

Avaliação da Oralidade em Língua Inglesa: a atribuição de notas como expressão da subjetividade do professor

ASSESSING SPEAKING IN ENGLISH: GRADING AS AN EXPRESSION
OF TEACHERS' SUBJECTIVITY

Gladys **QUEVEDO-CAMARGO***
Lucas Henrique **GARCIA****

Resumo: Tendo em vista a expressiva literatura em Avaliação da Aprendizagem de Línguas, este artigo tem por objetivo compreender e discutir, a partir de estudos que tratam da confiabilidade (BROWN; BAILEY, 1984; BROWN, 2004; SCHOFFEN, 2009) e da avaliação da oralidade (BACHMAN, 1990; SCARAMUCCI, 2000; BORGES-ALMEIDA, 2009; QUEVEDO-CAMARGO, 2014), de que maneira os professores de inglês de um instituto de idiomas atribuem notas para mensurar o desempenho oral de seus alunos, considerando a adoção (ou não) de critérios previamente estabelecidos. Realizamos uma pesquisa qualitativa interpretativista, cujos dados foram coletados por meio de entrevistas semiestruturadas com cinco professores. Diante das discrepâncias entre as notas obtidas pelos alunos na

* Professora Adjunta efetiva do Departamento de Línguas Estrangeiras e Tradução (LET) do Instituto de Letras (IL) da Universidade de Brasília (UnB). Docente permanente do Programa de Pós-Graduação em Linguística Aplicada (PGLA) na mesma instituição. Mestre (2007) e Doutora (2011) em Estudos da Linguagem pela Universidade Estadual de Londrina. Pós-Doutorado com pesquisa na área de avaliação pela UNICAMP (2013). Contato: gladys@unb.br.

** Programa de Pós-Graduação em Linguística Aplicada da Universidade de Brasília (UnB). Mestre em Linguística Aplicada – PGLA/UnB (2016). Desenvolve trabalho sobre avaliação da proficiência oral de professores de língua inglesa. Contato: lucashg2@hotmail.com.

avaliação da expressão oral, quais são os fatores que influenciam e reforçam essa constatação? Até que ponto a atribuição de notas na avaliação oral reflete o real desempenho do aluno? Em situações de turmas diferentes e professores distintos, como assegurar a confiabilidade da avaliação da produção oral de alunos de um mesmo nível de referência? Os resultados da investigação indicam a necessidade de adotar critérios previamente estabelecidos, uma vez que, por conta da excessiva subjetividade, os testes orais parecem não aferir o nível esperado de produção oral dos aprendizes e tampouco refletem a realidade de uso da língua.

Palavras-chave: Avaliação da oralidade. Subjetividade. Atribuição de notas.

Abstract: Taking into account the significant literature in Language Teaching Assessment, this article aims at understanding and discussing, based on studies about reliability (BROWN; BAILEY, 1984; BROWN, 2004; SCHOFFEN, 2009) and speaking assessment (BACHMAN, 1990; SCARAMUCCI, 2000; BORGES-ALMEIDA, 2009; QUEVEDO-CAMARGO, 2014), how English language teachers working in a language institute grade their students' oral performance, considering the adoption (or not) of previously established criteria. A qualitative-interpretivist study was conducted and the data was collected by means of semi-structured interviews with five teachers. In face of the discrepancies among the students' oral assessment grades, what are the factors that influence and reinforce such discrepancies? To what extent speaking assessment grading mirrors students' real performance? Where there are different groups and different teachers, how to ensure speaking assessment reliability of students at the same reference level? The findings indicate the need to adopt previously established criteria as, due to excessive subjectivity, oral tests do not seem to measure the students' expected level of oral performance, nor mirror language use reality.

Keywords: Speaking assessment. Subjectivity. Grading.

Introdução

A avaliação da produção dos aprendizes em língua estrangeira (doravante LE), apesar de considerada como de extrema relevância para o

sucesso da aprendizagem, parece ser desconsiderada no currículo das licenciaturas em língua inglesa no Brasil. Este fator se reflete claramente na prática em sala de aula, considerando-se que uma formação de professores deficiente implica em um círculo vicioso que leva a consequências negativas, tais como baixa proficiência e aprendizagem também deficiente (CONSOLO et al., 2008).

Quando analisamos com mais especificidade a avaliação das habilidades orais, Luoma (2004) e Borges-Almeida (2009) afirmam que a produção oral é uma das habilidades mais complexas em termos de estabelecimento de critérios e instrumentos adequados para proceder à avaliação.

Brown (2004) nos conscientiza, ainda, sobre os desafios nos procedimentos de atribuição de notas, que são influenciados necessariamente por uma série de aspectos, que vão desde a liberdade de escolha dada ao aluno até o formato das tarefas e a definição do que se quer mensurar. Ainda neste mesmo sentido, Tarone (1998) assevera que a própria natureza do construto “oralidade” em LE só pode ser definida em um contexto específico.

O que constatarem esses autores nos leva a refletir sobre a maneira como os professores de língua inglesa têm avaliado a produção oral de seus alunos, uma vez que, apesar da eminente dificuldade em aferir a realidade do desempenho oral, isto não diminui (ou não deveria diminuir) a importância de se refletir e problematizar a questão para fins de aprimoramento dos instrumentos.

Sendo assim, quanto aos conceitos básicos de avaliação, a confiabilidade é a que mais suscita discussão quando se trata de avaliação da produção oral, como indica Luoma (2004). A falta de confiabilidade dos instrumentos (ou, nesse contexto, testes orais) pode resultar em consequências desastrosas para uma série de contextos de ensino de línguas, especialmente para a realidade um instituto de idiomas, em que turmas de níveis equivalentes são avaliadas de maneira distinta por professores também diferentes. Discutir maneiras de aumentar a confiabilidade é, então, criar condições para que o processo avaliativo exerça de fato seu papel dinâmico e contínuo de retroalimentar e redirecionar a aprendizagem e permitir que se avance a partir dali (SCARAMUCCI, 1998).

