

AVALIAÇÃO DA APRENDIZAGEM DE ILE: UMA REALIDADE QUE CHOCA

LAURA MICCOLI
UFMG

RESUMO

A importância da avaliação e os problemas frequentemente enfrentados por professores em relação à avaliação em sala de aula são a motivação para este artigo. Seu objetivo é chamar a atenção para uma abordagem para o desenvolvimento de instrumentos de avaliação informada pelos resultados de um estudo de caso que analisou 28 testes. A literatura básica que guia a construção de testes válidos e confiáveis é revista para apoiar o argumento de que a avaliação em sala de aula pode ser abordada a partir de uma fundamentação teórica sólida. Este estudo de caso será apresentado assim como sua metodologia para a análise dos dados. Os resultados revelam que, embora uma abordagem comunicativa tenha sido uma das abordagens mais reconhecidas para o ensino de uma segunda língua durante os últimos 25 anos, no que refere à avaliação, a abordagem que ainda parece predominar para a construção de testes na região de Minas Gerais nos lembra do período conhecido como gramática-tradução. Esses resultados perturbadores são apresentados e discutidos neste trabalho, assim como suas implicações para professores e pesquisadores.

Palavras-chave: avaliação em sala de aula; textos; inglês como LE.

ABSTRACT

The importance of evaluation and the problems most frequently faced by teachers in relation to classroom assessment are the motivation for this paper. The objective is to make a case for a more informed approach to the development of classroom assessment instruments from the results of a case study that analyzed 28 classroom tests. The basic literature that guides the construction of valid and reliable instruments is reviewed to support the argument that classroom assessment can be approached from a sound theoretical framework. The case study research will be presented as well as its methodology for data analysis. Results reveal that although the communicative approach has been one of the most recognized approaches for the teaching of a second language in the last 25 years, as far as evaluation is concerned, the approach that seems to predominate in classroom test construction in a region of Minas Gerais reminds us of the period known as grammar-translation. Such disturbing results are presented and discussed in this paper as well as its implications for teachers and researchers.

Key-words: classroom assessment; texts; EFL.

INTRODUÇÃO

No Brasil, a importância da avaliação na educação está em evidência desde 1990. Iniciativas como SAEB – Sistema Nacional de Avaliação da Educação Básica: ENEM – Exame Nacional do Ensino Médio e o Provão – Exame Nacional de Cursos, revelam a política do Ministério da Educação para a implementação de um sistema contínuo de avaliação com propósitos diagnósticos e de computação de dados. O objetivo subjacente

ao sistema avaliativo educacional, em todos os níveis, é de aperfeiçoar a qualidade da educação oferecida no Brasil. Aperfeiçoar a qualidade das avaliações utilizadas em sala de aula constitui uma das motivações deste trabalho.

O Inglês como Língua Estrangeira (ILE) é uma área não avaliada pelos exames acima mencionados. Segundo o INEP - Instituto Nacional de Estudos Pedagógicos, a LDB faculta às instituições a escolha da língua estrangeira a ser incluída nos currículos de níveis fundamental e médio, justificando, assim, a não elaboração de provas de línguas estrangeiras em seus exames¹. Entretanto, uma vez que em documentos da *home page* do INEP encontram-se afirmações de que o desempenho dos aprendizes reflete a qualidade do ensino a que foram expostos, é curioso que avaliações de língua estrangeira não estejam incluídas nas baterias de testes que o instituto produz. Afinal, poderiam ser produzidas avaliações nas diferentes línguas estrangeiras para incluir nos exames, cabendo ao aluno a seleção da(s) língua(s) que a instituição adota em seu currículo.

No entanto, isso está fora do escopo deste trabalho. O que se coloca em questão aqui é que existe, de acordo com o INEP, uma correlação positiva entre o desempenho de aprendizes em instrumentos de avaliação e o ensino de boa qualidade. Todavia, essa correlação positiva depende de uma outra variável, não mencionada por Locatelly (2001) – a qualidade do instrumento de avaliação utilizado para medir o desempenho dos aprendizes. Se esse instrumento é bem construído, refletindo além de um construto válido, diretrizes básicas para a elaboração de avaliações, o mesmo irá gerar resultados confiáveis. Se este não for o caso, o instrumento de avaliação pode interferir no desempenho do aprendiz no teste a que se submete.

Temos evidência de literatura nacional abordando o tema avaliação no Brasil (Luckesi 1997, Sarmento 1997, Scaramucci 1997). Quanto à questão da avaliação da aprendizagem de ILE, há estudos que a documentam. Dentre vários destacamos Porto (2003), Neves (2002) e Lopes (2002). A pesquisa de Porto (2003) investigou as percepções de três professoras universitárias sobre avaliação de aprendizagem de ILE, evidenciando posturas tradicional ou transicional em relação à maneira como é realizada a avaliação da aprendizagem. Concluiu ainda que essas professoras, que formam professores de línguas, têm dificuldade em explicitar seus critérios e em elaborar provas. Neves (2002) investigou o discurso de professores universitários sobre avaliação, evidenciando inconsistências e posturas que reforçam questões de subjetividade e poder no exercício da avaliação. Lopes (2002) analisou testes comunicativos para evidenciar que, apesar dessa qualidade, os mesmos mantêm foco em questões mais voltadas para o conhecimento de língua do que para o uso do ILE. Apesar da qualidade de essas pesquisas, elas são incipientes frente à extensão territorial e aos desafios da educação no Brasil. Mesmo assim, nos permitem ter apenas uma informação limitada dos aspectos envolvidos no exercício da avaliação da aprendizagem de ILE. Dessa forma, é necessário um maior esforço na produção de pesquisas que tenham como foco a avaliação da aprendizagem de ILE e como são produzidas por professores.

