimento.

REFERÊNCIA BIBLIOGRÁFICA

- BACHMAN, L. F. **Fundamental considerations in language teaching**. Oxford : Oxford University Press, 1990.
- BACHMAN, L. F. What does language testing have to offer? **TESOL Quarterly**, 25:671-704, 1991.
- BACHMAN, L. F.; PALMER, A. S. **Language testing in practice**. New York : Oxford University Press, 1996.
- BAILEY, K. M. **Learning about Language assessment – dilemmas, decisions, and directions**. USA : Monterey Institute of International Studies, 1988.
- BRINDLEY, G. Describing language development? Rating scales and SLA. In : BACHMAN, L. F.; COHEN, A. D. (eds). **Interfaces between second language acquisition and language testing research**. Cambridge : Cambridge University Press, 1998.
- BROWN, J. D. Do English faculties rate writing samples differently? **TESOL Quarterly**, 25: 587-603, 1991.
- BROWN, H. D. **Teaching by principles**. New Jersey : Prentice Hall Regents, 1994.
- BROWN, H. D. **Principles of language learning and teaching: An interactive approach to language pedagogy**. 2 ed. White Plains NY : Longman, 2001.
- BROWN, H. D. **Language assessment – principles and classroom practices**. White Plains NY: Longman, 2004.
- CHARNEY, D. The validity of using holistic scoring to evaluate writing: a critical overview. **Research in the Teaching of English**, 18: 65-81, 1984.
- CLARK, J. L. D. Language testing: past and current status – directions for the future. **Modern Language Journal**, 67: 431-443, 1983.
- CRONBACH, L. J. Test validation. In : THORNDIKE, R. L. (ed). **Educational measurement**. 2. ed. Washington DC : American Council on Education, 1971.
- DAVIDSON, F.; HUDSON, T.; LYNCH, B. Language testing: operationalization in classroom measurement and L2 research. In : CELCE-MURCIA, M.(ed.). **Beyond basics: issues and research in TESOL**. Rowley M.A.: Newbury House, 1985.
- FCE Handbook**. University of Cambridge. Local Examinations Syndicate, UCLES, 2001.
- FITZPATRICK, R.; MORRISON, E.J.. Performance and product evaluation. In: THORNDIKE, R.L. (ed). **Educational Measurement**. 2ed.. Washington DC, American Council on Education, 1971. e In : FINCH, F.L. (ed). **Educational Performance Assessment**. Chicago: The Riverside Publishing Company, 1991.
- GREEN, T. Making the grade: score gains on the IELTS writing test. **Research and validation group-research notes**. University of Cambridge Examinations ALTE, May, 2004.

- GRONLUND, N. E. **Assessment of student achievement**. 6. ed. Boston : Allyn and Bacon, 1998.
- HAERTEL, E. Performance measurement. In : ALKIN, M. C. (ed.) **Encyclopedia of educational research**, 6 ed., 1992. p. 984-9.
- HAMP-LYONS, L. Newbury house TOEFL preparation – **preparing for the test of written English**. New York: Newbury house publishers, 1989.
- HAMP-LYONS, L. Second language writing: assessment issues. In : KROLL, B.(ed.). **Second language writing: research insights for the classroom**. Cambridge : Cambridge University Press, 1990. p. 69-87.
- HENNING, G. The influence of test and sample dimensionality on latent trait person ability and item difficulty calibrations. **Language Testing** 5(1): 83-99, 1988.
- HERMAN, J. L.; ASCHBACHER, P.R.; WINTERS, L. **A practical guide to alternative assessment**. Alexandria Va. : Association for Supervision and Curriculum Development, 1992.
- HILLOCKS, G., Jr. Synthesis of research on teaching writing. **Educational Leadership** 44(8): 71-82, 1987.
- HUGHES, Arthur. **Testing for language teachers**. Cambridge : Cambridge University Press, 1989.
- HUGHES, D. C.; KEELING, B. ; TUCK, B. F. Are untidy essays marked down by graders with neat handwriting? **New Zealand Journal of Educational Studies**, 18: 184-6, 1983.
- JACOBS, H. L.; ZINGRAF, S. A.; WORMUTH, D. R.; et al. **Testing ESL composition: A practical approach**. Rowley M.A : Newbury House, 1981.
- JONES, R. I. Testing: a vital connection. In : PHILLIPS, J. (ed.). The language connection: from the classroom to the world. **The ACTFL Review of Foreign Language Education**, vol. 9: 237-65. National Textbook Co, Skokie II, 1977.
- JONES, R. I. Performance testing of second language proficiency. In : BRIÈRE, E.J.; HINOFOTIS, F.B. (eds). **Concepts in language testing: some recent studies**, TESOL, Washington DC, 1979. p.50-7.
- JONES, R. I. Second language performance testing: an overview. In : HAUPTMAN, P.C.; LEBLANC, R.; WESCHÉ, M.B. (eds). **Second language performance testing**. Ottawa: University of Ottawa Press, 1985. p.15-24.
- JONES, N.; SHAW, S. D. Task difficulty in the assessment of writing: comparing performance across three levels of CELS. **Cambridge ESL Research and Validation Group**, Research Notes 11, University of Cambridge – ESOL Examinations, February, 2003.
- KENYON, D. Introductory remarks at symposium on development and use of rating scales in language testing, **14th Language Testing Research Colloquium**, Vancouver, February 27 – March 1, 1992.
- KERR, Philip. **American inside out - workbook – upper intermediate**. MacMillan, Oxford, 2003. p.58.
- LADO, R. **Language testing: the construction and use of foreign language tests**. London: Green and Co. Ltd., 1961.

- LINACRE, J. M. **Many-faceted Rasch measurement**. Chicago IL. : MESA Press, 1989.
- LINACRE, J. M.; WRIGHT, B. **Facets: Rasch measurement computer program, version 2.6**. Chicago: MESA Press, 1992.
- LUNZ, M. E.; STAHL, J. Judge consistency and severity across grading periods. **Evaluation and the Health Professions** 13:425-44, 1990.
- McINTYRE, P. N. **The importance and effectiveness of moderation training on the reliability of teacher assessments of ESL writing samples**. MA thesis, University of Melbourne, 1993.
- McNAMARA, T. **Measuring second language performance**. England : Longman, 1996.
- McNAMARA, T. **Language testing**. Oxford : Oxford University Press, 2000.
- McNAMARA, T. F. ; LYNCH, B. K. A generalizability theory study of ratings and test design in the writing and speaking modules of the ACCESS: test. In : WIGGLESWORTH, G. ; BRINDLEY, G. (eds) .**Access: Issues in English Language Test Design and Delivery**. National Centre for English Language Teaching and Research, Sydney : Macquarie University, 1993.
- MESSICK, S. The interplay of evidence and consequences in the validation of performance assessments. **Educational Researcher** 23 (2): 13-23, 1994.
- MOUSAVI, S. A. **An encyclopedic dictionary of language testing**. 3. ed.. Taiwan : Tung Hua Book Company, 2002.
- NICHOLLS, S.M.. A avaliação de rendimento e a elaboração de testes. In : **Aspectos pedagógicos e metodológicos do ensino de inglês**. Cap.13. Maceió AL. : EDUFAL, 2001. p.111-120.
- O'MALLEY, J. M.; PIERCE, L. V. **Authentic assessment for English language learners – practical approaches for teachers**. USA ; Addison-Wesley Publishing Company, 1996.
- POLLITT, A. Giving students a sporting chance: assessment by counting and by judging. In : ALDERSON, C.; NORTH, B. (eds). **Language testing in the 1990s**. London : British Council/MacMillan, 1991.
- RYANS, D. G.; FREDERIKSEN, N. Performance tests of educational achievement. In : LINDQUIST, E. F. (ed). **Educational measurement**. Washington D.C. : American Council on Education, 1951. p. 455-91.
- SHOHAMY, E.; GORDON C. M.; KRAEMER, R. The effect of raters' background and training on the reliability of direct writing tests. **The Modern Language Journal**, Spring, 1992.
- SKEHAN, P. **Individual differences in second language learning**. London: Edward Arnold, 1989.
- SLATER, S. J. Introduction to performance testing. In : Spierer, J. E. (ed). **Performance testing: issues facing vocational education**. Columbus OH. : National Center for Research in Vocational Education, 1980.
- SMITH, F. **Writing and the writer**. New Jersey : Lawrence Erlbaum Associates, 1994.
- SPOLSKY, B. **Measured words**. Oxford : Oxford University Press, 1995.

STIGGINS, R. J. Design and development of performance assessments. **Education Measurement: Issues and Practice** 6 (3): 33-42, 1987.

TOEFL Score user guide. **Educational testing service**, Princeton, NJ, 2001.

UR, P. **A course in language teaching**. Cambridge : Cambridge University Press, 1996.

WEIGLE, S. C. **Assessing writing**. Cambridge : Cambridge University Press, 2002.

WEIGLE, S. C. Using FACETS to model rater training effects. **16th Annual Language Testing Research Colloquium**, Washington D.C., March, 1994.

WEIR, C. J. **Communicative language testing**. London : Prentice Hall International, 1990.

WESCHE, M.B. Communicative testing in a second language. **Modern Language Journal** 67 (1): 41-55, 1983.

WESCHE, M.B. Second language performance testing: the Ontario test of ESL as an example. **Language Testing** 4 (1): 28-47, 1987.

WESCHE, M. B., Introduction . In : HAUPTMAN, P.C., LEBLANC, R.; WESCHE, M.B. (eds). **Second language performance testing**. Ottawa : University of Ottawa Press, 1985.

WESCHE, M.B. Performance testing for work-related second language assessment. In : SHOHAMY, F.; WALTON, R. (eds). **Language assessment for feedback: testing and other strategies**. Washington D.C. : National Foreign Language Center Publications, 1992. p.103-22.

APÊNDICES

APÊNDICE I - CRITÉRIO DE CORREÇÃO ADAPTADO PARA ESTA PESQUISA

Critério holístico de avaliação de escrita:

RESULTADO	CRITÉRIOS
6 Proficiente	<ul style="list-style-type: none"> • Redige parágrafos únicos ou múltiplos com introdução clara, idéias completamente desenvolvidas e conclusão; • Emprega tempos verbais apropriados e estruturas gramaticais e sintáticas variadas; emprega frases complexas com eficácia e utiliza vocabulário de transição adequadamente; • Vocabulário rico e preciso; • Apresenta erros de mecânica ocasionais (ortografia e pontuação), que não interferem no significado.
5 Fluente	<ul style="list-style-type: none"> • Redige parágrafos únicos ou múltiplos com idéia central e de apoio; apresenta idéias de forma lógica, apesar de algumas partes não estarem completamente desenvolvidas; • Emprega tempos verbais apropriadamente e estruturas gramaticais e sintáticas variadas; erros de estruturação de frases não interferem no significado; utiliza vocabulário de transição; • Vocabulário variado de acordo com o propósito do texto; • Apresenta poucos erros de mecânica que não interferem no significado.
4 Em Expansão	<ul style="list-style-type: none"> • Organiza idéias em uma ordem lógica e seqüencial com algumas idéias de apoio; inicia a redação de um parágrafo; • Experimenta com uma variedade de tempos verbais, mas não de forma consistente; apresenta erros de concordância verbal; utiliza frases compostas e complexas; utiliza vocabulário de transição de forma limitada; • Vocabulário adequado ao propósito do texto, no entanto às vezes estranho; • Emprega ortografia e pontuação convencionais; erros às vezes interferem no significado.
3 Em Desenvolvimento	<ul style="list-style-type: none"> • Redige frases sobre uma idéia; apresenta alguma seqüência, mas pode faltar coesão; • Redige frases simples no tempo presente; apresenta dificuldades com concordância verbal; <i>run-on</i> sentences são comuns; inicia o uso de frases compostas; • Vocabulário de alta freqüência; pode apresentar dificuldades com a ordem das palavras; omite terminações ou palavras; • Utiliza algumas regras de pontuação e ortografia; erros freqüentemente interferem no significado.
2 Iniciante	<ul style="list-style-type: none"> • Inicia a transmissão de idéias através da escrita; • Redige predominantemente expressões ou frases simples/padronizadas; • Vocabulário limitado e repetitivo; • Apresenta erros de ortografia.
1 Elementar	<ul style="list-style-type: none"> • Não há evidência de organização ou desenvolvimento de idéias; • Apresenta palavras simples, gravuras ou frases padronizadas; • Cópia de um modelo; • Pouca noção de ortografia e pontuação.