Tendo em vista este posicionamento, percebemos a necessidade de desenvolver um estudo que buscasse entender de que maneira os professores

de inglês atuantes em um instituto de idiomas concebem o processo de avaliação e atribuem, conseqüentemente, menções/notas para mensurar a produção oral de seus alunos, a partir da problemática inicial, que gira em torno da discrepância entre as notas atribuídas ao desempenho oral dos alunos de um mesmo nível de referência.

Para isto, procuraremos responder às seguintes perguntas que norteiam este estudo: (a) qual a possível explicação para a discrepância entre as notas de desempenho oral dos alunos de um mesmo nível de referência no referido instituto de idiomas?; (b) de que maneira e até que ponto a atribuição de notas na avaliação oral do professor reflete o real desempenho do aluno?; (c) em situações de turmas diferentes e professores distintos, como conceber o nível de confiabilidade da avaliação do desempenho oral de alunos que se encontram em um mesmo nível de referência?

As seções que compõem este trabalho estão organizadas da seguinte maneira: primeiramente, apresentamos o referencial teórico sobre avaliação e, especificamente, confiabilidade, ao qual nos afiliamos para desenvolver esta pesquisa; em seguida, fazemos considerações sobre a subjetividade do professor de línguas no momento de avaliar a produção oral. Na terceira seção, esclarecemos a natureza da pesquisa e os instrumentos metodológicos que viabilizaram a coleta de dados e sua posterior análise e discussão, estas últimas apresentadas na próxima seção. Finalmente, utilizamos as perguntas norteadoras supracitadas para explicar as considerações finais, retomando, a fim de respondê-las, as discussões apresentadas durante a análise dos resultados.

Considerações Conceituais: a confiabilidade e suas implicações

Segundo Brown (2004, p. 27), “um teste confiável é consistente e seguro”¹. Isto implica dizer que, se o mesmo teste é aplicado em diferentes ocasiões e produz resultados semelhantes, trata-se de um teste de alta confiabilidade.

¹ “a reliable test is consistent and dependable”.

Todas as traduções deste artigo são de nossa responsabilidade, sendo transcrito o trecho original nas subseqüentes notas de rodapé.

O autor elenca, ainda, uma série de características de um teste confiável, das quais duas nos são especialmente apropriadas: primeiro, é um teste que se mantém consistente entre duas ou mais aplicações; em segundo lugar, que possui rubricas² uniformes para a atribuição de notas.

A confiabilidade estaria, então, diretamente relacionada com a atribuição de notas, uma vez que um teste confiável permite que diferentes avaliadores em diferentes contextos atribuam notas similares.

Bachman (1990, p. 20) nos lembra, no entanto, que “a confiabilidade tem a ver com a consistência das medidas em diferentes momentos, formas de teste, avaliadores e outras características do contexto de aferimento”³.

Se considerarmos, por exemplo, o contexto de dois professores de língua inglesa que avaliam a produção de alunos que estão enquadrados em um mesmo nível, mas em grupos diferentes, espera-se que, na realidade de um instituto de idiomas, os testes aplicados (sejam eles orais ou escritos) sejam confiáveis o bastante para que as notas atribuídas sejam consistentes em ambas as situações, como considera Bachman (1990, p. 20):

Suponhamos que dois avaliadores dêem notas amplamente diferentes para o mesmo exercício de escrita. Na ausência de qualquer outra informação, nós não temos nenhuma base para decidir qual nota considerar, e podemos, conseqüentemente, considerar ambas não-confiáveis.⁴

É importante considerar, assim, que a confiabilidade não se restringe ao teste em si, isto é, não necessariamente se verifica apenas na observação da elaboração do teste e em suas características. Na verdade, o teste é apenas

² O termo é concebido aqui em sua acepção dentro da terminologia educacional, como prevê o *National Science Education Standards* (NATIONAL RESEARCH COUNCIL, 1996, p. 93), isto é, como “um padrão de desempenho para uma determinada população”.

³ “Reliability has to do with the consistency of measures across different times, test forms, raters, and other characteristics of the measurement context”.

⁴ “Suppose two raters gave widely different ratings to the same writing sample. In the absence of any other information, we have no basis for deciding which rating to use, and consequently may regard both as unreliable”.

um dos instrumentos que compõem o contexto da avaliação, como aponta o diagrama de Rea-Dickens e Gemaine (1993), em que os autores dividem este contexto em quatro macroseções com subdivisões, que vão desde os aspectos relacionados à sociedade de modo geral (como atitudes, expectativas, influências culturais e religiosas, entre outros) até chegar ao microcontexto, representado pela sala de aula, sendo que, neste último, a própria relação professor-aluno também é vista como fator de composição da realidade contextual da avaliação.

Nesse sentido, Brown (2004) salienta quatro fatores que podem influenciar significativamente o grau de confiabilidade de um teste. O primeiro trata da confiabilidade relacionada ao aluno (*student-related reliability*). O autor elenca uma série de fatores individuais – tais como doença, fadiga, ansiedade e estratégias para a execução de testes – como elementos a serem considerados.

Brown (2004) continua a série de fatores com a confiabilidade relacionada à administração do teste (*test administration reliability*) e à própria confiabilidade do teste (*test reliability*), sendo a primeira definida a partir das condições necessárias (ou a falta delas) para o bom andamento do teste em questão, enquanto a segunda considera a própria natureza e os propósitos do teste como fatores que podem implicar em erros no momento de mensurar a produção dos aprendizes (ou, nesse contexto, *test-takers*).

O quarto fator de influência, e o que mais nos interessa para fins de esclarecimento neste artigo, trata da confiabilidade relacionada ao avaliador (*rater reliability*), que pode ser dividida em (i) confiabilidade entre avaliadores distintos e (ii) de um mesmo avaliador em situações distintas (em inglês, *inter-rater e intra-rater reliability*).

O primeiro caso consiste em falta de confiabilidade quando dois avaliadores atribuem notas inconsistentes para um mesmo teste. As razões que motivam esta inconsistência são variados. Entretanto, o autor destaca que isto ocorra “possivelmente por falta de atenção aos critérios para atribuição de notas, inexperiência, distração, ou mesmo por pré-concepções”⁵ (BROWN, 2004, p. 21).