¹ Essa foi a resposta dada a um e.mail enviado ao INEP por Laura Miccoli em 2002 que indagava do porquê da não inclusão da língua inglesa no ENEM.

Além disso, há pouca literatura que oriente os professores que desejem desenvolver boas avaliações e sanar um problema recorrente - os testes nem sempre revelam os resultados esperados (Hughes 1989, Brown 2001). Por esse motivo, dois problemas afetam professores: ou os resultados nem sempre refletem o verdadeiro desempenho do aprendiz ou não representam nada mais que apenas valores numéricos que nem mesmo os professores conseguem interpretar. Tais questões levam os professores a não confiar nos resultados que esses testes revelam.

Para a maioria dos professores de ILE, a elaboração de testes é uma de suas responsabilidades, juntamente com a de ensinar. Isso significa que a maioria dos professores elabora os testes que aplicam. No entanto, há professores que não o fazem. Em algumas escolas são os supervisores. Em institutos de idiomas particulares, os testes geralmente são preparados por uma escola central que os distribui às suas filiais. Independentemente de quem elabora instrumentos de avaliação, persiste o problema dos professores não confiarem nos seus resultados, mesmo quando são eles os elaboradores. A conclusão lógica, então, é que professores não confiam em si próprios enquanto elaboradores de testes. Na verdade, aqueles que têm participado de cursos de avaliação² confirmam essa e outra hipótese – professores não gostam de testes. Não gostam de elaborá-los, não gostam de aplicá-los e não gostam de corrigi-los, porque acham que, na maioria das vezes, os resultados dos testes não correspondem ao desempenho dos alunos em sala de aula.

Em vista das questões acima, a motivação para investigação que relatamos neste trabalho é: (1) a crença de que avaliação e ensino são partes do processo de aprendizagem; (2) a crença de que a avaliação pode melhorar a qualidade do ensino (Pellegrini 2003); (3) a crença de que o bom desempenho em um teste não está relacionado somente com a qualidade do ensino, mas também com a qualidade dos instrumentos de avaliação utilizados para avaliar o desempenho dos aprendizes (Hughes 1989, Brown and Hudson 2002); (4) a evidência de que os professores não sentem confiança nos testes que aplicam (Porto 2003) e, (5) a avaliação é uma área de pesquisa produtiva em Linguística Aplicada na qual encontramos princípios para o desenvolvimento de bons instrumentos de avaliação da aprendizagem de uma língua estrangeira.

O ESTUDO DE CASO

A investigação foi realizada ao longo do primeiro semestre de 1999, a partir da coleta de avaliações elaboradas por seis alunos em processo de finalização de seu curso de graduação em uma das universidades do estado de Minas Gerais. Voluntariamente, eles atenderam à solicitação da professora e pesquisadora (Santos 2001), que informou a eles que o seu objetivo era investigar a questão da avaliação em sala de aula. Foram analisados os testes elaborados por eles mesmos. Cinco deles já trabalhavam em escolas públicas no estado de Minas Gerais e todos possuíam até cinco anos de experiência no ensino. Portanto,

² Tenho sido responsável por cursos de avaliação de aprendizagem de língua estrangeira nos níveis da graduação e pós-graduação na UFMG e em outras instituições de ensino superior.

não eram novatos no ensino de ILE. A pesquisa visou responder às seguintes questões: (1) o que as avaliações revelam sobre a maneira pela qual esses professores percebem o processo de avaliação e ensino? (2) Quais são suas preferências sobre o que testar e como testar?

Foi adotada uma abordagem qualitativa com a metodologia de estudo de caso para a coleta de documentos (Jonhson 1992, Santos *ibid*) – os testes elaborados e aplicados pelos professores participantes. Cada um deles forneceu quatro testes. Portanto, os dados foram gerados a partir de 24 testes com 138 questões. Estas foram analisadas de acordo com a literatura da área de avaliação em ILE, que será revisada no Marco Teórico.

Os procedimentos para a análise desses testes incluíram os seguintes passos: (1) identificação dos tipos de formatos de questões; (2) agrupamento de questões de acordo com o formato em comum; (3) categorização desse agrupamento; (4) levantamento, agrupamento das questões quanto ao seu conteúdo – gramática, vocabulário ou mista; (5) levantamento de problemas relativos às instruções e à redação das questões; e (6) levantamento de problemas relativos à aparência dos testes, ordenação espacial, letra cursiva e rasuras. Neste artigo nos restringimos aos resultados em relação aos formatos e aos conteúdos dos testes analisados. Na próxima seção, apresentamos o marco teórico que orientou a análise dos dados coletados.

MARCO TEÓRICO

A questão da avaliação da aprendizagem é uma área profícua e polêmica na área de Linguística Aplicada. Profícua por haver uma quantidade de pesquisas internacionais e nacionais que atestam a sua pertinência no processo de ensino e aprendizagem de ILE e polêmica por haver diferentes perspectivas para sua implementação – uma abordagem positivista e uma abordagem crítica (Pennicook 2004). A abordagem positivista vê o processo de avaliação como resultado de pesquisas e conhecimentos que asseguram a validade de instrumentos objetivos de avaliação, tais como as questões de múltipla escolha. A abordagem crítica à avaliação (Shohamy 2001) questiona o uso de testes como a única forma de se aferir o conhecimento, além de discutir a importância não só da responsabilidade daqueles que elaboram os instrumentos de avaliação, mas também pelos usos que são feitos desses instrumentos. Dessa forma, rejeita a noção de objetividade e neutralidade da abordagem positivista.