Adaptado do livro **Authentic Assessment for English Language Learners** (p. 22), desenvolvido por um grupo de professores de inglês como segunda língua (ESL) em escolas públicas do Fairfax Country, na Virgínia, EUA.

APÊNDICE II – QUESTIONÁRIO A SER RESPONDIDO PELOS CORRETORES

A. Introdução

1. Corrigir as composições em ordem numérica crescente.
2. Caso haja pausa entre as correções, anotar o número da última composição corrigida e os horários de término e reinício das correções.

B. Informação sobre o corretor

1. Idade: _____
2. Descrição breve do seu histórico profissional e experiência com a correção de composições.

3. Você se considera um corretor exigente ou permissivo? Justifique.

C. Aspectos sobre critérios de correção

1. Que tipos de critério de correção você já utilizou em sua experiência profissional?

2. Você compreendeu claramente o critério de correção adotado neste estudo?

3. Na sua opinião você teria utilizado um critério diferente? Justifique?

4. Quais são os aspectos positivos e negativos de utilizar um critério holístico?

5. Quais são as limitações deste critério?

D. Aspectos e impressões sobre o ato da correção

1. A que horas você corrigiu as composições?

2. Todas as composições foram corrigidas em um mesmo momento? Caso contrário, explique os momentos de correção.

3. Como você se sentiu:

a) no início da correção?

b) na metade da correção?

c) no final da correção?

4. O que você já havia feito durante o dia até o início das correções?

5. Em que dia da semana você efetuou as correções?

6. Em que local você efetuou as correções?

7. Não conhecer os alunos ou os seus nomes, fez diferença para a correção?

8. Qual a sua reação ao atribuir notas de 1 a 6, e não de 0 a 10?

9. Quais são os aspectos positivos e negativos de utilizar uma escala de correção numérica menor (de 1 a 6)?

ANEXOS

ANEXO 1 – CRITÉRIOS DE CORREÇÃO – FCE

Band 5	<p>Full realisation of the task set.</p> <ul style="list-style-type: none"> • All content points included with appropriate expansion. • Wide range of structure and vocabulary within the task set. • Minimal errors, perhaps due to ambition; well-developed control of language. • Ideas effectively organised, with a variety of linking devices. • Register and format consistently appropriate to purpose and audience. <p>Fully achieves the desired effect on the target reader.</p>
Band 4	<p>Good realisation of the task set.</p> <ul style="list-style-type: none"> • All major content points included; possibly one or two minor omissions. • Good range of structure and vocabulary within the task set. • Generally accurate, errors occur mainly when attempting more complex language. • Ideas clearly organised, with suitable linking devices. • Register and format on the whole appropriate to purpose and audience. <p>Achieves the desired effect on the target reader.</p>
Band 3	<p>Reasonable achievement of the task set.</p> <ul style="list-style-type: none"> • All major content points included; some minor omissions. • Adequate range of structure and vocabulary, which fulfils the requirements of the task. • A number of errors may be present, but they do not impede communication. • Ideas adequately organised, with simple linking devices. • Reasonable, if not always successful attempt at register and format appropriate to purpose and audience. <p>Achieves, on the whole, the desired effect on the target reader.</p>
Band 2	<p>Task set attempted but not adequately achieved.</p> <ul style="list-style-type: none"> • Some major content points inadequately covered or omitted, and/or some irrelevant material. • Limited range of structure and vocabulary. • A number of errors, which distract the reader and may obscure communication at times. • Ideas inadequately organised; linking devices rarely used. • Unsuccessful/inconsistent attempts at appropriate register and format. <p>Message not clearly communicated to the target reader.</p>
Band 1	<p>Poor attempt at the task set.</p> <ul style="list-style-type: none"> • Notable content omissions and/or considerable irrelevance, possibly due to misinterpretation of task set. • Narrow range of vocabulary and structure. • Frequent errors which obscure communication; little evidence of language control. • Lack of organisation, or linking devices. • Little or no awareness of appropriate register and format. <p>Very negative effect on the target reader.</p>
Band 0	<p>Achieves nothing; too little language for assessment (fewer than 50 words) or totally irrelevant or totally illegible.</p>

ANEXO 2 – CRITÉRIOS DE CORREÇÃO – TOEFL

Table 9.1. Test of Written English Scoring Guide

6	<p>Demonstrates clear competence in writing on both the rhetorical and syntactic levels, though it may have occasional errors.</p> <p><i>A paper in this category</i></p> <ul style="list-style-type: none"> • effectively addresses the writing task. • is well organized and well developed. • uses clearly appropriate details to support a thesis or illustrate ideas. • displays consistent facility in the use of language. • demonstrates syntactic variety and appropriate word choice.
5	<p>Demonstrates competence in writing on both the rhetorical and syntactic levels, though it will probably have occasional errors.</p> <p><i>A paper in this category</i></p> <ul style="list-style-type: none"> • may address some parts of the task more effectively than others. • is generally well organized and developed. • uses details to support a thesis or illustrate an idea. • displays facility in the use of language. • demonstrates some syntactic variety and range of vocabulary.
4	<p>Demonstrates minimal competence in writing on both the rhetorical and syntactic levels.</p> <p><i>A paper in this category</i></p> <ul style="list-style-type: none"> • addresses the writing topic adequately but may slight parts of the task. • is adequately organized and developed. • uses some details to support a thesis or illustrate an idea. • demonstrates adequate but possibly inconsistent facility with syntax and usage. • may contain some errors that occasionally obscure meaning.
3	<p>Demonstrates some developing competence in writing, but it remains flawed on either the rhetorical or syntactic level, or both.</p> <p><i>A paper in this category may reveal one or more of the following weaknesses:</i></p> <ul style="list-style-type: none"> • inadequate organization or development • inappropriate or insufficient details to support or illustrate generalizations • a noticeably inappropriate choice of words or word forms • an accumulation of errors in sentence structure and/or usage.
2	<p>Suggests incompetence in writing.</p> <p><i>A paper in this category is seriously flawed by one or more of the following weaknesses:</i></p> <ul style="list-style-type: none"> • serious disorganization or underdevelopment • little or no detail, or irrelevant specifics • serious and frequent errors in sentence structure or usage • serious problems with focus.
1	<p>Demonstrates incompetence in writing.</p> <p><i>A paper in this category</i></p> <ul style="list-style-type: none"> • may be incoherent. • may be undeveloped. • may contain severe and persistent writing errors.
0	<p>A paper is rated 0 if it contains no response, merely copies the topic, is off-topic, is written in a foreign language, or consists only of keystroke characters.</p>

TOEFL Score User Guide. Educational Testing Service, Princeton, NJ, 2001.

ANEXO 3 – MODELO DE CRITÉRIO DE CORREÇÃO HOLÍSTICO

Rating	Criteria
6 Proficient	<ul style="list-style-type: none"> • Writes single or multiple paragraphs with clear introduction, fully developed ideas, and a conclusion • Uses appropriate verb tense and a variety of grammatical and syntactical structures; uses complex sentences effectively; uses smooth transitions • Uses varied, precise vocabulary • Has occasional errors in mechanics (spelling, punctuation, and capitalization) which do not detract from meaning
5 Fluent	<ul style="list-style-type: none"> • Writes single or multiple paragraphs with main idea and supporting detail; presents ideas logically, though some parts may not be fully developed • Uses appropriate verb tense and a variety of grammatical and syntactical structures; errors in sentence structure do not detract from meaning; uses transitions • Uses varied vocabulary appropriate for the purpose • Has few errors in mechanics which do not detract from meaning
4 Expanding	<ul style="list-style-type: none"> • Organizes ideas in logical or sequential order with some supporting detail; begins to write a paragraph • Experiments with a variety of verb tenses, but does not use them consistently; subject/verb agreement errors; uses some compound and complex sentences; limited use of transitions • Vocabulary is appropriate to purpose but sometimes awkward • Uses punctuation, capitalization, and mostly conventional spelling; errors sometimes interfere with meaning
3 Developing	<ul style="list-style-type: none"> • Writes sentences around an idea; some sequencing present, but may lack cohesion • Writes in present tense and simple sentences; has difficulty with subject/verb agreement; run-on sentences are common; begins to use compound sentences • Uses high frequency words; may have difficulty with word order; omits endings or words • Uses some capitalization, punctuation, and transitional spelling; errors often interfere with meaning
2 Beginning	<ul style="list-style-type: none"> • Begins to convey meaning through writing • Writes predominately phrases and patterned or simple sentences • Uses limited or repetitious vocabulary • Uses temporary (phonetic) spelling
1 Emerging	<ul style="list-style-type: none"> • No evidence of idea development or organization • Uses single words, pictures, and patterned phrases. • Copies from a model • Little awareness of spelling, capitalization, or punctuation

Adapted from a rubric drafted by the ESL Teachers Portfolio Assessment Group, Fairfax County Public Schools, Virginia.

ANEXO 4 – MODELO DE CRITÉRIO DE CORREÇÃO HOLÍSTICO

COMPOSIÇÃO 1

My Roommate Rita

Rita, my roommate, is very inconsiderate. She leaves the kitchen in a mess. Yesterday, Rita made a sandwich and left the lettuce and mayonnaise on the breadboard.. Also, she frequently leaves one or two dirty pots or pans in the sink. Sometimes she eats my food without asking. Last week, she ate the red apple that I was going to take for lunch. This morning she used up my milk on her breakfast cereal. Besides, Rita uses the bathroom for an hour or more every morning. She showers, and washes and blowdries her hair. She even makes up her face in there. Of course, the bathroom is always in a mess. I have found towels on the floor and toothpaste on the washbasin. Finally, she seldom takes my phone messages. Last night, I asked her if anyone had called while I was out. But she couldn't remember if Jim or Joe had called. When she does take a message, she sometimes forgets to write down an important phone number. **In conclusion, Rita is a very thoughtless person.** If she doesn't change some of her ways soon, I am going to look for another roommate.

Understanding the Paragraph

Let's review the paragraph about Rita to make sure you understand its structure. The first sentence is the topic sentence. It introduces the topic and the controlling idea for the whole paragraph.

- **Rita, my roommate, is very inconsiderate.**

All of the main supporting sentences that follow the topic sentence must support it. Notice that each of the main supporting sentences has a main idea. So, for example, the first main supporting sentence points out one of the writer's main complaints.

- **She leaves the kitchen in a mess.**

The supporting details prove this statement by explaining that food is left out and that the kitchen sink is left dirty. So, this part of the paragraph is now complete.

The second main supporting sentence also supports the topic sentence, by pointing out another of the writer's main complaints.

- **Sometimes she eats my food without asking.**

Again, the supporting details must explain this main point. Both the apple and the milk have disappeared. These details complete this section of the paragraph.

Now, in the following blanks, add the supporting details for the rest of the main supporting sentences in the paragraph about Rita.

Third supporting sentence:

Rita uses the bathroom for an hour or more every morning.

1. _____

2. _____

Fourth supporting sentence:

The bathroom is always in a mess.

1. _____

2. _____

Fifth supporting sentence:

She seldom takes my phone messages.

1. _____

2. _____

3. _____

Write a one-paragraph composition in which you describe (1) a good or bad habit of one of your friends or family members.

COMPOSIÇÃO 2

WRITING

1. You are going to write a description of a place that you have visited on vacation recently. Think of a place you have been to.
2. Look at the list of topics below. Check ✓ each topic that you think it is necessary to include (because it gives essential information). Put a star * next to each topic that you think would be interesting to someone who is thinking of visiting this place.

What is it called? Where is it?

How do you get there?

Are there any interesting facts about this place (e.g., size, population, history)?

Is it famous or popular? Why or why not?

When did you go there?

What time of the year is it best to visit?

What is the weather usually like?

Why did you go there?

What can people do there?

Is there anything to do at night?

Are there interesting places to visit nearby?

What kind of scenery or buildings do you find there?

What are the people like?

Are there any interesting places to eat or drink?

Are there any interesting traditions or festivals?

How did you feel about your visit?

What advice would you give to someone who is going to visit this place?

3. Select the most interesting topics and make an outline. You should write four paragraphs. Decide which topics will go in each paragraph.

Paragraph 1

Paragraph 2

Paragraph 3

Paragraph 4

4. Write a description of the place you visited. Remember that the person who will read your description is thinking about going there too.

ANEXO 5 – COMPOSIÇÕES

1.

A friend of mine, Cris, is always listening music very high. For me it's OK, but the problem is that I like rock music and she likes country music. I respect her way, but I should respect mine too. I listen music in my bedroom and she listens in all the rooms of the house. It makes me very nervous. And we are very different people. I like to go out, and she disapprove this. We discut many times because of that. She thinks that "good girls" don't go out. I like to live with her, but if she doesn't change or tolerate thinks that I like, I'll look for another roommate.