⁵ “Possibly for lack of attention to scoring criteria, inexperience, inattention, or even preconceived biases”.

Já no segundo caso, isto é, de falta de confiabilidade de um mesmo avaliador, Brown (2004, p. 21) destaca que este fenômeno é comum em professores atuantes em sala de aula (*classroom teachers*) e é originado em uma série de fatores, tais como: falta de clareza na atribuição e aplicação de critérios de atribuição de notas, cansaço, ou mesmo falta de cuidado.

Dentre as distinções de confiabilidade, interessa-nos em especial analisar a confiabilidade do avaliador, o que se justifica pelo contexto desta pesquisa, que se centra na maneira como os professores-avaliadores de um instituto de idiomas procedem à atribuição de notas nos testes orais aos quais submetem seus alunos.

Ora, se existem fatores externos ao teste que influenciam diretamente os resultados obtidos após sua aplicação, o avaliador é, necessariamente, um deles. Schoffen (2009) define a confiabilidade com base no efeito desses fatores, pois, para a autora, “a confiabilidade de um teste reside em minimizar os efeitos na avaliação dos fatores externos ao teste, fazendo com que a habilidade linguística do candidato seja avaliada de maneira uniforme” (p. 23).

Partindo-se do mesmo pressuposto e corroborando a concepção de confiabilidade enquanto princípio que pressupõe a diminuição dos efeitos externos na avaliação, Brown e Bailey (1984) discutem maneiras que nos permitam fazê-lo, tais como maior especificação do instrumento e, fundamentalmente, dos critérios que dão suporte à grade de avaliação.

A literatura aponta, portanto, para a importância da acuidade do avaliador, para que ele tenha à sua disposição não apenas um instrumento de avaliação adequado ao nível de referência de seus aprendizes e à habilidade a qual se propõe mensurar, mas também uma série de critérios que justifiquem e embasem suas considerações e atribuições de notas, seja dentro da realidade de um mesmo grupo com o mesmo professor, ou no contexto de grupos diferentes, com professores-avaliadores diferentes, mas de mesmo nível de referência.

Estudos e investimentos que proporcionem aumento da confiabilidade dos instrumentos e dos agentes humanos envolvidos no contexto da avaliação permitiriam uma diminuição consistente nas discrepâncias entre os resultados obtidos pelos alunos e, conseqüentemente, fortaleceriam a concepção da avaliação como “mola propulsora da aprendizagem”, como proposto por Scaramucci (1998).

Quando tratamos de avaliadores e os concebemos como “agentes humanos” – excluindo-se, para o contexto desta investigação, a avaliação automática feita por máquinas de correção – é fundamental que discutamos, também, o grau de subjetividade que a agência destes indivíduos pode trazer ao processo de avaliar e suas consequências, sobretudo em se tratando de avaliação da produção oral, questão esta que discutiremos na seção seguinte.

Avaliação da Produção Oral: subjetiva por natureza?

Para iniciarmos a discussão sobre subjetividade da avaliação da produção oral em língua estrangeira, faz-se necessário compreender, primeiramente, o que se concebe como avaliação da aprendizagem. Neste sentido, Luckesi (1978) advoga que “a avaliação é um julgamento de valor sobre manifestações relevantes da realidade, tendo em vista uma tomada de decisão.”

Se considerarmos a definição de Luckesi e trouxermos para a realidade do que se concebe como avaliação da aprendizagem de línguas, os pontos levantados pelo autor nos abririam discussões imprescindíveis. Em razão do contexto da pesquisa aqui relatada, nos limitaremos a discutir apenas a primeira parte da definição, isto é, a avaliação vista como juízo de valor.

Analisando a definição do autor por um viés de subjetividade, é possível compreender que, se na própria concepção do que é avaliar, o julgamento de valor está presente, este fenômeno é uma realidade do processo avaliativo e deve, portanto, ser considerado como parte inerente a ele, como advoga Ferraz (1994, p. 2), ao assumir que, desde a escolha que se faz dos itens de um teste, por exemplo, até a linguagem que se utiliza, a presença da subjetividade é inevitável.

Aceitar a subjetividade em avaliação é condição para um aprofundamento dos problemas e uma melhor aproximação da realidade. Aceitar a subjetividade em avaliação é ainda a forma mais eficaz de tentar controlá-la, evitando a ilusão de que a objetividade é possível e de que o aluno é aquilo que o teste mede (FERRAZ, 1994, p. 2)

O primeiro passo para tratarmos a questão da subjetividade em avaliação é, desse modo, aceitar que ela existe. Isso não significa, contudo, que não devemos buscar meios desejáveis para relativizá-la, pois o conflito entre os diversos agentes subjetivos que permeiam o contexto da avaliação – como exposto na seção anterior – pode levar à não confiabilidade do teste ou de qualquer outro instrumento utilizado para o mesmo fim.

O próprio Luckesi, autor da definição de avaliação que utilizamos no início desta seção, nos esclarece o que o “juízo de valor” prevê nesse contexto, afirmando a importância da presença de critérios pré-estabelecidos, a partir dos quais a afirmação qualitativa sobre o que o instrumento aferiu é consolidada. O autor prossegue, definindo as implicações de estabelecer juízo de valor:

O julgamento de valor, por sua constituição mesma, desemboca num posicionamento de ‘não-indiferença’, o que significa obrigatoriamente uma tomada de posição sobre o objeto avaliado, e, uma tomada de decisão quando se trata de um processo, como é o caso da aprendizagem (LUCKESI, 1998, p. 33).

Xavier (1998) corrobora a visão de Luckesi e de Ferraz ao assegurar a não neutralidade da avaliação, que também não seria acidental. Para a autora, “ela [a avaliação] carrega as intenções, ideologias, crenças e concepções de ensino e aprendizagem do professor” (p. 115).