Há de se reconhecer que a adoção de qualquer uma dessas duas abordagens decorre de decisões que se inserem no campo das políticas de ensino e de avaliação. As conseqüências de uma ou outra postura atendem a interesses específicos que podem manter ou mudar o status quo. Dessa forma, como a abordagem crítica à avaliação decorre de uma discussão dos pressupostos da avaliação positivista, a precedência da abordagem positivista é inquestionável. Portanto, seu conhecimento é básico para aquele que pretende vir a elaborar uma avaliação. Como, no estudo de caso que aqui apresentamos, lidávamos com professores em formação, adotamos a abordagem positivista como marco teórico para o estudo, mas sabemos que a questão da avaliação pode vir a suscitar discussões muito mais amplas.

No paradigma positivista, os termos avaliação e teste são, às vezes, tomados um pelo outro (Brown 2001). No entanto, há distinção entre ambos. A maioria dos autores, inclusive Brown, usa avaliação como um termo mais genérico que teste. Avaliação, então, refere-se a qualquer uma das muitas maneiras utilizadas para coletar informações relacionadas com a aprendizagem. O teste, por sua vez, é uma dessas maneiras.

De acordo com Genesse e Upshur (1996), existem avaliações com testes e sem testes. Uma redação e uma entrevista são, na verdade, exemplos de diferentes instrumentos para avaliar habilidade escrita e oral. A maioria dos professores que utiliza esses dois tipos de testes avalia de forma tradicional. Procedimentos de avaliação diferentes do tradicional tais como as entrevistas não são comuns e são conhecidos como avaliações alternativas. Exemplos de instrumentos alternativos de avaliação são os portfólios e as conferências. Um portfólio é uma coleção das melhores produções dos aprendizes, selecionadas por eles mesmos. A conversa é uma oportunidade de interação entre professor e aluno, que é guiada por perguntas preparadas pelo professor ou com a colaboração do aluno. Ela pode usada para avaliar o portfólio ou qualquer aspecto do processo de aprendizagem. Nessa forma de avaliação, perguntas cuidadosamente desenvolvidas levam o aprendiz a refletir sobre a qualidade do trabalho que tem feito em sala de aula.

Em se tratando de testes tradicionais, a elaboração de um instrumento envolve decisões sobre as diferentes maneiras de se avaliar determinadas habilidades. Nessa fase do processo, o elaborador tem que estar ciente das diferenças na terminologia e nas abordagens para a elaboração de questões que serão incluídas em seu instrumento de avaliação.

A primeira decisão que o elaborador do teste enfrenta é a escolha de uma abordagem direta ou indireta para testar determinada competência ou habilidade. Uma abordagem direta na avaliação demanda do candidato o desempenho da habilidade que está sendo avaliada. Logo, uma redação é uma abordagem direta para avaliar a habilidade de escrita de um candidato. Se o objetivo é avaliar a habilidade oral do candidato, uma entrevista é apropriada. Uma abordagem direta em avaliação é possível em relação às habilidades de produção como a fala e a escrita, em que o teste permite uma observação real do desempenho do aprendiz³. Em contraposição, as abordagens para avaliar a compreensão oral e leitura não acessam o desempenho do candidato diretamente, ou seja, o elaborador do teste não tem como desenvolver um instrumento de avaliação que lhe permita observar o processamento do material escrito ou falado. É necessário elaborar um teste que garanta a demonstração de que o aprendiz executou a tarefa de leitura ou de compreensão com sucesso. Isso só pode ser feito indiretamente, através de testes de compreensão oral e de leitura, que enfocam no resultado do processo de compreensão oral ou de leitura. Então, a avaliação indireta avalia as competências que subjazem às habilidades que estão sendo testadas (Hughes 1989). Um exemplo típico de um teste indireto para avaliar escrita é aquele que pede ao candidato que identifique os erros em uma sentença.

A segunda decisão que o elaborador de testes terá que tomar será entre testar com itens isolados ou integrativos (Weir 1990) ou dentro desse contínuo. Uma questão de item

³ O real, nesse caso, é relativo, pois nem sempre o desempenho de um aluno em um teste oral ou escrito reproduzirá o seu desempenho em uma situação de não teste.

isolado avalia cada elemento da língua de uma vez. Em contraposição, para responder uma questão integrativa, o candidato não enfoca apenas um aspecto da língua de cada vez. Ele tem que por em uso outras habilidades para completar a questão. Uma questão integrativa, então, requer do aluno ou candidato o uso de mais de uma competência para chegar à resposta correta.

Juntamente com as decisões acima mencionadas, é importante que se faça distinção entre tarefa, questão e item. O termo tarefa está relacionado com a habilidade que está sendo testada, por exemplo, compreensão oral, leitura, fala, escrita, vocabulário e gramática⁴. O termo questão refere-se ao formato em que uma tarefa específica será apresentada. Finalmente, o termo item refere-se aos componentes da questão. Os exemplos abaixo ilustram essas diferenças.

1. She quickly _____ her lunch.
 - a. drank
 - b. ate
 - c. drove
 - d. slept

(Madsen 1983:16)

2. Read the text below and complete it with either “do” or “make”.

John didn't know what to _____ last night when he his brother told him that he would have to _____ the arrangements for his mother's vacation trip. He wanted to have said that was something his brother _____ all the time. Every time there was something to take care of, he managed to go away on a business trip. This _____ him mad, but he decided once more to keep quiet. So, he _____ say anything.