2.

Live with someone else is a very hard work. People think differently and you must know to respect other's ideas, although sometimes it is not so easy. For example, some people just cannot do something and not leaving the place in a mess. On the other hand, there are people who leaves everything perfectly cleaned and organized. Definitely I am not like this, not the first case, neither the second, but I like the things cleaned and on its places, so there are some habits from other people that really makes me lazy, such as leave some place I have just finished cleaning in a mess, or leave the sink dirty with toothpaste. But what I really dislike are people who leaves everything perfect. In conclusion, everybody has likes and dislikes and it is not possible to everybody agrees about everything.

3.

Biscala, my friend, is a bad person. She is always looking at what someone is doing. Last year she started talking about people around us and one day she complained about my way of acting in front of people. Besides, she is always complaining that she hasn't money to buy the things that she likes, so, when she is with me in the mall I start to feel sad when I have money to buy CD, DVD and she hasn't nothing. Finally, she complains that I don't care about her. Last month when we were talking about a lot of things in the computer, she always keep saying that I changed and I am not like before. In conclusion Biscala is a pain in the neck! I really like her, but if she goes on this way, she will keep me away from her, that's what makes me feel sad.

4.

My classmate Eva is extremely annoying. She is always talking about people. Last week, Eva dared to talk bad about my friends, in a way that really offended me. She never takes a hint when she is destroying people, really, behind their backs. I've never met someone like her, it really bothers me the way she acts and how horrible this bad habit is. Even though, I am here writing the most I can about her disgusting habit, it doesn't mean I act the same. I know how people are, but that is no reason for me to insert, everyday, in my classmate's head a report of other people's lives. I have to do something about this situation, otherwise, I will spend the rest of my college years listening to her terrible news, in which is always a critic to someone who she doesn't get along well.

5

Stella, my best friend, is an amazing person. Since we started studying together she shows her wonderful sense of humor. She is always happy and her smile is very beautiful. We have done many assignments together and those experiences were great because she is a very dedicated girl. One of her best qualities is the confidence that she has in herself. She does not know that but with all the problems she passes through she is always strong and willing to start again. In conclusion, there are many qualities and adjectives I could write about her, but they would not fit in a one-paragraph composition, because she is indescribable.

6.

I am glad I don't need to deal with this bad habit anymore, but in the beginning it was tough to manage. When I first started dating my husband, some two year ago, everytime we arranged for him to pick me up, at home or anyplace else, he would be ten to fifteen minutes late, it drove me mad. Due to his constant delays, we would always be late for appointments, leave friends waiting, miss the beginning of film so on and so forth. Everytime he showed up late, he would have to listen to a long lecture on how awful it was for me to always be late for things, specially considering how methodic I am, a person who always early for everything. No to mention all the times he left me waiting at his station and airports, whenever I traveled to meet him. His common excuse was that either he miscalculated the time to drive or where I was, or the traffic was very bad. Finally after a lot of talking on how he need to manage better his time and look more often at his watch, he understood that his bad habit was affecting me and my life, so he decided to be more alert and promised to change. Indeed he did, he would never again be more than five minutes late, which I decided to accept for the sake of a happy marriage.

7.

My father is a very smart man. For start, he's always open for discussion. Everytime I want to do something, even though he doesn't agree, he will listen to me and accept if I have a good argument. He will never take something as an undeniable true. Again, all you need is good arguments to convince him that your idea should be viable. He's also a tireless researcher. As professor he's always working hard to improve and he's never satisfied with his acknowledge. That's why I admire my father so much, because of his open mind and because he is always open to new ideas.

8.

My Super-Mom.

My mother Lisa is the most special person I know. She is already sixty-one years old and she still works. She manages three offices and she does it very well. She is in charge of many important things, such as money and investments, and she deal with that easily. She is also a very reliable person. She has been working at the same place for more than forty years. She works with somebody else's money and people trust her a lot. For this forty years she has never had a problem in her job. Besides being an excellent worker, my mom is also a wonderful mother. In spite of working full time, she has always taken care of our house and of her family. When my brother and I were still in school, she managed to attend all the school meetings, even if she had an important work at work. On the weekends, she always cooks what we like and spends a lot of time with her family. In conclusion, my mother is a wonderful person and an excellent mother. I wish I could be a mother and a person like her and I also wish I could manage my time like she does.

9.

My older sister, Jemifa, is the kind of person who thinks that she does everything right, and like to give orders to everyone. To do that, she use to say "Don't leave your bed room or bathroom a mess", "Do not leave towels and pants on the bathroom's floor"; and some others things like that. But the thing, is that she doesn't do what she say, all the thing that she use to say "Do not do", she does. She annoys not just me, but everyone in the house, it's really sucks get in my room, and find it out in a completely mess, with clothes on the floor. The problem, is that she didn't realize that she does all this things and say that it isn't the truth.

10.

Mr. Trouble

My brother is a very careless person. He likes to hang out with his friends four or five nights a week, but he doesn't tell anybody. My mom tries to call his cell phone, but he doesn't answer. She gets really worried, of course. My father has told him a thousand times not to come home late, but he never listens, and that pisses my father off.

Besides, he's often involved in fights and discussions. Once he had an argument with a guy on a street, and the guy threw his car's fire extinguisher on my brother's car! A few weeks ago, he came home with bruises and refused to explain what had happened. He was forbidden to go out for a whole month.

Although he's a nice boy, my brother gets into trouble very easily. He has destroyed my father's motorcycle. His car was arrested because of the loud music. A friend of his was arrested selling marijuana. In sum, he has not the happiest stories to tell.

That doesn't mean he's going to be like that all his life. He's at University now, he has a new girlfriend and a new group of friends. My mom and dad says that he's getting better everyday.

Well, at least that's what we hope!

11.

Daniel, my friend, is so organized. Everything must be in the right place. If something in a room is gets messy about there, in his room is always everything in perfect order, if you mess up something you need to organize it immediately, because if you don't do that he gets mad with you. Even when we all gonna go out, we need all to be on time, if we said "on 20:00 we will be there", is on 20:00 that we need to go. But with all this organization he is a really nice guy. And we all like him a lot.

12.

my father. Paul, is driving me crazy. He doesn't leave with us anymore - actually, he left when I was 5:00 but he keeps telling me what to do. I had a conversation with him last month about military career. He knows I don't like the idea, but he is still trying to convince me that joining the army - in this case it was about navy - is the best for me, because I won't be without work. I will study for free and have a wonderful life. What he cannot understand is that we already had this conversation when I was 10:00 and I made myself clear about it. Just because he joined the air force, he tries to make me one of them. Also, he frequently asks me about my study works and giving me "tips" about how I must do this and that in university. He thinks I am still a baby and that I need his help or advices. OK, I don't have a nice relation with him. He was never a father to me. Maybe I should free him and tell him this. Maybe I should tell him I don't want his advices and I don't want him causing me anymore troubles.

Writing:

My friend Carol is one of my best friends. She is my age and we do almost everything together. Sometimes, however, she can be very annoying. Last week, for instance, my mother came to visit and decided to cook me dinner. She called me seven times during dinner saying that she needed to talk. I told her I would call her later, but it didn't matter. We are both very hardworking people, but she seems to be complaining all the time that she deserves to earn more for what she does. Since I work all day myself, I never say anything because I think a lot of people feel the same way. We always lend each other things, such as books, clothes, and makeup. It seems that I never do get my things back, even if I return whatever I borrowed from her. Well, I guess everyone has their own problems. Aside from all that, she is as loyal as a friend can be and by the end of the day that's all that really matters.

14

Absent-minded Bruno

My brother Bruno is a very disheveled and funny teenager. He usually forgets easily what I or my mother ask him to do. After lunch my mother asks Bruno to wash the dishes and leave them to work. At night when she arrives the dishes are still on the table and Bruno is sleeping. On Tuesdays and Thursdays I wash clothes and ask Bruno to take the clothes out of the washing machine and put them to get dry after his English class. When I arrive at night the clothes are still in the washing machine and I can hear Bruno playing his guitar. Sometimes he also forgets to go to school in the morning and I have to remember him. But Bruno is also an extremely funny boy. He can make jokes of anything, but the difficult thing Bruno is used to do to make us laugh is imitating. If someone tells him about a person he can imitate in a comic way. His best imitation, elected by the family, is when he imitates Homer Simpson, a cartoon character. Bruno can always make anyone laugh even though we are always asking him ten or twelve times (and almost losing patience) to do something.

15.

Kleyton, my brother-in-law, is such an inconvenient person to live in an organized family. He leaves every place of the house in an incredible mess. Every time he takes a shower he leaves the bathroom all soaked. He also forget to take up his clothes. Some afternoons, he uses to take a nap in the T.V. room, and when he wakes up, the couch is all scattered. Last week he was preparing dinner and he forgot to turn off the microwave oven, so it was turned on the whole night, wasting power and money. In addition, he did not wash the dishes that he used, and the worst problem was that he forgot to cover the pans of the food and put it in the refrigerator, as the result was that all that food got spoiled. There are much more bad habits he has, such as never making up her bed or arranging his bathroom, not even tidying up his clothes. The main problem is that while he is not organized with his own things and does not care with the house organization, he disturbs every member of the family. In conclusion, if Kleyton does not change his habits it would become a more serious problem because it involves not only him, but the whole family.

16.

My wife is a very organized person. She likes to see everything in its place. For example, Tatiane (my love) appreciates to put her things and clothes separated of mine, she uses to wake up and sleep at the same time everyday and, obviously, she has a schedule where she writes down what is necessary to do tomorrow or next week. What my love likes to do also is cleaning the house. After lunch or dinner the stove is always clean, the dishes are washed after every meal and the bathroom has a wonderful smell. In conclusion, she takes care of our home and I think I'll need to do the same, otherwise, Tatiane will be my ex-wife.

17.

My lazy sister

Dayane is my sister, she's very intelligent, but she's so lazy. She doesn't even want to hear what I have to say to her, because she knows I'll complain about her attitudes.

On weekends, my mother always makes a lot of food, because many relatives go to my house. Everybody helps, makes something, but my sister stays in her bed. Dayane prefers to watch TV or listen to the radio. It's against Dayane's way. Sometimes I think the problem is with me, because I like to see everything organized, and papers separated in colors, books separated for size. Money with less value comes first, then comes money with more value in my wallet. So that's the reason why my sister and I fight when we stay at home together.

There's one thing I can't deny, she's very intelligent. She studies hard, she is an example of a student, she makes in all tests she does in first or second place, including the ones for jobs. She's really very industrious. I think she's gonna be a very important person in the future, she intends to make a career about biotechnology. Although she's very lazy, my family and I support her in everything she does, but surely she must start to do something to change this feeling, because she's gonna grow up, marry and will have to do a lot of things, as work for hours and after that watch her baby and other things. By now it would be a great notice if she not changes but improves her mind.

18.

My cousin Marcos

Marcos, my cousin-bro, is completely organized. He always wakes up earlier than me and my real brother and he soon opens the windows of my house, when he sleeps there. There are things that I get amazing, for instance, to wash the breakfast dishes and tidy up his bed. It's a little bit funny the way he wakes me, because he knows I sleep a lot and that's why sometimes with his jokes make me get nervous, in spite of this I accept his jokes. Seldom Marcos leaves a mess, however he organizes it. In actuality I consider him like a brother who has a great personality and in spite of he is twenty-four years, he is like a child and an adult with his responsibilities.

19.

My best friend Ara Paula

Ara Paula is a wonderful girl. So, let's explain why she is wonderful!!

Firstly, because she is a kind and helpfull person. When we worked together at Vaug's Air Company, she used to help, not only me but also the people at the job!! And if I was unhappy, she told me funny stories or invited me to go a party. Secondly, she will be my god mother, because she was my "cupid": has presented her friend to me, so, we have been dating about 2 years and half and we intend to get married! All in all, Ara has a big heart and because of it, I wish the best for her!

20.

My brother

Miguel, my brother is very disorganized. He leaves the bathroom a mess. The toothpaste and his brush teeth are always in the washbasin and after he takes a shower I have found his socks and pants on the floor. His bedroom is dirty. His clothes never are in the right place, and he leaves the wet towel in his bed. Where he passes he turn on the light. Yesterday he went to my bedroom to take a cover and forgets to turn off the lights. Finally he takes my thing without ask. Last week he took my radio and he didn't care of the radio that fell down and now doesn't work anymore. Miguel is a person very difficult to live with and if he doesn't change someday the house will be ruin.