Uma vez que a subjetividade se consolida, então, como elemento essencial a ser considerado no processo de avaliação, como garantir a confiabilidade dos resultados dos instrumentos que utilizamos para avaliar e, ao mesmo tempo, ter a dimensão de que eles coexistem em caráter subjetivo?

Uma investigação que busque uma possível consideração sobre esse questionamento nos leva indiscutivelmente à análise de um elemento-chave no que se concebe como avaliação: a nota.

A literatura é rica no que concerne à temática da atribuição de notas na avaliação educacional (PERRENOUD, 1993; LUCKESI, 1998; XAVIER, 1998; BYRAM, 2002; entre outros), mas os autores divergem quanto à importância (e até mesmo a necessidade ou não) de se avaliar tendo como resultado a nota (ou menção).

Ainda que as notas sejam “um juízo de valor questionado” (XAVIER, 1998), o fato é que, se considerarmos o contexto da avaliação no Brasil e todos os agentes envolvidos (escola, pais, alunos, professores, instituições governamentais, entre outros), a atribuição de uma nota às produções parece indispensável no processo de ensino-aprendizagem – ainda que não numérica.

Nesse sentido, concordamos com a posição de Freire (1997 apud XAVIER, 1998) ao assumir que “o que não é possível é deixar de atribuir alguma graduação. O fundamental é diminuir o arbítrio do professor, mas não negar a necessidade de uma certa valoração à prática do aluno.”

O que nos parece consistente na literatura, no entanto, é a necessidade de buscar práticas que diminuam ou minimizem os efeitos da subjetividade (ou arbítrio, para Freire) – existente e que, de fato, influencia o resultado do processo de avaliação e a consequente atribuição de notas – do professor-avaliador.

Quando voltamos nossa análise para a subjetividade na avaliação da produção oral em língua estrangeira, observamos, na literatura (NUNAN, 1992; TARONE, 1998; BROWN, 2004; BORGES-ALMEIDA, 2009; QUEVEDO-CAMARGO, 2014; entre outros) a convicção de que a oralidade é, de fato, complexa, entre outros fatores, por sua natureza e pelas crenças que permeiam esta habilidade.

Borges-Almeida (2009, p. 124) nos lembra, em consonância com o que afirmam Brown (1994) e Scaramucci (2000) que “a concepção de que o falar é ‘maior’ que qualquer outra habilidade linguística quando se trata do uso da língua é uma representação cultural”, uma vez que a crença de que saber a língua é “falar” a língua ainda se faz presente.

Sendo a expressão oral a habilidade que está diretamente relacionada às noções tanto gerais quanto populares de “fluência” na língua estrangeira (SCARAMUCCI apud BORGES-ALMEIDA, 2009), sua avaliação deveria demandar preocupações igualmente significativas, uma vez que um teste oral precariamente elaborado ou pouco confiável pode ter efeito retroativo⁶

⁶ Não nos aprofundaremos na discussão sobre o efeito retroativo das avaliações para não perdermos de vista o objetivo do artigo. No entanto, neste trecho, compreendemos o termo em sua concepção mais ampla, isto é, como influência que determinado teste exerce em seu contexto de aplicação, seja esta influência positiva ou negativa, como apontam Alderson e Wall (1993).

negativo tanto para o aprendiz quanto para o processo de ensino-aprendizagem em si.

Como assegurar, no entanto, que os professores-avaliadores exerçam menos influência na atribuição de notas de um determinado teste oral? Uma possível solução seria a elaboração cuidadosa de tarefas que fossem desenvolvidas e aplicadas para contextos específicos (BROWN, 2004), aumentando a confiabilidade do instrumento e, ao mesmo tempo, respeitando a natureza contextual da oralidade (TARONE, 1998).

Em contrapartida, é fundamental salientar que Borges-Almeida (2011, p. 134) suscita a problemática de que é “um equívoco acreditar que o saber avaliar seja uma competência nata que não tenha de ser necessariamente desenvolvida, como se fosse um brinde que levamos quando adquirimos o pacote da formação de professor.”

Daí a necessidade de realizar pesquisas que investiguem de que maneira os professores atuantes em seus respectivos contextos vêm avaliando seus alunos, não a fim de demonstrar falta de habilidade ou de cuidado com o processo por parte do professor, mas, sim, com o objetivo de diagnosticar possíveis inconsistências e agir para que elas sejam posteriormente superadas.

A breve revisão da literatura realizada ao longo desta seção nos permite considerar que a produção oral em língua estrangeira implica em determinado nível de subjetividade. Contudo, o que os autores sugerem é que a natureza da habilidade é subjetiva, o que não significa que sua avaliação não deva seguir critérios pré-estabelecidos e nem mesmo que se trata de um juízo de valor a ser atribuído sem fundamentação.

Na realidade, ainda que não haja consonância quanto à maneira de mensurar e atribuir notas ao final do processo de avaliação, o que nos parece consonante é a aceitação da subjetividade do que se busca medir e, a partir desta aceitação, a realização de um trabalho que preveja medidas que reduzam os efeitos desta subjetividade na realização do teste. Dessa maneira, devemos refletir sobre maneiras de avaliar a expressão oral que considerem um grau adequado de confiabilidade e que, conseqüentemente, contribuam para a diminuição da subjetividade em um processo, em si, considerado de natureza subjetiva.

Considerações Metodológicas

A pesquisa aqui apresentada é uma investigação de natureza qualitativa interpretativista. O instrumento utilizado para a coleta de dados foi uma entrevista semiestruturada e o contexto em que se inserem os participantes desta investigação é um instituto de idiomas privado, localizado em uma cidade do Distrito Federal.

Um volume considerável de estudos em pesquisa qualitativa aponta para a importância da natureza interpretativista das investigações (como CHIZZOTTI, 2006; DENZIN; LINCOLN, 2006; GERGEN; GERGEN, 2006), uma vez que a interpretação e o caráter subjetivo são inerentes às atividades humanas, não podendo, portanto, serem excluídos do processo de pesquisa em ciências sociais aplicadas.