No primeiro exemplo, a tarefa que se requer do aprendiz é a de decidir sobre a palavra apropriada. Assim, essa questão testa o conhecimento de vocabulário. A tarefa é apresentada em um formato de múltipla escolha. Uma vez que neste texto há apenas esta questão para a tarefa acima, ela contém apenas um item. Essa questão de múltipla escolha é um exemplo típico de uma questão de item isolado.

No segundo exemplo, temos uma questão em formato de resposta limitada. Essa questão tem cinco itens e a tarefa requerida do aprendiz é a de ler e de completar o texto com o item lexical apropriado (vocabulário). Esse exemplo ilustra uma questão mais integrativa, pois exige do aluno mais que o reconhecimento da resposta correta, no caso, o processamento da informação do texto para o preenchimento correto das lacunas.

⁴ Embora vocabulário e gramática sejam habilidades que subjazem ao desempenho das habilidades de falar, escrever, ouvir e ler, elas estão sendo consideradas como habilidades separadas visto que elas podem ser testadas separadamente (Hughes 1989). De fato, testes famosos como o TOEFL e outros exames de proficiência dedicam seções à avaliação dessas habilidades.

Depois de decidir sobre as tarefas, o elaborador de testes tem que tomar decisões sobre o formato de tais questões, uma vez que uma tarefa pode ser testada de diferentes maneiras. Existem basicamente três formatos principais: questões fechadas, resposta limitada e questões abertas (Genesse e Upshur, 1996). Questões fechadas são aquelas que apresentam uma sentença ou texto que deverá ser completado com base nas alternativas fornecidas pelo autor. A tarefa do candidato é basicamente reconhecer a alternativa correta. Esse tipo de tarefa “não envolve produção lingüística e, portanto, não pode ser usada para avaliar produção oral ou escrita diretamente” (Genesse e Upshur, 1996:168). Logo, elas são mais adequadas para a avaliação de habilidades de compreensão. Os formatos típicos associados com questões fechadas são: múltipla-escolha, cloze com alternativas, questões de elicitación (uso de gravuras ou objetos) e questões de julgamento (falso ou verdadeiro).

Questões de resposta limitada referem-se a questões que não fornecem alternativas. Nesse tipo de questão, é tarefa do candidato produzir a resposta correta ou apropriada sem a ajuda das alternativas. Esse tipo de questão requer mais que apenas o reconhecimento da resposta correta. Na verdade, essas questões requerem que o candidato produza ou aplique o conhecimento que tem (Genesse e Upshur, 1996). Os típicos formatos associados com resposta limitada são: elicitación e cloze sem alternativas, ditado ou imitação, transformação (do presente para o passado), organização de sentenças ou diálogos (colocar palavras ou sentenças em uma ordem que faça sentido) e *matching* (ligar uma sentença ou palavra etc., de uma coluna a de outra coluna). A avaliação com questões de resposta limitada é mais subjetiva que a de questões fechadas, pois envolve o julgamento individual de quem corrige.

Nas questões abertas, as alternativas não são limitadas (Genesse e Upshur, 1996). O aluno ou candidato tem liberdade para prover a resposta de várias maneiras, o que significa que o autor do teste não tem como saber com antecedência o que o candidato irá produzir. Eles são especialmente adequados para avaliar quaisquer habilidades lingüísticas que requerem do aprendiz a produção lingüística. Exemplos típicos de questões abertas são: entrevistas orais, encenações, redações, tarefas de completar, relatórios orais ou escritos. Essas são tarefas bastante subjetivas e cabe ao examinador fazer julgamento sobre o tipo de produção gerado pelo candidato.

As decisões acima devem ser tomadas dentro dos parâmetros de validade e confiabilidade. Tais critérios têm importância fundamental na avaliação. Não é necessário que professores façam análises estatísticas para determinar a validade e a confiabilidade de seus testes, no entanto, eles devem ter consciência de que esses critérios são importantes para que os instrumentos de avaliação que elaboram funcionem bem.

A validade de um instrumento de avaliação é a extensão pela qual o teste reflete a característica que se quer avaliar (Genesse e Upshur, 1996). Então, dizemos que um teste é válido se ele testa o que queremos testar e não outra coisa diferente. Há referência a três tipos diferentes de validade na literatura (Hughes 1989, Genesse e Upshur *ibid*). Primeiro, há a validade de conteúdo, que se refere ao conteúdo de um instrumento de avaliação. Se o conteúdo do teste está de acordo com o conteúdo das especificações, então pode-se afirmar que o teste tem validade de conteúdo. Validade de construto refere-se ao construto que o teste reflete. Por exemplo, para testar produção oral, precisamos de uma boa definição

ou construto sobre o que é produção oral. O teste tem validade de construto quando espelha os atributos do construto. Finalmente, há a validade de face, que se refere à aparência de um teste (Weir, 1990). Dizemos que um teste possui validade de face quando ele apresenta uma combinação de atributos de um teste bem projetado.

Confiabilidade tem a ver com a consistência ou estabilidade dos resultados apresentados pelo instrumento de avaliação (Genesse e Upshur 1996). Na verdade, tem a ver com a confiança que depositamos nos resultados de um teste. Então, podemos dizer que um instrumento de avaliação é confiável à medida que seus resultados são considerados estáveis, ou, colocando de outra forma, não sujeitos a variações. Na literatura são encontradas referências a três tipos de confiabilidade (Hughes, 1989; Genesse e Upshur *ibid*). A primeira, a confiabilidade de corretor, refere-se aos procedimentos seguidos para (1) garantir que a pessoa encarregada de coletar a informação saiba o que quer testar e (2) seja treinada para a correção. A segunda, a confiabilidade de objeto, diz respeito aos procedimentos conduzidos para minimizar os efeitos de que um teste mascare as verdadeiras habilidades de quem o responde. Por exemplo, cuidar para que a qualidade de uma fita ou de um local para a realização de uma prova que avalie a escuta e a fala garante a confiabilidade de objeto daquela avaliação. O último tipo de confiabilidade é a confiabilidade de instrumento. Um instrumento é mais confiável se tiver sido testado antes de sua aplicação ou se for verificado seu potencial de otimização pelo uso de diferentes tipos de questões bem como diferentes abordagens para a coleta de informação.