21.

one-paragraph composition

SANDRA, My mother, is a very reasonable person. She always consider the situation I pass through in a properly way. Every time I have doubts in life, she is there to say something useful to me. She really knows how to act, even when she is trying to overcome her own problems, and that is why she is so reasonable. When I have money problems she is always the first one to realize it, and try to find out a solution for it. People say that is because she is a mother, and mothers are usually prepared to face this kind of situation, but I am sure that she is a special one. In conclusion, there is not such a reasonable person as my mother.

22.

I live with three nice girls in an apartment. I really enjoy to live with them. I moved two months ago and it was not difficult to feel at home there. The girls are very respectfull and we have a procedure in the house for the housecleaning. Each week everybody has to clean a part of the apartment. One week the restroom, the other the kitchen, the livingroom and take the garbage out. It really works. Everyone is responsible and when one of us does not do this job, we have to pay eight reais. It is really great to live with people like that. We can learn a lot and we become like a family.

23.

my middle brother, Wallace, is very organized and friendly. He keeps his things in its place. As soon as my father iron his clothes he store them on this closet, as I do. When the kitchen is cleaned and he uses a cup, for example, he washes it keeping everything organizing and clean. If you ask him to help you, he does. When our parents are travelling, he offers help to cook and clean the house. If he is watching some interesting TV program, and I call him to help me, he comes. Usually people do not do this. They usually continue watching the program and help others later. He is very funny, as he makes friend easily. He laughs at everything and any problem. When he notices somebody is upset, he makes this person to forget the trouble telling jokes. In conclusion, my middle brother is easy-going. People can count to him in good and bad situations. Because of this, people admire him.

24.

Cirel, my classmate and one of my best friends, is so admirable. By the first time you meet her you can notice that she has a very strong personality. She knows exactly what she wants and she works hard to achieve her objectives. She is also a very strong person, when life shows its difficult moments, she does not allow them to let her down, and she keeps going being always dedicated. By the way, Cirel is very dedicated and responsible too. She gives her best on everything she does: job, school, friendship, etc. At college she is the leader. She always has the best ideas and she knows how to organize things. Besides that, she is a very friendly person. She will always be there for you, and you can always count on her. She is also very funny. She is often in a good mood, and always making us laugh. I could write so much about her because she is so special. I'm so glad to have her as a friend and as part of my life.

25.

My father is a very moody person. He never says "hi" to people. Last night, a friend of mine came to our house to check his e-mails and my father didn't even look at him. He is always "busy". If you ask him something, he'll say to you that he can't do it now, because he is watching an important TV show. He never gives me a ride. Last week, I was going to the same direction as him, and he told me it wasn't possible cuz he was late. I know that people can't be emailing all day long, but that is too much. The problem is that I can't kick him out and look for another route. Certainly, I am the one who has to leave.

26.

My uncle Albert makes use of bad words too often. He speaks up his mind with f-words in such a rate that it seems like he does not have a previous concern about who is around, whether they are old ladies with a formal vocabulary or children that are still learning the language. He expresses himself thoroughly but frequently by informal means. One of these days he was talking to the chief executive officer of the company where he works at, he could not say a word, though. At first he thought he was a bit nervous but after a few minutes evaluating the situation he noticed that his lack of speaking came from an exaggerated use of impolite language. Not only this but he also does not try to improve his vocabulary storage perhaps because he thinks it would not be helpful once he sees himself as a mere truck driver. In other words he does not consider himself capable to be a gentleman in terms of speaking. In conclusion, Uncle Albert ought to practice his subconscious formal speech in daily chats in order to, as we all know, get a better professional profile and a less messy talking life.

27.

I'm going to write about a great friend that I have
~~about a great friend.~~

Juca is the best person I have ever met. We are friends
 since we were 10^{years old} and I really love him and his
 family. He got married six years ago and he has a
 boy (Gui), I am his (parent in law - padrinho) but I like
 to say why I love him a lot? Because he's just the
 trustfull person, friend, brother, honest, kind ^{in this world.}
 Do you know why he is all of this? Because he
 face the life ^{early} ~~with no stress~~. We know when we have or
 haven't got a friend. I know I like him and
 his son as not only friends but family. I'm
 luck to have them always by me.
 I'm sorry not writing enough about them but
 you have to know these boys to see and
 to understand what I mean.

28.

My best friend Elton,

Elton, my best friend, doesn't have responsibility. He plays drums with me in our band. Always I have to find a place to play or training our songs. He always say that doesn't have money to pay the studio or to buy his own drum or equipments which are basics to start playing, but he has money to buy other things that aren't important.

Also, he frequently has come late at studio, making us lose money and time. Last week, we have to ask to one friend of mine who is also drummer, ~~to~~ borrow his cymbals ~~and~~ his double pedals for us.

Everybody asks to me. "Why did you keep this man playing with you?" I don't know what to say, but I'm ^{sure} that he is the best drummer that I have seen, because he plays with his heart, with emotion, happiness, anger and rage.

He also has great ideas when we are composing something, but the best characteristic that I love in him is his loyalty with me and with the other guys.

Elton has a lot of bad habits, but his few good habits are more important to me, and one of his good habits is being my friend.

ANEXO 5 – COMPOSIÇÕES 2

2b.

I WAS USED TO LIVE SINCE I WAS THREE YEARS OLD IN A SMALL CITY IN SANTA CATARINA, CALLED JOACABA. IT IS SITUATED IN THE WEST OF THE STATE. THERE ARE NO RIVERS THERE BUT THE HOT CLIMATE IS LIKE THE CITIES WHICH THERE ARE. IT IS VERY HOT IN THE SUMMER, WITHOUT WIND BECAUSE IT LOOKS LIKE A HOX, VERY DIFFERENT FROM HERE IN CURITIBA. BUT IN THE WINTER IT BECOMES AS WARM OR COOLER THAN HERE.

THERE ARE NOT MANY INTERESTING THINGS TO DO THERE. TO VISIT, IT IS A NICE PLACE IN THE END OF THE YEAR WHEN THE PEOPLE PUT A LOT OF BEAUTIFUL THINGS, LIGHTS AROUND THE CITY, THE STORES ARE OPENED UNTIL TEN P.M. TO THE PEOPLE BY PRESENTS, AND THERE ARE A LOT OF SHOWS DOWNTOWN, LIKE DANCE, MUSIC, ETC. THE CARNIVAL THERE, IS CONSIDERED THE THIRD BEST OF SANTA CATARINA BECAUSE THERE ARE THE PEOPLE GOING THROUGH THE PRINCIPAL STREET WITH REALLY MARVELOUS ENTACTS AT SATURDAYS AND MONDAY'S NIGHTS, AND THERE IS THE YOUNG PEOPLE WHO STAY DOWNTOWN DANCING ALL THE NIGHT FOR FIVE DAYS.

I GO THERE ONCE A MONTH BECAUSE MY PARENTS LIVE THERE. THEY SAY THAT JOACABA IS THE BEST CITY TO LIVE BECAUSE IT IS SMALL, WITH 25 THOUSANDS PEOPLE, IT IS CALM AND THE PEOPLE ARE VERY POLITE, FUNNY AND HAPPY. OF COURSE EVERYBODY KNOWS ABOUT THE LIFE OF EVERYBODY, BUT IT IS A NICE PLACE.

THERE IS A GOOD UNIVERSITY THERE, CALLED UNOESC, IT IS VERY BIG AND IT IS ALWAYS GROWING. THERE IS A MOVIE THEATER AND A THEATER WHICH IS NOT VERY USED. THERE IS A SPACE SHOPPING THERE TOO. THE PROBLEM OF MY CITY IS THE PRICE OF EVERYTHING. EVERYTHING IS EXPENSIVE, I THINK IT IS BECAUSE IT IS REALLY SMALL. BUT THERE IS A GOOD PART, THE ASB; THERE IS ONLY ONE, AND EXPENSIVE, BUT IT IS VERY GOOD. THERE ARE ANOTHER ONES AND THINGS LIKE THAT WHICH ARE GOOD TOO.

MY CITY IS A VERY GOOD ONE. I LIKE IT VERY MUCH AND ALL THE PEOPLE WHO WENT THERE FOR THE FIRST TIME AND TALKED TO ME LIKED TOO. IT IS A GOOD PLACE AND THE PEOPLE ARE REALLY VERY KIND. TO THE YOUNG PEOPLE I RECOMMEND THE CARNIVAL OF JOACABA. IT IS MARVELOUS.

5b.

Montreal is the name of a city in Canada. There is possible to speak either English or French. The best time to visit this city is in spring. All the flowers start to grow, the city looks like a rainbow and the longest day of the year starts to get dark around 21h30-22h00pm. In the summer is really delicious. Everyone goes to the park with bathing suits and rollers. The restaurants and bars put chairs and tables on the sidewalk. The temperature is not very hot too. Fall is amazing. All the leaves are falling and the streets get full of them. The asphalt is almost impossible to be seen. The worst season is winter. Montreal is the biggest city underground because it is one of the coldest cities in the world. The day is cloudy and it has light only until 4h00 o'clock in the afternoon.

There are many beautiful parks to visit in Montreal. The cinemas are built in huge buildings and each floor shows a different kind of movie. The nightclubs are different. They open at 5 o'clock in the afternoon. You do not need to pay to get in, but you pay at the time you order for something plus taxes to the waiter. They close very early and after this there is still fun in other nightclubs that open only when the other ones close. The gyms are 24 hours so there are no excuses for not having time to exercise.

Near Montreal there is a city called Quebec. It is a French city. They do speak English but they do not like to. There is an old part that looks like decades ago. And there is also the new part where there is a lot of technology. The people in Montreal are nice. They trust a lot in each other. They are very honest but do not believe in everyone, especially foreigners. They only become friends when they know the person very well. They are a little cold, not many hugs and kisses. They show their feelings in a different way. The food is great. The traditional food in Montreal is: "poutine". It is a kind of French fries with a cheese cream on it. But the cream can also be something else. There is a restaurant called Moneypick in English and La Marche in French. This place looks like a city because you have little tents from different countries and the cooks are native from their countries. In June starts the jazz festival. Lots of people insoube themselves to play any instrument and some streets are closed for that and the

participating does not need to be a famous musician.

Montreal is wonderful. It looks like Curitiba because of the parks and flowers. The city is very clean and safe. If one day you decide to visit there, go in June, respect the rules because they are very severe and visit as many places as possible. Try to rent a car and travel to somewhere near so you can see how pretty is the view in the road and how careful are the drivers. This city is also interesting because you can learn two languages: English and French. To travel to somewhere else is also good to see how good the city we live is nice.

8b.

There are many beautiful places that I have been to. I love traveling and every time I can I pack my things, get together with some friends and take a trip. I think that, among all the places I have already been to, the most beautiful is Praia do Rosa in Santa Catarina. Actually, not only this beach but also other ones located in the south of the state, like Itiraquera, Mirim and Silveira are wonderful places.

Praia do Rosa is a very known beach because of its waves. Surfers love going there. It is a small beach but it is full of beautiful and interesting places nearby. First of all, most of the hotels there are located at the top of a mountain from where you can see the sea and the beach. The view from this mountain is terrific. You wake up in the morning, open the window and you see a paradise outside. A lot of trees and down there the ocean and the beach. There is a delicious restaurant there called "Tigre Asiático". They serve all kinds of Asian food. The food is great and the decoration of the place is very beautiful. Even the waiter are dressed in typical clothes. All that makes you feel you are in Asia.

During the months of September and October, a lot of people go to Praia do Rosa in order to see the "Franca" whale. You can ride a boat and get very close to the whales. You get to see these huge animals very close to you and it is fantastic. They also have a project there in order to

If you go to Praia do Rosa, there is one last place you should visit and it is called Itiraquera. This beach is located near a lagoon and you can see the place where this lagoon finds the sea and this is very nice. There are many trees on this beach and it is a very large beach. You can walk a lot and never get to its end. The place is also very quiet and calm. Few people know about it so it is still a not explored place. It is a very pleasant place to rest and think, specially if you are with someone you like.

There are many other beautiful places that I have been to but this beaches are the ones I have liked the most. There you are able to live in peace, at least for some time, and enjoy yourself around nature. They are places that make you feel better, calmer and happier. Places that you can travel to any time of the year and have a great time.

12b.

Empório São Francisco is the "bar" of my life. It is near to Tom's Final and Tio Tex (2 another options if you want to drink and have fun) at Largo do Ordem, Alto São Francisco. It is a famous place, in fact, it is a famous place in the rock scene. Not the hard rock style but the soft one, like Casa das Máquinas, Mutantes, Secos e Molhados and legends of rock, of course, like Led Zepellin, Jimmy Hendrix and Beatles.