Chizzotti (2006) discorre sobre a interpretação como parte fundamental do que se concebe como pesquisa qualitativa, pois seu objetivo seria interpretar o sentido dos acontecimentos tendo como base o significado atribuído pelas pessoas ao utilizar a língua para falar ou fazer (isto é, operar). O autor é contundente ao afirmar que “qualquer paradigma deve recorrer à intuição humana e à inferência interpretativa” (p. 28).

No que diz respeito ao método de coleta dos dados, Silveira (2002, p. 119-120) assegura que a entrevista é amplamente utilizada nas pesquisas em ciências humanas, principalmente na área de educação. No entanto, concordamos com a ressalva da autora de que, sob a luz de estudos em áreas da linguística, em especial da Análise da Conversação e da Sociolinguística Interacional, o papel da entrevista na pesquisa vem sendo questionado, uma vez que esta deve ser vista como um “evento discursivo complexo”, no sentido de envolver muito mais que a simples dicotomia pergunta-resposta.

Sobre esta redefinição dos objetivos da entrevista, Rosa e Arnoldi (2006) destacam a importância da entrevista semiestruturada nos estudos qualitativos, tendo em vista o caráter de flexibilidade conferido a este instrumento, o que nos permite observar a verbalização do pensamento do entrevistado e maior fluidez por parte do entrevistador, características estas que consideramos de fundamental importância para as pesquisas em avaliação, especialmente quando se observa questões relativas à subjetividade no processo avaliativo.

Foram entrevistados cinco professores de inglês atuantes em um instituto de idiomas localizado em uma cidade do Distrito Federal, aos quais foram atribuídos nomes fictícios, com o objetivo de resguardar o anonimato. O tempo de experiência dos participantes em sala de aula varia entre 3 e 20 anos. As entrevistas foram devidamente gravadas em arquivo de áudio e transcritas para a realização da análise e de suas posteriores discussões.

Análise dos Dados e Discussão dos Resultados

Apresentaremos, nesta seção, excertos das entrevistas realizadas com os professores, a fim de discuti-los e problematizá-los em seguida. Para fins de organização, seguiremos a sequência do roteiro de entrevista (Anexo) em nossa discussão.

Desse modo, analisaremos – de forma contrastiva e com base no referencial teórico explicitado nas seções anteriores – as respostas dadas pelos professores para cada uma das perguntas da entrevista, de forma a compreender de maneira mais profunda as perspectivas adotadas por estes participantes frente aos questionamentos propostos.

As duas primeiras perguntas do roteiro de entrevista nos foram úteis para contextualizar o tempo de atuação em sala de aula de cada professor e a sua análise e/ou classificação geral do processo de avaliação.

Todos os professores classificaram o processo de avaliação ao qual submetem seus alunos como “bom”, utilizando-se de adjetivos que variam de “adequado” e “satisfatório” até “ótimo” e “completo”.

As respostas dadas corroboram e se justificam com as explicações fornecidas para a pergunta 3 do roteiro de entrevista (*Há recomendação da instituição de ensino quanto a como se deve avaliar oralmente os alunos?*), uma vez que todos afirmaram que a coordenação do instituto de idiomas em questão lhes dá “liberdade” para avaliar os alunos oralmente, ainda que as provas escritas sejam padronizadas de acordo com o material didático adotado, como demonstram os excertos abaixo.

Élida: De todos os lugares que já trabalhei, o [nome da instituição] foi onde me senti mais livre para avaliar como acho mais adequado. É ótimo! Tirando a prova escrita, que é padrão.

Carlos: *A recomendação é que a gente faça o aluno se comunicar na prova oral, pra poder atribuir ponto, né? Mas, assim, de modo geral, temos toda a liberdade pra avaliar.*

Podemos inferir dos excertos acima que a instituição não estabelece critérios a serem cumpridos para a atribuição de notas nos testes orais, tampouco aponta direcionamentos sobre a natureza do teste em si. No entanto, o fato de estabelecer um padrão para os testes escritos e não fazer considerações sobre os testes orais confirma a afirmação de Bygate (1987), de que a habilidade de produção oral é subvalorizada se comparada à produção escrita, exatamente por ser considerada como um mero improviso ou vista de maneira coloquial, o que pode implicar em menos seriedade e prestígio.

Quanto à quarta pergunta do roteiro de entrevista (*Em quais momentos do período/semestre a aplicação de testes ocorre? Há feedback para o aluno após a realização/correção dos testes?*), os participantes afirmaram que os testes são bimestrais, isto é, ocorrem em dois dias (um para cada habilidade: escrita e oral) e na mesma semana, sempre ao final de cada bimestre. Quanto ao *feedback*, elemento considerado de extrema importância para um processo de avaliação ser considerado formativo (TARAS, 2005) e não apenas somativo, quatro dos cinco professores entrevistados afirmaram discutir o resultado da prova escrita em sala (e apenas no primeiro bimestre, já que, no segundo, os alunos não têm mais aula após a realização das avaliações), mas não da prova oral.

Márcia: *Sempre busco dar feedback. É importante para a aprendizagem do aluno, né? Só é difícil fazer isso no final do semestre, já que eles entram de férias depois da prova.*

Carlos: *Eu sempre corrijo a prova escrita com eles em sala. É bom porque faço uma revisão dos tópicos gramaticais e de vocabulário antes de continuar e dar conteúdo novo. [...] Com a prova oral, confesso que não faço.*

Isabel: *A nota já é parte do feedback, né? Mas eu acabo conversando sim com eles na aula seguinte. [...] Isso com a prova escrita. Com a prova oral não tem muito o que dar de retorno e é muito pessoal, varia de aluno pra aluno.*

Podemos observar nos excertos acima a dificuldade dos professores em incluir o aluno em suas práticas avaliativas, seja por uma questão institucional (as férias logo após a realização dos testes) ou de falta de percepção quanto à complexidade da avaliação da produção oral.

A subjetividade já se expressa a partir daqui, pois observamos uma reação em cadeia: a instituição não estabelece recomendações sobre a prova oral, o que se reflete no fato de que os professores não concebem a noção de *feedback* para os alunos quanto à prova oral, imprimindo um status de inferioridade à oralidade, ainda que o instituto de idiomas se diga adepto da abordagem comunicativa do ensino de línguas.