RESULTADOS

Nesta seção, apresentamos os resultados obtidos na análise dos aos formatos e dos conteúdos das 138 questões encontradas nos 24 testes elaborados pelos seis professores em formação que tomaram parte neste estudo de caso.

No que diz respeito ao formato, a análise de questões revelou uma preferência por duas categorias de formatos: Resposta Limitada e Questão Fechada. A categoria Resposta Limitada tem cinco subcategorias: *complete sem alternativas; transforme; escreva; traduza e ordene*. A categoria Questões Fechadas tem duas subcategorias: *múltipla escolha e complete com alternativas*.

Os dados analisados revelaram uma preferência por três tipos de conteúdo, os quais foram categorizados como *Gramática, Mista e Vocabulário*. Na categoria *Gramática* encontramos questões cujo conteúdo enfoca estruturas gramaticais e palavras funcionais. Na categoria *Mista* encontramos questões cujo conteúdo se atem a aspectos gramaticais, aspectos de uso da língua e vocabulário. Na categoria *Vocabulário*, encontram-se as questões que avaliam conhecimento de vocabulário. Na Tabela 1, a seguir, são apresentados os dados referentes à análise de formato e conteúdo. Estão incluídos o número de questões analisadas e a porcentagem de questões em cada categoria e subcategoria.

Tabela 1: O Formato e o Conteúdo das Questões

		Conteúdo				
		Gramatical	Misto	Vocabulário	Total	Percentual
Formato						
Resposta Limitada	Complete sem Alternativas	15	2	12	29	21%
	Transforme	27	0	0	27	19%
	Escreva	4	15	0	19	14%
	Traduza	0	15	1	16	11%
	Ordem	5	0	0	5	4%
	Total de Resposta Limitada (RL)	51	32	13	96	69%
	Percentual de RL	36%	23%	10%	69%	
Resposta Fechada	Múltipla-Escolha	10	7	8	25	18%
	Complete com Alternativas	10	6	1	17	13%
	Total de Respostas Fechadas (RF)	20	13	9	42	31%
	Percentual de RF	15%	10%	6%	31%	
	Total de RL e RF	71	45	22	138	100%
	Percentual de RL e RF	51%	33%	16%	100%	

A tabela 1 revela os seguintes dados: no campo vertical encontram-se as duas grandes categorias de acordo com o formato: Resposta Limitada e Questões Fechadas. No campo horizontal estão as três categorias referentes ao conteúdo: Gramática, Mista e Vocabulário.

Quanto ao formato, as questões estão distribuídas em subcategorias. Na primeira subcategoria – *Complete sem alternativas* encontramos 15 questões de conteúdo gramatical, 2 de conteúdo misto e 12 questões de vocabulário. Portanto, das 138 questões analisadas, 29 foram elaboradas no formato *Complete sem alternativas*, representando 21% do total. Na segunda subcategoria – *Transforme* – havia 27 questões de conteúdo gramatical e nenhuma questão era mista ou de vocabulário, que representam 19% das 138 questões analisadas. Na terceira subcategoria de formato – *Escreva* – encontramos quatro questões de conteúdo gramatical, 15 de conteúdo misto e nenhuma que enfocasse vocabulário, ou 14% das questões analisadas. Na quarta categoria – *Traduza* – nenhuma questão enfoca conteúdo gramatical. Quinze apresentam conteúdo misto e 1 enfoca vocabulário. Dessa forma, encontramos 11% das questões analisadas no formato Traduza. Na quinta e última subcategoria de questões de Resposta Limitada – *Ordem* – encontramos cinco questões de conteúdo gramatical. Em relação às 138 questões analisadas, esta subcategoria corresponde a apenas 4% das questões.

A análise da categoria Questões Fechadas apresentou os resultados que passamos a relatar. Na primeira subcategoria – *Múltipla Escolha* – 10 questões são de conteúdo gramatical, 7 mistas e 8 de vocabulário. Então, as 25 questões de múltipla escolha representam 18% dos dados. A segunda subcategoria – *Complete com alternativas* – inclui 10 questões de conteúdo gramatical, seis mistas e 1 de vocabulário. O total de questões dessa subcategoria é de 17 questões, o que representa 13% do total. A partir desses dados, no que diz respeito ao formato, encontramos evidências de que a maioria das questões analisadas pertence à categoria Resposta Limitada, que representa 69% das 138 questões analisadas. As Questões Fechadas representam, então, 31% do total.

Quanto à análise do conteúdo na categoria Resposta Limitada, os dados indicam que o conteúdo gramatical representa 36%, o conteúdo misto representa 23% e o conteúdo vocabulário 10%. Sobre a análise de conteúdo na categoria Questões Fechadas, os dados indicam que o conteúdo gramatical foi o foco de 15% das questões, o conteúdo misto de 10% e o de vocabulário 6%. Considerando os dados acima, a análise de conteúdo revelou que o conteúdo Gramatical foi o foco de 51% das questões analisadas; as questões Mistas estiveram em foco em 33% delas, e as de vocabulário em 16%.