If you want to have a sit you better get a table reservation because about 11:00 o'clock it's already impossible for you to sit and also, at this time, the air inside of it is all cloudy by the smoke's cigarette. Sorry I forgot that they now have that air-conditionated thing, so you won't have this problem anymore. As soon as you get there, stop at the bar and ask for Renato to get you a beer, then go: jump around with your beer in your hands (no glasses for those who are not in tables) and have fun singing with the greatest band of Curitiba. Sexofone is on at Saturday's night.

Start your evening at Empório. It is a small bar in front of it, 3 tables and the greatest "espeterinho" of the area nearby. Go there, have the first beer then you are ready to enter at Empório. If you get there early, look at the walls, it is full of musicians portraits. Really beautiful thing made by local artists. There are also many candles the light the kind of the red walls.

When you get tired of singing and clapping, have a sit (at this time I am sure that will have a table available) and drink a really nice speciality of the house. It is called B-52. It is a mix of 3 drinks plus Absinto, they put fire on it and you drink it with a tube. Try it. I forgot to say they every Sunday they have a special show, like a tribute to a special band. If you have nothing to do at night, any night, go there and have fun.

15b.

Two years ago in my vacation of July, I went to Caxias do Sul, at Rio Grande do Sul, in Brazil to visit some relatives.

The city is not very big, but it is known by "The city of grapes". There ~~was~~ were only two shopping centers that time. I visited both which are good and not very big though. My cousins, my sister and I went to the cinema in one of them to watch "Jurassic Park 3". It was very funny. Then we had a snack of friend-fries, some meat and soda. Instead of going to the shopping center you can go to a night club. There are many with any kind of music you want. I went to one that played only rock and Techno music. It was really cool. The people there were very warm and I could make some friends.

Besides that they are traditional due to their typical music and dancing and also because their food. They have some C.T.G (Centro de Tradições Gaúchas) festivals where men and women wear in traditional costumes. Because they were colonized mostly by Italians their food is made with a lot of pastry. I could enjoy their "Sopa d'Alhoine" and the "Tortelli", both are wonderful. Also the way they dance at C.T.G is very similar as Italians dance.

Also the view there is amazing. There are many hills and green pastures and many trees. It is really beautiful. Depending on the weather it can become more interesting: if it is winter, sometimes you can see the snow, if it is spring the view become more colourful and if it is summer you can enjoy some summer clubs and beaches there. It was winter when I went, but I can say that ⁱⁿ any weather you like the most, you can have a great time there.

21b.

There is a very nice place to visit on vacation close to Curitiba. This place is São Paulo, the biggest city in Brazil! There, you can find all the good things you find abroad, in cities like New York, London and Paris as well.

You can find in São Paulo people from all over the world as well as their culture, and it makes São Paulo such a interesting city to visit to.

People are used to saying that São Paulo seems to be a stressed city, but for sure it depends on people's point of view, because if you relax and try to find the good points of São Paulo, you will certainly find lots of it.

Despite of the heavy traffic and crowded places you can find the best restaurants, movies, shopping malls in a city that does not sleep. If you have never taken a subway riding, it is a good opportunity to do it in São Paulo.

So, there are lots of good things to say about São Paulo, but you will only realize that going there and appreciating this amazing city.

22b.

I am going to talk about a country that I not only visited but lived too. I have been five times in Germany and lived there for three years. My parents are Germans and moved to Brazil more than 30 years ago and they had to go back to Germany every five years because of my father's work. We also have all our relatives living there. Just two cousins live in Rio Grande do Sul.

The two first times I went to Germany I was a child and I don't really remember that much. So, I am going to talk about the year that I spent there when I was fifteen years old.

First I did not want to go to Germany because it was a strange country and all my friends were in Brazil. The beginning was not easy. Everything is different: the people, the weather, the culture and food. I had to go to school and the system is very different as in Brazil. I had classes from 7.30 a.m. until 3.30 p.m. Sometimes and I had not only Math, Physics, History and this kind of subjects like here in Brazil. I had different kinds of sports, I learned how to make ceramic flower-pot and a lot of interesting activities.

Germany is a very beautiful country. It does not have beaches like in Brazil but it has amazing castles and very old cities that still have old buildings and wonderful Churches. The winter is very beautiful when the snow falls on the houses and everything becomes white. The autumn is also a very beautiful season with all the trees losing their leaves.

German people are well known because they are more cold in comparison to the Latin people. But they are reliable and always on time and they are very open to learn more from other cultures. Germans like to travel a lot. Some of them travel to a different country every vacation. They like to taste different kinds of food and bring the recipe home. You can find any kind of food and drinks in Germany. It is really worth to know this little country in Europe. At summer, winter, autumn or spring, there are always nice places to go out and have fun.

23b.

Minas Gerais is a beautiful and interesting state because there people can visit historical cities, they can taste typical food. And also, people who live there is very friendly.

Visiting the historical cities of Minas Gerais, tourists learn much about the Brazil history because it was on these places that began the history of our country. An example, Ouro Preto was the first capital of Brazil. These cities have an old architecture:

- There are much gold in the churches, there many draws on its walls and roof. Some of them keep caves above their floor.
- Streets are built from stones.
- There is not a garden in front of the houses.

Belo Horizonte is more modern because it is capital of Minas Gerais. There, people can have fun going at nightclubs, malls, theaters, movies.

Minas Gerais is very famous because cheese bread that is more popular and traditional food of there. But there are other dishes as delicious as cheese bread such as "queijada", "cabelo" Ticken, "mixidinho", "tutu de feijão" and a lot of candies, for example pumpkin candy, coconut candy, milk candy.

And for people who like to drink there are many "cachacas" for them have fun and for them taste a traditional "cachaca mineira".

Visiting Minas Gerais people can stroll by many different places and they can meet the locals, who are very funny and friendly people.

24b.

In September my boyfriend invited me to spend Brazil's Independence Day in a very nice resort called Aguatã. I was so excited to go because I heard the place was so much fun, and I also needed to relax because I was very tired from work. We left on September 6th early in the morning, because the resort is in Carmo do Rio Preto - a city in Paraná, and the trip takes five hours. The day was so beautiful and the trip was calm.

We arrived there in the afternoon, and the first thing we decided to do was to take a look at the swimming pool because it was very hot. There were many pools, one with a big slide, another one heated, and one with a bar inside it. We got the keys of our room, put on some swimsuits and went to the pool. It was awesome! The room was also very nice, there was air-conditioning, a television and a nice bathroom. After the pool, we had dinner in the hotel restaurant and the food was delicious! There were many options, we enjoyed it a lot.

The next day we discovered that there were many activities to do for fun, such as: walking around the resort, fishing, horseback riding, volleyball, tennis, climbing, dance classes, canoeing, and much more. We spent the day having so much fun doing all these activities. At night there was a very good concert with a guy playing Brazilian Popular music. We also enjoyed an Italian dinner. This was another beautiful, sunny and nice day.

The next morning was already time to return to Curitiba. It rained a lot during the trip, but that was not a problem because I had a very interesting book to read called *Wuthering Heights* and I read the whole book during the trip. When we arrived in Curitiba it was already dark, but I could not wait the next morning to call my friends and tell them about this wonderful place called Aguatã where you can relax, do many different activities, eat delicious kinds of food and feel so happy!

26b.

EVERY YEAR LOTS OF TOURISTS HEAD A SMALL CITY IN ITAJAI VALLEY, CALLED RIO DO SUL. THIS FASCINATING CITY HOLDS OVER SIXTY THOUSAND INHABITANTS SPREAD AMONG ITS SEVERAL MOUNTAINS AND FORESTS, BIG PART OF WHOM TRACE EUROPEAN BACKGROUND. JUST 100 MILES AWAY FROM SANTA CATARINA CAPITAL, FLORIANÓPOLIS, VISITORS FROM DIFFERENT COUNTRIES SPEND EITHER A WEEK OR THE WHOLE SUMMER TIME LOOKING FOR FUN IN THIS GREAT PLACE.

SOME OF THE VISITORS COME ON THE VERY COLD SEASON WHILE OTHERS PREFER THE HIGHEST HOT TIME IN DECEMBER. THE ONES WHO WOULD RATHER ENJOY THE WINTER IS BECAUSE ITAJAI-AÚ RIVER FLOW GETS HIGHER FOR WATER RAFTING FANS, MEANWHILE SUMMER SPORTS LIKE RAFTING AND SWIMMING ARE MORE PRACTICED IN THE SUMMER.

EXTREME SPORTS ARE ALSO SEARCHED BY MOUNTAIN CLIMBING, HIKING AND CYCLING PRACTICERS. NOT ONLY THIS, BUT NIGHT LIFE IS PRETTY EXCITING. THERE ARE SEVERAL BARS YOU CAN VISIT SUCH AS GUINNESS AND BELLA. IF YOU PREFER A CALMER PLACE DON GIOVANNI RESTAURANT WILL LEAD YOU TO LOVE MOMENTS IN A SAFE AND UNUSUAL ENVIRONMENT.

YOU CANNOT MISS THIS FRIENDLY AND WELL-STRUCTURED PLACE. ALONG WITH THE OLD GERMAN STYLE HOUSE YOU WILL FIND PURE AIR AND A GREAT FESTIVAL IN OCTOBER CALLED KEGELFEST. SINGLES AND COUPLES ENJOY FROM EXTREME SPORTS PRACTICING TO A WARM LOCAL POPULATION CONTACT IN AN ONLY PLACE, RIO DO SUL.

28b.

On last summer, I had the gracefull oportunity to visit the capital of Santa Catarina, called Florianópolis, this name is because this city was establish by Marechal "Floriano" Peixoto.

Florianópolis is located on south of Brazil, four hours from Curitiba by car or bus, but if you want to go by plane, the flight takes half an hour. The city is divided in a place that the people there call continent, and another place called Island, where there are the best beaches of Brazil, for example Joaquina's beach where occur the Surf world championship, praia dos Ingleses, where you can rest on the sand, play with the kid, swimming in a safe place and also, eat fruits of sea and all kinds of fishes, but if you are thinking in peace, relax, you can go to the beach called Armação do Sul, in this beach there is a place which you can stay naked.

The best time of the year to visit is on summer, where you can go to the beaches, go fishing, pratice sandboard and eat on the restaurants, or have a fun at nigh in a lot of pubs on Barra da Lagoa. If you want to profit your tourism, go by car or you can rent a car because the beaches are far from each others, and you must stay in Florianópolis one week at least to know all the city.

My trip was wonderfull, and I had been there for two weeks, and now when I want to go to the beach, I think Florianópolis "the magic land", and I am waiting the next holyday to go there.

ANEXO 6 – QUESTIONÁRIOS RESPONDIDOS PELOS CORRETORES

A. Introdução

1. Corrigir as composições em ordem numérica crescente.
2. Caso haja pausa entre as correções, anotar o número da última composição corrigida e os horários de término e reinício das correções.

B. Informação sobre o corretor – a

1. Idade: 41

2. Descrição breve do seu histórico profissional e experiência com a correção de composições.

Formação Universitária – Licenciatura em Letras – Inglês e Licenciaturas PUC Campinas-SP; Professora de Inglês em Cursos de Idiomas 1988-2004; Supervisora 1999-2004; Participação em diversos eventos na Área de Educação e mais especificamente na Área de Língua Inglesa; Experiência em correção de composições em Cursos de Idiomas/Processo de Escrita

3. Você se considera um corretor exigente ou permissivo? Justifique.

Observo com certo apreço questões relacionadas ao Conteúdo, onde o escritor tem em foco o leitor fazendo com que o mesmo tenha facilidade e interesse na leitura. Para isso ocorra a necessidade de coerência e coesão são fatores indispensáveis na escrita.

C. ASPECTOS SOBRE CRITÉRIOS DE CORREÇÃO

1. Que tipos de critério de correção você já utilizou em sua experiência profissional?

Critério holístico e analítico

2. Você compreendeu claramente o critério de correção adotado neste estudo?

Sim

3. Na sua opinião você teria utilizado um critério diferente? Justifique?

Não. O critério da correção holística é mais preciso quando se observa coerência e coesão. O critério analítico é mais específico para direcionar o escritor a observar questões tal como organização: introdução, desenvolvimento e conclusão; gramática, vocabulário, pontuação, ortografia.

4. Quais são os aspectos positivos e negativos de utilizar um critério holístico?

O critério analítico tem como aspecto positivo a maior precisão ao apontar questões tal como (1) gramática (2) vocabulário (3) ortografia (4) pontuação (5) etc. Em se tratando de artigos, ensaios, etc. o critério holístico é mais funcional, pois tende a enfatizar aspectos relevantes para uma boa compreensão do texto.