Essa realidade nos permite questionar, então, como são atribuídos os resultados referentes a esses testes, o que nos leva a analisar as respostas dadas à quinta pergunta do roteiro (*Quais critérios você utiliza para atribuir uma determinada nota ao aluno no teste oral?*), que revelam uma série de considerações desconexas entre os professores.

Todos os professores participantes citaram aspectos gerais para avaliar o desempenho oral dos alunos, como: gramática, vocabulário, pronúncia e, em alguns poucos casos, adequação ao tema ou conteúdo. No entanto, ao serem perguntados sobre a atribuição de notas associadas a cada critério, bem como sobre suas especificidades, os professores demonstraram insegurança e incoerência, o que nos permite concluir que, excluindo-se a organização mínima em aspectos gerais, os professores avaliam cada aspecto com excesso de subjetividade e falta de parâmetros ou referências, como confirmam os excertos seguintes:

Márcia: Os critérios que eu uso são aqueles clássicos, né? Avalio a gramática, que a gente viu em sala, a pronúncia, o quanto o aluno fala, se ele se expressa bem, se usa as palavras corretamente. [...] Não específico muito não. Distribuo a pontuação nesses critérios mesmo.

Isabel: Meus critérios são: vocabulário, pronúncia, principalmente pronúncia, na verdade. [...] Tem também gramática, que ele tem que pronunciar bem, mas usar as construções corretas. É isso. [...] Ah, dependendo do conteúdo, eu até avalio especificamente, tipo se ele usou 'phrasal verbs', mas só, eu acho.

Élida: *Meu principal critério é adequação ao que eu pedi, sabe? Porque tem aluno que viaja, que fala muito bem, tudo certinho, mas não tem nada a ver com o que a gente está discutindo. Assim como tem aluno que fala as coisas sem, assim, não tão bem enquanto pronúncia, mas fala super bem! [...] Prefiro avaliar globalmente, porque aí a gente se aproxima mais da realidade do uso da língua.*

Carlos: *Bom, como eu disse antes, meu critério é a comunicação. Quero que o aluno se expresse, interaja e use mesmo a língua. Não fico me prendendo a esses detalhes de microestrutura, não. [...] Comunicou? Falou bem? Pra mim, é o que importa!*

Jorge: *Eu avalio pronúncia (claro), gramática (menos) e vocabulário. Isso pra mim é fundamental. [...] Acho que se a gente especificar demais, a prova fica muito criteriosa e pouco comunicativa, entende?*

A fala dos professores-avaliadores é preocupante no sentido de que contribui para a falta de confiabilidade da avaliação da produção oral. Se concebemos um teste confiável como aquele que possui consistência e segurança quanto aos resultados em diferentes situações (BROWN, 2004) e que minimize a influência de fatores externos no processo (SCHOFFEN, 2009), a ausência de critérios ou a falta de consistência destes implica na possível baixa confiabilidade (ou até mesmo “não confiabilidade”) do teste aplicado pelos professores entrevistados.

Especialmente no que concerne à confiabilidade de um mesmo avaliador (*intra-rater reliability*), podemos questionar se a atribuição de notas é equilibrada entre os aspectos gerais citados pelos professores como critérios e, ainda, se esta atribuição é balanceada entre os alunos, uma vez que fatores extremamente subjetivos, como quem se concebe como “bom” aluno ou aluno “ruim”, podem afetar criticamente a confiabilidade do avaliador (BROWN, 2004).

Percebe-se, também, a crença de que o estabelecimento de critérios é um empecilho para a comunicação em situação de testes, como observado nos excertos de Élida, Carlos e Jorge. Cabe considerar, no entanto, que a natureza subjetiva da oralidade, como exposta por Tarone (1998), não implica em subjetividade de critérios, pois os critérios do professor não influenciam

(ou não deveriam influenciar) a produção do aluno, especialmente quando se trata de testes comunicativos⁷, em que a variedade de contextos para uso da língua são determinantes para a atribuição de um resultado satisfatório.

Para realizar a última pergunta do questionário (*Como você dividiria a pontuação dos seguintes critérios (uso adequado de estruturas gramaticais, vocabulário preciso e adequado ao contexto, fluência, precisão de pronúncia e adequação ao tema/ conteúdo) para alunos do nível “X”?*⁸), tivemos que separar os professores de acordo com o nível de referência em que estão inseridos os grupos para os quais lecionam.

O objetivo do questionamento seria compreender de que maneira a confiabilidade entre avaliadores (*inter-rater reliability*), aliada à subjetividade na atribuição das notas e à ausência de critérios pré-estabelecidos e/ou devidamente especificados, pode influenciar a discriminação de aprendizes pertencentes a grupos diferentes, porém de mesmo nível de referência por meio da nota obtida nos testes orais.

Na divisão do grupo de professores participantes, Élidea, Jorge e Isabel (grupo I) lecionavam, no semestre em questão, grupos de mesmo nível: Intermediário II B (o equivalente, na realidade do instituto de idiomas, ao nível B2 do Quadro Comum Europeu de Referência). Já Isabel, Márcia e Carlos (grupo II) tinham em comum grupos de nível Avançado B (ou nível C1 do QCER). O Quadro 1, a seguir, é referente às respostas dadas pelos professores do grupo I.

⁷ De acordo com Brown (1993), é característica dos testes comunicativos a exigência de uso da língua em determinado contexto. Na realidade, o autor reafirma a importância de se medir uma variedade de habilidades linguísticas que podem ser vistas como critérios, dependendo da situação de teste – como coesão e conhecimento sociolinguístico, por exemplo – considerando-se que o aluno/aprendiz deve ser avaliado levando em consideração as funções da linguagem.

⁸ Os critérios apresentados nesta questão foram elaborados após a realização das entrevistas com todos os professores. A partir do insumo dado por eles ao responderem a questão de número 5 do roteiro de entrevista, elaboramos as fichas referentes aos dois grupos e fizemos uma nova rodada de entrevistas, apenas para que fosse respondida a última questão.