DISCUSSÃO

Nossa discussão será iniciada a partir da segunda pergunta da pesquisa: qual é a preferência dos professores sobre o que testar e como testar? Em seguida, a partir dos resultados apresentados acima, tentaremos responder à primeira pergunta da pesquisa – o que as avaliações revelam sobre como os professores percebem o processo de aprendizagem?

Quais são as preferências dos professores sobre o que testar?

A análise do conteúdo das questões elaboradas pelos participantes responde a pergunta sobre “o quê”. A partir da análise dos dados, fica claro que os professores testam predominantemente a gramática, visto que 71% das questões enfocam conteúdo gramatical. Vocabulário é o segundo foco da avaliação, correspondendo a 22% do conteúdo das questões elaboradas. Finalmente, a mistura desses dois conteúdos como foco corresponde a 45% dos dados.

Esses dados parecem indicar que esses professores ainda consideraram o aprendizado de ILE baseado principalmente no conhecimento de gramática e de vocabulário. Quatro justificativas parecem contribuir para esses resultados.

A primeira justificativa pode estar no material de ensino. Por um lado, pode ser que o material adotado ainda enfoque gramática e vocabulário como habilidades primordiais envolvidas no aprendizado de ILE. Esse seria o caso de professores de escolas particulares, que normalmente fazem uso de material didático. Entretanto, como a maioria dos professores deste estudo trabalha em instituições públicas, o problema pode estar na ausência de material. A falta de material didático pode levar os professores a fazerem uso de gramáticas como principal fonte de material a ser usado para a sala de aula, cujo custo é bem menor que

materiais didáticos importados. Em ambos os casos, o resultado é avaliações cujo foco recai sobre o conhecimento da língua, e não sobre a aplicação desse conhecimento a situações que promovam o uso da língua estrangeira. Uma pesquisa mais ampla poderá elucidar essa questão.

Uma segunda justificativa para a preferência dos professores em avaliar gramática e vocabulário pode estar relacionada com suas experiências anteriores como aprendizes. Se eles foram testados apenas em conhecimento gramatical e de vocabulário, eles podem estar apenas reproduzindo uma abordagem a que foram expostos. Se esse for o caso, isso levanta questões importantes sobre o ensino que predomina nos cursos de graduação para professores. É difícil admitir que no século XXI ainda encontramos o foco do ensino direcionado para a gramática e o vocabulário, especialmente após a revolução comunicativa dos anos 70.

As terceira e quarta justificativas são as que consideramos prováveis. A terceira, mais provável, pode ser assim explicada: embora muitos professores conheçam a abordagem comunicativa para o ensino, talvez seja difícil implementá-la, dada a falta de treinamento em conduzir salas de aula com grande número de alunos com o foco nessa abordagem. Ainda, a falta de conhecimento sobre como elaborar avaliações comunicativas pode ser uma justificativa.

A quarta e última justificativa para esses dados é uma hipótese menos provável que precisaria ser verificada empiricamente: embora os professores ensinem adotando a abordagem comunicativa em suas salas de aula, eles continuam a elaborar testes com foco na gramática e no vocabulário por não possuírem conhecimento técnico de como elaborar instrumentos de avaliação comunicativos.

Em qualquer das duas justificativas acima, há a necessidade de pesquisar mais sobre avaliação e, além disso, incluir disciplinas que tratem da avaliação da aprendizagem nos cursos de graduação. Para aqueles que já estão formados devem ser planejados cursos de educação continuada que ofereçam uma oportunidade para esses professores atualizarem suas práticas de ensino e avaliação. Essas duas medidas, se implementadas, podem levar ao aperfeiçoamento da qualidade dos instrumentos de avaliação desenvolvidos por professores.

Quais são as preferências dos professores em como testar?

A análise do formato das questões fornece os dados para responder esta pergunta. Os dados apresentados na sessão dos resultados indicam que a preferência dos professores está em elaborar testes que incluam apenas dois tipos de questões gerais – Resposta Limitada e Questões Fechadas. Esses professores não incluem em seus testes questões que permitam os aprendizes mostrar produção livre. Nos testes analisados, não foi encontrada nenhuma evidência de tarefas de escrita que vão além de responder a uma pergunta feita pelo professor. Portanto, pode-se dizer que esses professores preferem questões controladas à produção livre dos alunos para avaliá-los. A implicação decorrente desta preferência parece indicar que a prática da avaliação é reflexo do que os professores fazem em sala de aula. Dessa forma, há a necessidade de pesquisas que busquem traçar a relação entre o ensino e a avaliação.

Outra evidência apresentada pelos dados sobre Formato é que, embora os professores prefiram métodos mais controlados de avaliação, dos dois tipos, o mais freqüente é o Resposta Limitada, o que indica que eles têm maior preferência pela produção controlada do que por tarefas que requeiram apenas o reconhecimento de respostas corretas. Esse resultado é positivo uma vez que parece indicar que os professores querem encontrar evidência de produção mesmo que limitada.

Três justificativas podem ser apresentadas para essa preferência dos professores. A primeira justificativa para essa preferência pode ser que, considerando que os professores geralmente lecionam grandes grupos de alunos, é mais prático testar por formatos controlados, uma vez que será mais rápido para eles corrigir testes neste formato do que os de questões abertas. No entanto, uma segunda possível razão pode ser que os professores tenham uma proficiência limitada de ILE, o que os levaria a testar em formatos que requerem de seus alunos mostrar conhecimento da língua, mas não a aplicação real da língua em tarefas comunicativas como a escrita, por exemplo.