5. Quais são as limitações deste critério?

O critério holístico tem como limitações quando utilizado para correções com objetivos específicos de sintaxe, morfologia.

D. ASPECTOS E IMPRESSÕES SOBRE O ATO DA CORREÇÃO

1. A que horas você corrigiu as composições?

Pela manhã.

2. Todas as composições foram corrigidas em um mesmo momento? Caso contrário, explique os momentos de correção.

Todas as composições foram lidas e alguns comentários foram feitos nessa primeira leitura.

3. Como você se sentiu:

a) no início da correção?

Procurou observar quantas composições são ao todo antes da correção. Fiquei um tanto surpresa (28), pois tinha como determinação fazer a primeira leitura das 28 composições em um mesmo momento – pela manhã.

b) na metade da correção?

O tema facilitou, pois senti maior envolvimento com os escritores ao descreverem como eles vêem e julgam os hábitos de pessoas mais próximas.

c) no final da correção?

Tarefa realizada! O que me impressionou nas leituras (1) como as pessoas têm problemas de relacionamento (2) as conclusões dos parágrafos foram melhores que as introduções.

4. O que você já havia feito durante o dia até o início das correções?

Primeira Leitura: Acordei às 6:30, tomei café da manhã, cheguei no trabalho as 08:00, li um artigo “The English Teacher as a “VIP” in Student Motivation,” conferi os e-mails do dia e dei início as correções.

Segunda Leitura: Resultado de 1 a 6. Ao fazer a segunda leitura, enfocando o resultado/pontuação de 1 a 6, observei que a organização do parágrafo é lógica, conteúdo o que chamou mais a atenção foi o uso e o emprego correto de palavras que venham a enriquecer o conteúdo, exemplo: (#4) ...extremely annoying..., ...dared to talk..., Eventhough, (#6) ...tough to manage. Due to his Constant delay..., ...for the sake of a happy marriage. (#26) ...in such a rate..., ...he could not say a word, though, ...his lack of speaking came from..., “In conclusion, Uncle Albert ought to practice his subconscious formal speech in delay chats in order to, as we all know, get a better professional profile and a less messy talking life. (Just great!)

5. Em que dia da semana você efetuou as correções?

Primeira leitura: 04 de outubro de 2004 – segunda-feira

Segunda Leitura: 06 de outubro de 2004 – quarta-feira

6. Em que local você efetuou as correções?

No trabalho

7. Não conhecer os alunos ou os seus nomes, fez diferença para a correção?

Sim. Pois o não conhecer faz com que você observe o conteúdo/a mensagem. Vice não tem a imagem do escritor “falando”. Você não sabe das limitações que o escritor possa ter, como limitação de linguagem, comportamentos, ou mesmo físico.

8. Qual a sua reação ao atribuir notas de 1 a 6, e não de 0 a 10?

Quanto menor a variação na escala de notas, mais precisão ao atribuir as notas.

9. Quais são os aspectos positivos e negativos de utilizar uma escala de correção numérica menor (de 1 a 6)?

Aspectos positivos: precisão, funcional, poucas variações

Aspectos negativos: compreensão por parte não apenas dos alunos, mas dos pais quanto a utilização da escala de 1 a 6

A. INTRODUÇÃO

1. Corrigir as composições em ordem numérica crescente.
2. Caso haja pausa entre as correções, anotar o número da última composição corrigida e os horários de término e reinício das correções.

B. INFORMAÇÃO SOBRE O CORRETOR - B

1. Idade: 35

2. Descrição breve do seu histórico profissional e experiência com a correção de composições.

Sou professora de inglês há quinze anos e tenho uma experiência considerável no que diz respeito a correção de redações: parágrafos e ensaios. Todavia, há aproximadamente dois anos não tenho trabalhado com turmas que produzam redações, muito embora durante este tempo eu tenho trabalhado preparando trabalhos de redação para serem desenvolvidos na sala de aula de curso de língua para níveis intermediários.

3. Você se considera um corretor exigente ou permissivo? Justifique.

Considerando os dois pólos, acredito que tendo a ser permissiva, mas sempre procuro dar feedback escrito sobre os erros cometidos assim como sobre o conteúdo e sua adequação. Acredito que seja importante o aluno e expressar suas idéias, mas entendo que dependendo do nível, avançado por exemplo, não se pode ignorar alguns problemas de escrita, principalmente se estes tornarem o texto incompreensível. Sinto-me impotente em relação à correção quando não é possível dar ao aluno feedback, no caso de testes que não retornam ao aluno. Nesta situação reconheço que tendo a ser mais permissiva.

C. ASPECTOS SOBRE CRITÉRIOS DE CORREÇÃO

1. Que tipos de critério de correção você já utilizou em sua experiência profissional?

Crítérios holístico e analítico. Também já corrigi redações sem fazer uso de nenhum critério preparado previamente, apenas decidindo a nota conforme a quantidade de erros e qualidade do conteúdo.

2. Você compreendeu claramente o critério de correção adotado neste estudo?

Sim, mas tenho dúvidas sobre os diferentes níveis, especialmente no que diz respeito à pontuação. Entendo que este seja um critério para alunos de nível intermediário-alto e avançado. Gostei dos termos: proficiente, fluente, em expansão e em desenvolvimento. Acredito que estes termos auxiliam aluno e professor a entender melhor o estágio no qual se encontram. Além disso, há um lado muito positivo em relação ao lado afetivo/emocional: estar “em desenvolvimento” não significa, para mim, “insuficiente” ou “fraco”.

3. Na sua opinião você teria utilizado um critério diferente? Justifique?

Não. Acredito que este critério seja coerente, útil e abrangente.

4. Quais são os aspectos positivos e negativos de utilizar um critério holístico?

Positivo: avalia-se a capacidade do aluno de se comunicar, sem prender o corretor a minúcias.

Negativo: o aluno muitas vezes necessita de um feedback sobre as minúcias. Maior explicitação de problemas de língua que podem interferir no texto, na sua compreensão.

5. Quais são as limitações deste critério?

A única limitação que vejo é em relação a nota “1 a 6”, por motivos culturais.

D. ASPECTOS E IMPRESSÕES SOBRE O ATO DA CORREÇÃO

1. A que horas você corrigiu as composições?

Durante o dia, após o almoço em dias diferentes.

2. Todas as composições foram corrigidas em um mesmo momento? Caso contrário, explique os momentos de correção.

Não. Em dias diferentes, pois não foi possível me concentrar em todas as redações de uma só vez. Houve um momento em que desisti de corrigi-las pois senti que o resultado não seria satisfatório.

3. Como você se sentiu:

a) no início da correção?

Entusiasmada, mas pensando no trabalho de corrigir 28 redações.

b) no metade da correção?

Preocupada com o meu critério, se as notas eram coerentes. Preocupada em não ter muitas notas “4”

c) no final da correção?

Satisfeita por ter terminado, mas ainda preocupada com as notas. Ainda preocupada coma nota “4”

4. O que você já havia feito durante o dia até o início das correções?

Trabalhos diversos: preparação de provas, organização de um teste e workshop. Assim como no último dia aplicação de testes para 70 pessoas e um workshop para 11 pessoas.

5. Em que dia da semana você efetuou as correções?

05 de outubro de 2004 (terça-feira); 06 de outubro de 2004 (quarta-feira); e 08 de outubro de 2004 (sexta-feira)

6. Em que local você efetuou as correções?

Na sala onde trabalho como supervisora.

7. Não conhecer os alunos ou os seus nomes, fez diferença para a correção?

Acredito que sim, mas tenho a curiosidade de saber se as histórias relatadas são reais. Acredito que não conhecer faz grande diferença, pois os alunos não estão previamente “rotulados” como bons, proficientes, insuficientes.

8. Qual a sua reação ao atribuir notas de 1 a 6, e não de 0 a 10?

Observando o critério não há maiores limitações, mas algumas vezes me perguntei o que o “4” representaria numa escala de “0 a 10”. O seis (6) seria um “10”, mas quanto ao “4” tive dúvidas.

9. Quais são os aspectos positivos e negativos de utilizar uma escala de correção numérica menor (de 1 a 6)?

Positivo: a diferença entre um aluno 6 e um aluno 3 não é assustadora, mas alguma maneira indica que quem é “3” pode fazer avanços consideráveis.

Negativos: a questão cultural: desde a infância nos acostumamos com esta escala. Em minha experiência, quando fui avaliada com a escala alfabética: A-F, sempre precisei relacionar com a escala numérica de 1 – 10 para “entender” melhor a nota atribuída.

A. INTRODUÇÃO

1. Corrigir as composições em ordem numérica crescente.
2. Caso haja pausa entre as correções, anotar o número da última composição corrigida e os horários de término e reinício das correções.

B. INFORMAÇÃO SOBRE O CORRETOR - C

1. Idade: 28

2. Descrição breve do seu histórico profissional e experiência com a correção de composições.

Professora de língua inglesa há quatorze anos, conseqüentemente, corrigindo composições pelo mesmo tempo..

3. Você se considera um corretor exigente ou permissivo? Justifique.

Me considero exigente em relação ao teor ou conteúdo de uma composição e um pouco mais permissiva em relação a estrutura e mecânica.

C. ASPECTOS SOBRE CRITÉRIOS DE CORREÇÃO

1. Que tipos de critério de correção você já utilizou em sua experiência profissional?

Já trabalhei em instituições onde não se aplicava critério algum e há oito anos trabalho com critérios minuciosos.

2. Você compreendeu claramente o critério de correção adotado neste estudo?

Sim.

3. Na sua opinião você teria utilizado um critério diferente? Justifique?

Não, essa forma vem a ser muito prática para o uso dos corretores.

4. Quais são os aspectos positivos e negativos de utilizar um critério holístico?

Acho negativo não avaliar cada um dos elementos da escrita separadamente. Mas por outro lado, dessa forma toma menos tempo.

5. Quais são as limitações deste critério?

Nem sempre se pode traçar uma média no uso dos diferentes elementos da escrita de forma a enquadrar o redator em apenas uma das classificações da tabela.

D. ASPECTOS E IMPRESSÕES SOBRE O ATO DA CORREÇÃO

1. A que horas você corrigiu as composições?

15:40.

2. Todas as composições foram corrigidas em um mesmo momento? Caso contrário, explique os momentos de correção.

Houve pausa as 17:00 e retomada as 19:00.

3. Como você se sentiu:

a) no início da correção?

Confiante e certa com relação à classificação das composições.

b) no metade da correção?

A medida que lia mais e mais composições com problemas diferentes e padrões diferentes ficava em dúvida na hora de classificá-las.

c) no final da correção?

Sensação de que poderia não ter sido igualmente justa com todos e devesse rever as classificações.

4. O que você já havia feito durante o dia até o início das correções?

Coisas de ordem pessoal e lecionado uma aula de revisão para prova.

5. Em que dia da semana você efetuou as correções?

Terça-feira

6. Em que local você efetuou as correções?

Sala dos professores da instituição onde leciono.

7. Não conhecer os alunos ou os seus nomes, fez diferença para a correção?

Não, creio que não.

8. Qual a sua reação ao atribuir notas de 1 a 6, e não de 0 a 10?

Dificultou muito mais, quanto maior a escala, mais podemos marcar a diferença de qualidade entre uma composição e outra.

9. Quais são os aspectos positivos e negativos de utilizar uma escala de correção numérica menor (de 1 a 6)?

Talvez seja mais preciso, no entanto, me senti mais exigente, talvez injusta com alguns e é uma variedade muito pequena. Existem muitos diferentes graus de qualidade de escrita.