Quadro 1 –Divisão de pontuação quanto a critérios de um mesmo nível (B2)

GRUPO I – INTERMEDIÁRIO II B			
	ÉLIDA	JORGE	ISABEL
Uso adequado das estruturas gramaticais	2.0	3.0	1.5
Vocabulário preciso e adequado ao contexto	3.0	2.0	3.0
Fluência	1.0	3.0	2.0
Precisão de pronúncia	3.0	1.0	2.0
Adequação ao tema/conteúdo	1.0	1.0	1.5

Fonte: Elaborado pelos autores.

Percebemos, a partir das respostas dadas à última questão, que os professores do grupo I, que lecionam para o mesmo nível de referência (B2), não apresentam uniformidade na distribuição das notas em nenhum dos aspectos. Isto implica dizer que seus testes orais, ainda que sejam aplicados para alunos de mesmo nível, suscitam realidades diferentes, pois a atribuição de notas apresenta divergências evidentes quanto ao que se pretende mensurar e quanto à importância de cada um dos critérios pré-estabelecidos.

É importante ressaltar, também, que os critérios citados pelos referidos professores na questão 5 da entrevista também não são consonantes ou uniformes, o que nos permite inferir que a discrepância entre as notas é, na realidade, ainda mais evidente.

Analisemos, agora, o quadro elaborado a partir das respostas dadas pelos professores que compõem o grupo II, que têm em comum alunos de nível C1 de referência.

Quadro 2 – Divisão de pontuação quanto a critérios de um mesmo nível (C1)

GRUPO II – AVANÇADO B			
	MÁRCIA	CARLOS	ISABEL
Uso adequado das estruturas gramaticais	2.0	1.0	1.5
Vocabulário preciso e adequado ao contexto	2.0	1.0	3.0
Fluência	2.0	5.0	2.0
Precisão de pronúncia	2.0	1.0	2.0
Adequação ao tema/conteúdo	2.0	2.0	1.5

Fonte: Elaborado pelos autores.

O quadro acima coincide com as respostas obtidas pelo grupo I, pois, neste caso, também os professores não coincidiram em nenhum dos aspectos. É interessante destacar, especialmente, que a valorização do critério “vocabulário preciso e adequado ao contexto” é bastante discrepante, assim como Carlos parece valorizar muito mais o critério “fluência” do que as outras professoras.

Uma vez mais, os dados indicam um nível alto de subjetividade na distribuição das notas em critérios, o que aponta para a falta de confiabilidade dos testes orais aplicados pelos professores em questão, uma vez que alunos de mesmo nível estariam sendo avaliados com base em atribuições e pesos significativamente diferentes para os mesmos critérios.

Consideramos, então, que a confiabilidade entre avaliadores em um mesmo contexto (*inter-rater reliability*) também seria prejudicada, pois as notas atribuídas aos alunos não seria consistente e os fatores externos (como a troca de professores, por exemplo) influenciariam no resultado obtido pelo aluno em um teste oral que se propõe a aferir produção oral em um mesmo nível de referência.

As respostas dadas pelos professores apontam não apenas para a existência da subjetividade no processo de avaliação – que é aceita e considerada pela literatura – mas, também, para um contentamento em utilizá-la, a ponto de não haver preocupação eminente em diminuir os efeitos desta na atribuição de notas. Entretanto, a subjetividade é apontada como

um fator externo que deveria ser minimizado, como asseveram Luckesi (1998) e Schoffen (2009). Observamos, portanto, com a análise dos resultados e as discussões suscitadas, que a confiabilidade e a subjetividade na avaliação da produção oral por parte dos professores parecem ser aspectos inversamente proporcionais.

Considerações Finais

Procuramos compreender, durante o desenvolvimento desta pesquisa e da conseqüente elaboração deste artigo, de que maneira os professores de inglês de um instituto de idiomas avaliam e atribuem menções/notas para classificar a produção oral de seus alunos.

Fizemos um percurso do que abordam os teóricos sobre a questão da confiabilidade e da subjetividade, em que constatamos que um teste confiável pressupõe a redução da influência de uma série de fatores externos ao teste em si, mas que podem acarretar em mudanças significativas no resultado obtido pelos alunos.

Dentre estes fatores, destacamos a importância da confiabilidade relacionada ao avaliador (*rater-reliability*), o que se refletiu na seção seguinte, em que discutimos a subjetividade por parte do avaliador desde a escolha do instrumento, passando pela própria concepção do que é avaliar, até o momento de atribuição de notas.

A partir dos apontamentos teóricos, procedemos à análise dos dados e discussão dos resultados, que corroboraram com a posição dos autores analisados no sentido de que permitiram a observação das influências que a subjetividade excessiva traz para o processo de avaliação da produção oral no referido instituto de idiomas. Subjetividade esta que é encorajada pela própria coordenação do instituto e que se reflete na fala dos professores e na ausência de critérios pré-estabelecidos para o momento de avaliar a habilidade oral, que é vista como menos prestigiosa que a avaliação da escrita.

Observamos, também, que mesmo com a adoção dos critérios citados pelos próprios professores participantes da pesquisa, no que concerne à distribuição das notas para cada critério os professores se mostraram inconsistentes, uma vez que em nenhum momento as notas atribuídas a cada critério foram equivalentes.

É importante salientar que não advogamos pela eliminação da subjetividade na avaliação, mas concordamos com os autores que afirmam a necessidade de minimizar os efeitos que esta subjetividade pode causar no teste em si.

Preocupa-nos, portanto, o fato de que os professores entrevistados confundam a evidente subjetividade, que é parte inerente do processo avaliativo, com “liberdade” e até mesmo com falta de critérios para proceder à avaliação, pois as consequências de um teste não confiável e com influência excessiva de subjetividade podem ser expressivamente negativas, quando consideramos, por exemplo, o contexto de um instituto de idiomas em que professores diferentes avaliam de maneira também diferente e com base em percepções igualmente distintas alunos que se encontram em um mesmo nível de referência.