A razão final e mais provável para essa preferência pode ser a falta de conhecimento, por parte dos professores, de outros formatos que podem ser usados para avaliar aprendizagem de língua. Por exemplo, o uso do formato que requer transferência de informação, o uso de gráficos ou textos para contextualizar a língua bem como o uso de testes temáticos são maneiras de elaborar testes de maneira mais desafiadora para os estudantes e, ao mesmo tempo, oferecer-lhes a oportunidade de aprender enquanto fazem o teste.

No entanto, os professores precisam ter conhecimento de que há diretrizes para a elaboração de bons instrumentos de avaliação que sejam motivadores, válidos e confiáveis. Porém, até onde sabemos, poucos programas de graduação universitária no Brasil incluem avaliação como curso para que os professores em formação tenham contato com essas diretrizes. Na verdade, os cursos de avaliação não são obrigatórios nem mesmo no mestrado, tornando-se, então, opcionais para aqueles que têm interesse pessoal sobre o assunto.

O que os testes revelam sobre como os professores percebem a avaliação e o processo de aprendizagem?

Não há dúvidas de que a análise dos testes elaborados pelos professores que participaram deste estudo de caso revelou uma abordagem tradicional de avaliação e uma visão do processo de aprendizagem bem distante das concepções de aprendizagem de língua que predominam atualmente.

A abordagem tradicional de avaliação é revelada pelo conteúdo dos testes cujo foco é na gramática e no vocabulário. Testes comunicativos enfocam não apenas as habilidades subjacentes, mas também, a produção oral, a compreensão oral, a leitura e a escrita. Além do mais, o uso de formatos controlados, como a resposta limitada e as questões fechadas, é evidência da aplicação de uma abordagem tradicional na qual a evidência de conhecimento dos alunos está em reconhecer ou em preencher lacunas cujo foco é isolado de contexto ou de aplicação a situações comunicativas.

No que diz respeito à concepção do processo de aprendizagem, os testes revelam uma visão da língua como uma questão de conhecimento de gramática e seu vocabulário. No entanto, aprendizagem de uma língua atualmente é considerada muito mais que isto. A visão atual sobre o processo de aprendizagem de línguas deixou a crença de que aprender línguas é apenas uma questão de conhecimento de estruturas gramaticais e vocabulário para trás desde o final dos anos 70, quando a abordagem comunicativa de ensino de línguas começou a focar as publicações de Allwright (1977) Brumfit and Johnson (1978) Widdowson (1978). Aprender uma língua significa desenvolver competências, tais como a competência gramatical, a competência discursiva, a competência pragmática e a competência estratégica (Canale e Swain, 1980). Portanto, a gramática é apenas um dos aspectos da aprendizagem de línguas.

Nenhum dos testes analisados avalia a compreensão ou a produção oral ou as habilidades de leitura e escrita, o que deixa em aberto a possibilidade de que essas dimensões não existem nas aulas desses professores, dada a premissa de que os testes refletem o que foi coberto durante um curso.

A possibilidade de esses resultados serem confirmados em outros estudos futuros abriria a agenda de pesquisas sobre avaliação de aprendizagem de ILE no Brasil com o status de um tema urgente. Outros projetos de pesquisa deveriam ser conduzidos para indicar se a situação aqui apresentada é isolada ou comum. No entanto, embora urgente, o tema a ser investigado mostra-se extenso, pois prevemos um longo caminho a ser percorrido no sentido de melhorar a situação da avaliação da aprendizagem de ILE no Brasil.

CONCLUSÃO

A pesquisa ora apresentada é considerada um passo inicial para a compreensão do que os professores incluem em seus testes e como eles abordam os instrumentos de avaliação. No entanto, como qualquer projeto de pesquisa, este tem pelo menos três limitações.

A primeira refere-se ao número de participantes em nosso estudo. Apesar do número significativo de questões analisadas, o número de participantes é pequeno para se traçarem conclusões definitivas. A discussão acima deve ser considerada do ponto de vista de um estudo de caso. Portanto, seria desejável que outros projetos de pesquisa similares a este fossem desenvolvidos para sabermos se o que foi evidenciado aqui sobre a abordagem à avaliação de ILE é a norma ou a exceção.

Uma segunda limitação refere-se à falta de entrevistas que nos permitissem conhecer os princípios subjacentes à construção de cada teste analisado. Perguntas como: por que o professor incluiu este foco ou por que ele escolheu este ou aquele formato, continuam sem resposta. Nesta discussão, muitas possibilidades foram levantadas, mas elas não podem ser mais que especulações se não houver dados que lhes dêem suporte.

Uma limitação final está no foco do instrumento de avaliação desvinculado do ensino que o acompanha. Um estudo etnográfico que enfocasse o papel da avaliação como parte

do processo de ensino mostraria um quadro mais preciso de como a avaliação está envolvida no processo de ensino e aprendizagem de uma língua estrangeira.

No entanto, apesar das limitações acima, nosso estudo apresenta um quadro metodológico para a categorização de questões de testes baseado na literatura da área, fornecendo uma base metodológica para estudos mais aprofundados ou para replicações deste estudo com diferentes populações de professores. Além disso, os resultados revelam uma visão da avaliação utilizada por professores de ILE em formação. Temos consciência de que o quadro apresentado é limitado em sua abrangência. Ainda assim, é digno da atenção de professores e pesquisadores para que se possa vislumbrar uma mudança na realidade que, acreditamos, seja similar⁵ àquela que aqui descrevemos. Um maior número de estudos nos dará uma visão mais precisa dessa realidade.