ANEXO 7 – TRANSCRIÇÃO DA DISCUSSÃO SOBRE AS NOTAS DADAS ÀS COMPOSIÇÕES DE NÚMERO 1 PELOS CORRETORES A E B

- Mediador: A tarefa a ser feita é discutir as notas atribuídas por vocês corretores A e B às composições dos alunos do Curso de Letras da PUCPR, nível pós-intermediário e avançado. Conforme estabelecido no estudo, as turmas de Letras são divididas em 3 níveis, G1, G2 e G3, que apesar de discriminar o grau de habilidade dos alunos não é comparável aos cursos de línguas, que dividem seus alunos em até 12 níveis tão discriminatório assim. Por meio da discussão e replanejamento das notas, vocês devem tentar alcançar um denominador comum.
- Mediador: À primeira composição o corretor A atribuiu nota 2 e o corretor B nota 4.
- Corretor B: O que eu considerei, algo que me chamou atenção no critério foi a divisão de habilidades, fluente, em desenvolvimento, em expansão, iniciante e elementar. Eu atribui a nota 4 porque eu entendi que este escritor é uma pessoa em expansão, apresentou alguns problemas de ortografia, ele muda de assunto, tem alguns problemas, mas para mim se encaixa dentro dos critérios apresentados aqui. Uma questão importante para mim é que talvez revendo eu pudesse mudar realmente de opinião, mas...
- Corretor A: Ao avaliar as redações eu tive em mente como objetivo o grau de proficiência, sabendo que são 3 níveis, mas considerei o nível dos alunos Upper-Intermediate e Advanced. À composição 1 eu dei a nota 2, ele é iniciante, isso significa que ele apresenta um tópico, que é “listening to music very high”, quando ele desenvolve esse tópico ele coloca que o companheiro de quarto aprecia outro tipo de música, então eu fiquei em dúvida como leitora, se foi a música ser alta ou se foi o gosto musical. Quando ele coloca isso, ele sai fora um pouco do

tema e tem dificuldades de vocabulário, apresenta erros de ortografia, o vocabulário dele é limitado e um pouco repetitivo, então eu considero ele como iniciante. Ele desenvolve sim, mas foge um pouco do assunto e como leitora eu não me senti atraída pela composição.

Mediador: Se vocês tivessem que atribuir uma nota apenas para a correção 1, qual seria esta nota?

Corretor B: Eu baixaria para 3, mas acho que não chegaria em 2, eu acho que considerando a questão “em expansão, em desenvolvimento”, eu o consideraria como em desenvolvimento, ele tem possibilidade de melhorar com um feedback apropriado e reestruturar, considerando que isso foi uma única versão do que ele escreveu.

Corretor A: Levando em questão o critério holístico, você tem aí vários itens, então é a combinação de todos esses itens, mas o escritor tem que ter em mente principalmente o leitor, quem está lendo a redação. Então aqui eu mantenho a nota 2, ele é iniciante.

Mediador: A nota do corretor A é mantida – 2, e a do corretor B é alterada para 3. No entanto, ainda assim não foi possível chegar em um denominador comum, e a nota 2.5 não faz parte da escala. Composição 2, corretor A atribuiu a nota 2 e B a nota 3.

Corretor B: Uma dificuldade dessa redação foi visualizar, a letra. Houve essa limitação, mas mesmo assim foi possível identificar alguns erros de ortografia, erros gramaticais, foi percebido também que ele faz uso de algumas palavras de transcrição, como “por exemplo”. Ao meu ver a nota foi 3, pois ele está em desenvolvimento, redige frases sobre uma idéia, apresenta alguma seqüência, falta coesão em algumas coisas, realmente. Tem problemas com ortografia e outras questões, mas seria um escritor em desenvolvimento.

Corretor A: Foi difícil ler pela questão do “handwriting”. Os exemplos que ele cita são exemplos muito simples, em outras palavras são pobres, ele tem

muito erro de ortografia, erros de gramática e pontuação, então houve por parte do leitor, dificuldade em ler. Mencionando novamente o critério holístico, ele tem várias questões a serem consideradas, mas a questão importante aqui é atrair o leitor.

Mediador: Composição número 3, corretor A nota 3 e corretor B nota 5.

Corretor B: Eu percebi que, quem escreveu fez uso de três idéias diferentes, fez uso de vocabulário de transição, esse foi um dos motivos e a linguagem é simples, mas leva a mensagem, ele consegue se comunicar, esse foi um dos fatores que eu considerei, então acabei atribuindo a nota 5, ele redige parágrafos únicos, fica focado na idéia central, emprega os tempos verbais, não é tão variado mas é preciso, ele utiliza o que precisa utilizar para essa situação.

Corretor A: Eu atribui a nota 3, ele está em desenvolvimento, ele redige frases sobre uma idéia, as frases são frases simples, tem senso de organização, mas está em desenvolvimento por questões de linguagem, por questões de muitas vezes coerência e ao redigir.

Mediador: Existe alguma alteração nas notas das composições número 2 e 3 de vocês?

Corretor B: Não, talvez o número 3 possa passar para 4, revendo, pois eu fico com a questão da comunicação, ele consegue elaborar o que ele tem que falar e a mensagem. Acho que um limitador que eu tenho é talvez como eu tenha nivelado os alunos como intermediário e upper-intermediate, essa questão que deve estar influenciando.

Corretor A: Eu mantenho a nota 3, em desenvolvimento.

Mediador: Composição número 4, os dois corretores atribuem a nota 5, eu gostaria de saber o porque?

Corretor A: 5, fluente, o que chamou muita atenção aqui, apesar da organização das idéias e de redigir frases sobre uma idéia e desenvolver essa idéia, foi a questão do vocabulário, ele usa aqui um vocabulário rico, variado, que atraiu o leitor.

- Corretor B: Concordo com a colocação de que ele usa vocabulário mais elaborado, muito embora existam algumas questões de ortografia. Num determinado momento ele se desvia dos assuntos para justificar as suas colocações, mas ele tem o tópico e desenvolve bem e o vocabulário, é uma das coisas que chama a atenção.
- Mediador: Composição número 5, corretor A nota 4, corretor B nota 5.
- Corretor A: Nota 4, em expansão, ele organiza as idéias aqui em uma ordem lógica, mantém o desenvolvimento das idéias e os exemplos em cima de uma idéia. Ele usa vocabulário adequado, apesar de ser simples, mas em alguns momentos ele elabora melhor o vocabulário, isso daí também atraiu o leitor.
- Corretor B: Ele apresenta uma combinação entre a sentença inicial e a conclusão, elas combinam, o que ele falou no começo ele retoma no final, apesar de apresentar um pouco de falta de palavras de transição e usar vocabulário simples, o escritor consegue transmitir a mensagem, se organizar de maneira simples, mas atinge o seu objetivo.
- Mediador: Composição número 6, os dois corretores atribuem a nota 6 da escala.
- Corretor A: Nota 6, ele é proficiente. Desde o início ele atrai o leitor, cria uma expectativa para que você continue a leitura, só pelo início da composição o vocabulário usado é rico. Ele tem alguns erros de mecânica e gramática, mas isso não interfere no significado, a conclusão foi extremamente bem desenvolvida, ele fechou nota 10.
- Corretor B: Ele apresenta uma seqüência lógica de idéias, existe uma riqueza de detalhes, não apenas isso, mas de vocabulário. Tem um bom início, tem um “topic sentence” na segunda sentença, termina com uma conclusão, então acredito que aqui é realmente a questão da seqüência de idéias, um vocabulário rico e a maneira como ele trabalhou. Percebi que apresenta alguns erros de ortografia, mas nada que vá influenciar na leitura, no entendimento do texto.
- Mediador: Composição número 7, corretor A nota 2 e B nota 4.

Corretor A: Nota 2, iniciante, quando ele menciona a palavra "smart" ele coloca como exemplo "open-minded" e "tireless", então eu fiquei questionando se esses exemplos estão realmente definindo o tópico que é "smart", "he's a smart man." Quando ele fecha o parágrafo aqui, ele coloca 2 razões para justificar a abertura, o "topic", essas duas razões são iguais, "because of his open mind" e "because he's always open to new ideas", para mim elas são a mesma coisa.

Corretor B: Eu acredito que uma questão importante aqui é a questão da coerência, ao meu ver ele começa com o tópico que é o pai, ele vê o pai de uma maneira positiva e ele vai até o final colocando as qualidades positivas do pai, e ele apresenta quais são essas qualidades. Faz uso, tenta fazer uso, de algumas "transitional words", mas acaba errando, no entanto, tenta fazer uso de técnicas que aprendeu. Considerando mais alguns erros que apresente, eu poderia abaixar a nota para 3, mas ainda mantenho que apresenta uma determinada coerência, não é iniciante, mas em desenvolvimento. De 4 para 3.

Mediador: Composição número 8, corretor A - 4, corretor B - 6.

Corretor A: Aqui o escritor está em expansão, nota 4, ele usa uma boa organização, faz uso de exemplos para mostrar o "topic sentence", apesar de os exemplos serem apropriados eles são muito simples para o nível upper-intermediate/advanced, então nota 4.

Corretor B: Eu mantenho a nota 6, porque eu acredito que esse escritor apresenta coerência, vocabulário bem desenvolvido, as idéias refletem o que a pessoa quer falar. Ele mantém uma certa coerência sobre a questão da mãe, exemplifica com questões da vida familiar, que podem ser simples, mas vêm a contribuir para que ele ilustre quem é essa mãe, é rico em detalhes, faz uso de algumas "transitional words", então nesse caso, eu mantenho a nota 6. Ela é proficiente.

Mediador: Composição número 9, corretor A nota 2 e B nota 3.

- Corretor A: Nota 2 - iniciante, ela redige com expressões ou frases simples, apesar destas frases atingirem o objetivo que é o tópico, o vocabulário é muito limitado e muita dificuldade para o leitor entender a escrita, o “handwriting”.
- Corretor B: Uma anotação que eu tenho aqui é que realmente poderia ser uma nota menor, eu atribui a nota 3. A redação é confusa, o vocabulário inapropriado, não apenas por usar palavras extremamente informais, que não seria o caso aqui. Apresenta alguns problemas com tempos verbais, então uma anotação que eu já tinha é que poderia ser uma nota menor, então eu poderia baixar esta nota para 2.
- Mediador: A nota de denominador comum seria 2. Composição número 10, corretor A nota 4 e B nota 6.
- Corretor A: Nota 4, em expansão, ele organiza as idéias em uma ordem lógica, ele experimenta fazer uso de tempos verbais diferentes, apesar de que algumas das formas não são apropriadas, o que me chamou atenção é o uso do vocabulário de transição, que atraíram o leitor, mas está em expansão, ele pode ser fluente, mas ainda está em expansão.
- Corretor B: A nota atribuída é 6, mantenho e acredito que o texto é bem organizado e confesso que uma das coisas que ajuda muito é a questão da apresentação, por incrível que pareça, mas se você observar a redação anterior e essa, a apresentação ajuda na organização e atrai um pouco mais o leitor, favorece e o leitor simpatiza mais com a redação. Apresenta organização dos parágrafos, o vocabulário é mais elaborado, apresenta uso de vocabulário de transição e descreve de maneira realmente boa, então atribuo a nota 6.
- Mediador: Composição número 11, corretor A nota 2 e B nota 3.
- Corretor A: Nota 2, iniciante, tive dificuldade na leitura, as palavras não estão legíveis. Ele tem aqui frases simples de apoio ao tópico, o vocabulário é muito limitado para o nível, ele poderia desenvolver o tópico com mais riqueza de detalhes, exemplos para atrair o leitor.

- Corretor B: Ele realmente apresenta problemas de estrutura, de organização, a linguagem não é tão elaborada e apresenta uma influência evidente da língua portuguesa, L1 dele, considerando esses fatores a nota poderia ser 2.
- Mediador: A nota atribuída pelos 2 corretores é 2. Composição 12, corretor A 4 e B 6.
- Corretor A: Ele organiza as idéias numa ordem lógica, tem aqui vocabulário adequado para o propósito do texto, faz erros que não interferem no significado, mas não atrai o leitor com os exemplos colocados.
- Corretor B: Eu mantenho a nota 5, mas tenho uma ressalva porque ele poderia ter feito uso de vocabulário de transição, pois ao mudar de assunto o texto apresenta quebras.
- Mediador: Composição número 13, corretor A - 4 e corretor B - 5.
- Corretor A: Ele organiza as idéias e apresenta vocabulário apropriado para desenvolver o tópico. Dá alguns exemplos para ilustrar melhor esse tópico, ele está em expansão e poderia até ser fluente pela questão da conclusão que atraiu muito, que me chamou muito a atenção. Nota 5.
- Corretor B: Eu mantenho a nota 5, mas tenho uma ressalva porque ele poderia ter feito uso de vocabulário de transição, pois ao mudar de assunto o texto apresenta quebras.
- Mediador: A nota do corretor A é 5 e do B 5 também, composição número 14, corretor A 2 e corretor B 4.
- Corretor A: A primeira nota que eu dei foi 2 – iniciante. Uma das questões aqui é que ele não dá uma conclusão para o parágrafo, na segunda leitura, eu pensei na nota 3, pelo fato de ele apresentar um bom vocabulário para ilustrar o tópico mencionado. Usa vocabulário de transição, apesar de limitado, faz alguns erros de gramática, que às vezes até podem prejudicar, mas atraiu o leitor, então seria nota 3, apesar de não apresentar conclusão no parágrafo.
- Corretor B: Um limitador dessa redação foi a apresentação, estava difícil de ler.