Quanto às implicações deste estudo para professores e pesquisadores, acreditamos na necessidade de se incluir, em cursos de graduação, disciplinas na quais se trate da questão da avaliação da aprendizagem de ILE de forma específica⁶. Essa especificidade permite aos alunos entenderem a distância que existe entre uma avaliação elaborada para dar uma nota e uma avaliação que pretende informar o aluno e o professor sobre o processo de aprendizagem. Na primeira, o objetivo é dar uma nota que pouco pode informar ao aluno sobre o seu processo de aprendizagem ou ao professor sobre o processo de ensino. Na segunda, esse objetivo determina uma série de decisões que extrapolam a checagem do conhecimento do conteúdo. Nela, o professor elabora a avaliação para que o aluno possa ter um retorno sobre a sua aprendizagem a partir de seu desempenho. Esse retorno vem através dos comentários do professor que devem acompanhar os resultados, personalizando-os. Afinal, o que significa um mesmo 80 para dois alunos diferentes? Que eles aprenderam a mesma coisa? Que eles aprenderam de forma igual?

Como bem ilustram as experiências relatadas por Pellegrini (2003), avaliar deve servir para um melhor ensino e uma melhor aprendizagem. Alunos que recebem de seus professores mais que notas, mas avaliações mais globais e personalizadas têm melhor desempenho (Pellegrini, *ibid*). Mas, para que essa avaliação voltada para o aprimoramento do ensino e da aprendizagem aconteça em curto prazo, é necessário pensar em programas de educação continuada. Professores que já estão em exercício precisam de cursos de atualização para realizar avaliações fundamentadas em conhecimentos sobre a natureza de avaliações bem elaboradas. Acreditamos que esse conhecimento levaria à elaboração de melhores avaliações, sejam elas tradicionais e objetivas ou críticas e alternativas e à redução da frustração do professor em relação à avaliação, mencionada na introdução deste estudo.

⁵ Em setembro de 2002, no XII Encontro da Associação dos Professores do Espírito Santo, onde uma versão deste artigo foi apresentada, os participantes reportaram acreditar que os resultados deste estudo podem refletir uma realidade que extrapola o grupo investigado.

⁶ Segundo alguns alunos que cursaram a disciplina “Tópicos de Prática de Ensino: Avaliação” no primeiro semestre de 2003 na Faculdade de Letras da UFMG, a disciplina que tratou da questão avaliação na Faculdade de Educação tinha um enfoque geral, sem a especificidade que lhes foi apresentada na disciplina da Faculdade de Letras.

Além disso, é necessário que mais pesquisadores voltem seus olhos para a questão da avaliação em salas de aula de todos os tipos. Sem um olhar mais amplo sobre a avaliação da aprendizagem de ILE que permita sua documentação, manteremos a noção de que o problema da aprendizagem está só no ensino, como afirma Locatelli (2001), sem a devida atenção ao papel da avaliação tanto para o ensino como para a aprendizagem.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALLWRIGHT, R. (1977). Language learning through communication practice. *ELT Docs*, 76, n. 3, p. 2-14.
- BROWN, H. D. (2001). *Teaching by Principles*. New York: Addison Wesley Longman.
- BRUMFIT, C.J., JOHNSON, K. (1978). *The communicative approach to language teaching*. New York: Oxford University Press.
- CANALE, M.; SWAIN, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, p.1-47.
- LOPES, C. R. (2002). *A avaliação no ensino comunicativo de língua inglesa: uma análise de testes escritos*. Dissertação de Mestrado não publicada. Letras: UFG.
- GENESEE, F; UPSHUR, J. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- HUGHES, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- JOHNSON, D. (1992). *Approaches to research in second language learning*. New York: Longman.
- LOCATELLI, I. *O desafio de avaliar a educação básica no Brasil*. <http://www.inep.gov.br/imprensa/artigos/iza01.htm>, 2001.
- LUCKESI, C.C. (1997). *Avaliação da aprendizagem escolar*. São Paulo: Cortez.
- MADSEN, H.S. (1983). *Techniques in testing*. New York: Oxford University Press.
- NEVES, M. (2002). *Processo discursivo e subjetividade: vozes preponderantes na avaliação da oralidade em língua estrangeira no ensino universitário*. Tese de Doutorado não publicada. IEL. Unicamp: SP.
- PENNYCOOK, A. (2004). *Critical applied linguistics*. In: DAVIES, A. e C. Elder (eds.) *Handbook of Applied Linguistics*. Oxford: Blackwell.
- PELLEGRINI, D. (2003). *Avaliar para ensinar melhor*. Revista Nova Escola, n.159, p.26-33, SP: Abril.
- PORTO, C. V. (2003). *Percepções de professoras de Letras/Inglês sobre avaliação de aprendizagem: um estudo de caso*. Dissertação de Mestrado não publicada. Programa de Pós-Graduação em Estudos Lingüísticos da Faculdade de Letras, UFMG. Belo Horizonte: MG.
- SARMENTO, D.C. (Org.) (1997). *O discurso e a prática da avaliação na escola*. Campinas: Pontes.

MICCOLI - A avaliação da aprendizagem de ILE: uma realidade que choca

- SCARAMUCCI, M. V. R. (1997). Avaliação de rendimento no ensino-aprendizagem de língua Estrangeira. In: ALMEIDA FILHO, J. C. (Org.). *Parâmetros atuais para o ensino de português língua estrangeira*, p.75-88.
- SHOHAMY, E. (2001). *The power of tests: a critical perspective on the use of language tests*. London: Longman.
- WEIR, C. J. (1990). *Communicative language testing*. Hemel Hempsted: Prentice Hall Regents.
- WIDDOWSON, H. D. (1978). *Teaching language as communication*. Oxford: Oxford University Press.