Acredito que apresentou uma boa conclusão e é em expansão, organiza as idéias, fala do irmão, exemplifica, apesar de que o maior problema foi a apresentação do texto em si.

Mediador: Composição número 15, corretor A-3 e corretor B-4.

Corretor A: Primeira leitura nota 3, segunda leitura nota 4, pelo fato de apresentar vocabulário de transição adequado. Ele dá vários exemplos para ilustrar o tópico, tem erros de ortografia, mas que não interferem, então é 4, em expansão.

Corretor B: Ele faz uso de bom vocabulário, apresenta alguns erros, mas faz uso de boas estruturas, que era difícil não dar a ele uma nota 6, mas devido aos erros mantenho a nota 5.

Mediador: Composição número 16, corretor A nota 3 e B nota 5.

Corretor A: Quando eu li esta composição eu imaginei a pessoa, por ela ser bem carinhosa com a esposa, pela questão da escrita mesmo, apesar da linguagem ser muito simplificada, com exemplos muito simples. Ela atraiu o leitor pela questão da sutileza e carinho pela pessoa.

Corretor B: Ele faz bom uso do vocabulário, consegue fazer uso do humor, apresenta coerência na organização dos fatos que aparecem. Poderia e tem condições de melhorar esse texto, mas ele é fluente.

Mediador: Composição número 17, corretor A-3 e corretor B-5.

Corretor A: Ele organiza a composição, faz uso de vocabulário adequado, apesar de que as frases não tem pontuação, frases longas (run-on) e cansa o leitor.

Corretor B: Acredito que o redator é coerente, pois apresenta pontos positivos e negativos, então é fiel à “topic sentence”. Faz bom uso de vocabulário, mas usa elementos informais como contrações, não apresenta vocabulário de transição. Acredito que a nota pode ser 4.

Mediador: Então corretor A nota 3 e B nota 4. Composição número 18, corretor A nota-2 e B nota-3.

- Corretor A: Foi difícil a leitura por não haver uma seqüência lógica, um dos exemplos foge do tópico, a linguagem é um tanto simplificada para o nível do aluno.
- Corretor B: Então vou relatar que aqui foi o momento em que eu retornei a correção, há vários momentos em que eu retomo. Ele apresenta vocabulário informal, o texto é confuso, com palavras não apropriadas e apresenta quebras sem transição. Revendo acredito que possa passar para 2.
- Mediador: Nota corretor B-2 e corretor A-2 também. Composição 19, corretor A-2 e B-4.
- Corretor A: Faz uso de linguagem informal, está sempre indagando para ilustrar o tópico, muito simples e no final foge do tópico.
- Corretor B: O que levou a nota 4 é a questão do desenvolvimento das idéias, de como ela exemplifica . Faz uso de conectores indicando ordem, usa exemplos para ilustrar, as idéias são simples e o vocabulário também, então a nota pode ser 3.
- Mediador: Nota corretor B-3 e corretor A-2. Composição n.º 20, corretor A-2 e B-3.
- Corretor A: O que atraiu na leitura dessa composição é que os exemplos são muitos comuns, então o leitor tem que estar atraído pelo que o compositor está redigindo, no caso ele é iniciante. Usa linguagem muito simplificada, as ilustrações, os exemplos são muito comuns, apresenta problemas de linguagem e algumas partes são incompreensíveis.
- Corretor B: Ele descreve o irmão, exemplifica bem com exemplos simples, com vocabulário pouco elaborado. Ele poderia usar exemplos simples, mas com vocabulário atraente. Existe uma estrutura deficiente, mas que pode ser melhorada.
- Mediador: Composição 21, corretor A-3 e B-6.
- Corretor A: Ele redigi frases sobre uma idéia, com repetição da mesma palavra

várias vezes. Os exemplos que ilustram o tópico são poucos, então nota 3.

Corretor B: Eu considerei o vocabulário, pois embora ele faça uso de repetição, procura usar palavras não tão frequentes. A estrutura tem problemas gramaticais e faz uso de transição, procura exemplificar e é consistente na caracterização da sua mãe. Ela atinge o objetivo de comunicar quem é essa mãe e como ela é. Revendo ela poderia ser 5, porque poderia apresentar elaboração maior nas sentenças.

Mediador: 21 corretor A-3, corretor B-5, composição número 22, corretor A-2 e B- nota 3.

Corretor A: Iniciante, por não atrair o leitor, por haver vocabulário limitado, redige frases sobre uma idéia, mas é 2.

Corretor B: Analisei questões como o aluno foge do assunto, no lugar de descrever uma pessoa ele descreve a rotina do apartamento e não as pessoas. Nota 3 porque com feedback, esse aluno poderia rever isso e mudar. Bom uso de vocabulário. mas considerando fuga do tema e falta de coerência, ele pode ser iniciante, nota 2.

Mediador: Corretor B-nota 2 e A-nota 2. Composição 23, A nota 2 e B nota 4.

Corretor A: Iniciante na primeira leitura , pois no início da composição fala sobre o irmão, depois muda para o pai e então usa o pronome "he", sem saber na leitura até a 2ª frase no final, uma questão. Na 2ª questão, ele menciona frases redundantes para o leitor. Na segunda leitura eu achei que ele poderia ser nota 3, pois usa bons exemplos para ilustrar o tópico, apesar dos problemas.

Corretor B: Nota 4, apesar de apresentar falta de vocabulário de transição e utilizar vocabulário diferente do usual. A meu ver apresenta coerência, pois descreve o irmão bem, mas falta a questão da transição. Atribuo a nota 3.

Mediador: Corretor A nota 3 e B nota 3. Composição nº 24, corretor A-3 e B-6.

Corretor A: A primeira leitura foi 3 pelo fato dela usar linguagem bem

simplificada. Na segunda leitura chamou a questão o tópico, ela coloca que a melhor amiga dela é admirável e ela exemplificou com 7 exemplos para justificar isso e detalhou mais. Então eu mudo de 3 para 4, pelo fato dela ilustrar todos os exemplos.

Corretor B: Ela utiliza linguagem simples, inicialmente atribui a nota 6, mas acredito que a nota deveria ser 5, pois ela deveria rebuscar um pouco mais sua linguagem. Não apresenta problemas gramaticais pela simplicidade da linguagem.

Mediador: Corretor A nota 4 e B nota 5. Composição 25, corretor A-3 e B-4.

Corretor A: Aqui ele redigi frases sobre uma idéia, no final da composição não sei se a pessoa do início é a mesma do final do texto. Gostei muito da conclusão, mas precisaria rebuscar um pouco mais o vocabulário para atrair o leitor.

Corretor B: Apresenta coerência, pois fala do lado negativo da pessoa analisada e no uso final faz uma boa conclusão, referindo-se a essa pessoa. Não faz uso de palavras de transição e deixa algumas questões sem explicação. Inicialmente considerei como um escritor em expansão, mas acredito que é em desenvolvimento. Sendo, assim nota 3.

Mediador: Corretor A, nota 3 e B nota 3. Composição número 26, corretor A- nota 6 e B-6.

Corretor A: Teve em mente o leitor, que se interessou em ler a composição. Gramática e vocabulário com estilo rebuscado, conclusão muito boa em relação ao tema, proficiente, nota 6.

Corretor B: Ele faz ótimo uso de vocabulário, estrutura, existe coerência, existe transição e realmente a maneira como ele desenvolve atrai o leitor.

Mediador: Composição 27, corretor A-2 e corretor B-3.

Corretor A: Tive dificuldade na leitura pela questão do “handwriting”, apesar dele se ater ao tema, tem problemas de linguagem que podem interferir, no desenvolvimento da composição. Faz uso de português, os tempos

verbais não são adequados, coloca exemplos sem ilustração.

Corretor B: Inicialmente nota 3, de fato nas minhas anotações vejo que ele apresenta problemas de vocabulário, conclusão fraca, faltam palavras de transição, em determinados momentos ele explica o texto por não poder escrever muito sobre o assunto. Revendo acredito que ele seja 2.

Mediador: Corretor A-2 e B-2. Composição número 28, corretor A-3 e B-4

Corretor A: O que me chamou atenção foi a questão dele estar julgando de uma forma "mesquinha" o colega, isso realmente interferiu na nota dada. Além disso, apresenta problema de linguagem, mas teve boa organização, o que realmente interferiu foi o problema de julgamento, eu poderia até considerá-lo em expansão, então fica com nota 3, mesmo assim.

Corretor B: Apresenta mais do que um parágrafo, se esforça para desenvolver e apresentar pontos negativos do amigo. Tem problemas de vocabulário e transição. Não me envolvi tanto com a questão do julgamento, mas me ative mais as idéias, atribui a nota 4, mas ele pode melhorar para 5.

Mediador: O que vocês consideraram como aspectos e fatores importantes na sua correção?

Corretor A: Fazemos uso do critério holístico, então levo em consideração a questão da organização, do vocabulário, da gramática, ortografia, pontuação, a mecânica, mas todos os itens tem que estar em harmonia para realmente fazer com que o leitor se interesse pela composição. Então um fator que eu considero muito importante é que, ao redigir o escritor tenha em mente o leitor e tente atraí-lo para se interessar pela composição através de exemplos não muito comuns.

Mediador: Você acha que dentro dos critérios você conseguiu identificar aspectos que apoiam essa sua visão de redação?

Corretor A: Por ser um critério holístico com todos os itens juntos, mas tem que ter harmonia. Se o que eu considero está dentro do critério? Acredito

que sim, porque se ele usa vocabulário e gramática adequados, vai conquistar o leitor. Acho que sim, que todos os itens do critério holístico compõe minhas idéias.

Corretor B: Eu acredito que a apresentação é um fator importante, redações borradas ou pouco legíveis atrapalham a leitura. Uma boa divisão de parágrafos e espaçamento é um atrativo inicial. Por se tratar de uma escrita em língua estrangeira exposta à um modelo com “Topic Sentence, Supporting Sentences” e Conclusão, exijo a estrutura e coerência. É necessário o uso de “transitional words”. Os elementos da narração são importantes e não me preocupa a simplicidade, pois às vezes não é possível rebuscar muito, mas a linguagem pode ser um bom molho para um arroz e feijão.

Outra questão que influenciou muito é que se trata de uma única correção, o aluno não vai ter feedback, não vai reescrever e esse fator me constrangeu de dar notas inferiores. Na minha correção inicial, considerando a escrita única, atribui notas maiores, muito embora minhas anotações demonstrem que eu percebi os problemas. Talvez seja uma questão de reavaliar o critério proposto. Numa segunda leitura percebi, as anotações e que talvez tenha sido um tanto complacente com os alunos e percebi a necessidade de modificar algumas notas. Algumas eu mantive, pois os alunos atingiram o objetivo, que é de se comunicar.

Os critérios refletem os apontamentos que fiz, mas acho que não estive tão em sintonia com o critério por ser holístico, porque em alguns aspectos a composição pode ser 6 ou 5 e eu não encontrei a harmonia dita pelo corretor A. Achar o equilíbrio entre as notas é a questão mais difícil, pois os alunos não se enquadram genuinamente, ou especificamente, em apenas um dos níveis.

ANEXO 8 – OFÍCIO 323/04/CEP-PUCPR – AUTORIZAÇÃO DO COMITÊ DE
ÉTICA EM PESQUISA



Pontifícia Universidade Católica do Paraná
Pró-Reitoria de Pesquisa e Pós-Graduação

Curitiba, 03 de setembro de 2004.

Of. 323/04/CEP-PUCPR

**Ref. “A busca pela minimização dos efeitos da subjetividade na
correção da escrita em língua estrangeira”.**

Prezado (a) Pesquisador (es),

Venho por meio desta informar a Vossa Senhoria que o Comitê de Ética em Pesquisa da PUCPR, no dia 01 de setembro do corrente ano aprovou o Projeto Intitulado **“A busca pela minimização dos efeitos da subjetividade na correção da escrita em língua estrangeira”**, pertencente ao Grupo III, sob o registro no CEP nº 397 e será encaminhado a CONEP para o devido cadastro. Lembro ao senhor (a) pesquisador (a) que é obrigatório encaminhar relatório anual parcial e relatório final a este CEP.

Atenciosamente,

A handwritten signature in black ink, appearing to read 'Emílio', is written over a faint horizontal line.

Prof. Dr. Emílio José Scheer Neto.

Coordenador do Comitê de Ética em Pesquisa

Ilma Sra

Ane Cibele Palma