Para que a prática avaliativa exerça de fato seu papel como “mola propulsora da aprendizagem”, é de suma importância que repensemos nossas práticas enquanto professores de línguas, a fim de compreendermos a avaliação como parte do processo de ensino-aprendizagem, não apenas com fins somativos ou como um aspecto final, um produto que deve ser analisado e pensado *a posteriori*, mas que possamos de fato concebê-la como um importante fator a contribuir para o sucesso do aprendiz no seu percurso de aprendizagem da língua inglesa.

Anexo

Roteiro de Entrevista Semiestruturada – Professores do (nome do instituto de idiomas)

1. Há quantos anos você dá aulas de língua inglesa em contexto de instituto de idiomas?
2. Como você classificaria o processo de avaliação ao qual submete seus alunos?
3. Há recomendação da instituição de ensino quanto a como se deve avaliar oralmente os alunos?
4. Em quais momentos do período/semestre a aplicação de testes ocorre?
 - 4.1. Há feedback para o aluno após a realização/correção dos testes?

5. Quais critérios você utiliza para atribuir uma determinada nota ao aluno no teste oral?
6. Como você dividiria a pontuação dos seguintes critérios (uso adequado de estruturas gramaticais, vocabulário preciso e adequado ao contexto, fluência, precisão de pronúncia e adequação ao tema/conteúdo) para alunos do nível “X”?

Referências

ALDERSON, J. C.; WALL, D. Does washback exist? *Applied Linguistics* 14, p. 115-129, 1993. (Republido de *Working Paper Series 11*, CRILE, Lancaster University).

BACHMAN, L. F. *Fundamental considerations in language testing*. Oxford: Oxford University Press, 1990.

BORGES-ALMEIDA, V. *Precisão e complexidade gramatical na avaliação de proficiência oral em inglês do formando em Letras: implicações para a validação de um teste*. 2009. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual de São Paulo, São José do Rio Preto.

BORGES-ALMEIDA, V. Com a palavra, o aluno formando em Letras: percepções sobre proficiência e avaliação da produção oral. In: BERGSLEITHNER, J. M.; WEISSHEIMER, J.; MOTA, M. B. (Org.) *Produção oral em LE: múltiplas perspectivas*. Campinas: Pontes, 2011. p. 130-152.

BROWN, H. D. *Principles of language learning and teaching*. 3. ed. New Jersey: Prentice Hall Regents, 1993.

BROWN, H. D. *Teaching by principles: An interactive approach to language pedagogy*. New Jersey: Prentice Hall Regents, 1994.

BROWN, H. D. *Language assessment: principles and classroom practice*. White Plains: Pearson Education, 2004.

BROWN, J. D.; BAILEY, K. M. A categorical instrument for scoring second language writing skills. *Language Learning*, n. 34, p. 21-42, 1984.

BYGATE, M. *Speaking*. Oxford: Oxford University Press, 1987.

BYRAM, M. *Developing the intercultural dimension in language teaching: a practical introduction for teachers*. Strasbourg: Council of Europe, 2002.

CHIZZOTTI, A. *Pesquisa qualitativa em ciências humanas e sociais*. Petrópolis: Vozes, 2006.

CONSOLO, D. A. et al. Exame de proficiência para professores de língua estrangeira (EPPL): proposta inicial e implicações para o contexto brasileiro. In: CLAFPL, 2., 2008, Rio de Janeiro. *Anais...* Rio de Janeiro: PUC-Rio, 2008.

DENZIN, N. K.; LINCOLN, Y. S. Introdução: a disciplina e a prática da pesquisa qualitativa. In: DENZIN, N. K.; LINCOLN, Y. S. (Org.) *O planejamento da pesquisa qualitativa: teorias e abordagens*. Porto Alegre: Artmed, 2006. p. 15-41.

FERRAZ, I. Instrumentos de avaliação: diversificar é preciso. In: I. E. E. (Ed.). *Pensar avaliação, melhorar a aprendizagem*. Lisboa: IEE, 1994.

GERGEN, M. M.; GERGEN, K. J. Investigação qualitativa: tensões e transformações. In: DENZIN, N. K.; LINCOLN, Y. S. (Org.) *O planejamento da pesquisa qualitativa: teorias e abordagens*. Porto Alegre: Artmed, 2006. p. 367-388.

LUCKESI, C. C. Avaliação educacional: pressupostos conceituais. *Tecnologia Educacional*, Rio de Janeiro, p. 5-8, 1978.

LUCKESI, C. C. *Avaliação da aprendizagem escolar*. 7. ed. São Paulo: Cortez, 1998.

LUOMA, S. *Assessing speaking*. Cambridge: Cambridge University Press, 2004.

NATIONAL RESEARCH COUNCIL. *National Science Education Standards*. Washington: National Academy Press, 1996.

NUNAN, D. *Research methods in language learning*. Cambridge: Cambridge University Press, 1992.

PERRENOUD, P. *Práticas pedagógicas, profissão docente e formação: perspectivas sociológicas*. Lisboa: Dom Quixote, 1993.

QUEVEDO-CAMARGO, G. Proficiência linguística e avaliação da produção oral em língua inglesa na educação básica. In: EL-KADRI, M. S.; PASSONI, T. P.; GAMERO, R. (Org.). *Tendências contemporâneas para o ensino de língua inglesa: propostas didáticas para a educação básica*. v. 7. Campinas: Pontes, 2014. p. 245-272.

ROSA, M. V. F. P. C.; ARNOLDI, M. A. G. C. *A entrevista na pesquisa qualitativa*. Belo Horizonte: Autêntica, 2006.

SILVEIRA, R. M. H. A entrevista na pesquisa em educação: uma arena de significados. In: COSTA, M. V. (Org.) *Caminhos investigativos II: outros modos de fazer pesquisa em educação*. Rio de Janeiro: DP&A, 2002.

SCARAMUCCI, M. V. R. Avaliação: mecanismo propulsor no ensino/aprendizagem de língua estrangeira. *Contexturas*, São José do Rio Preto, v. 4, p. 115-124, 1998.

SCARAMUCCI, M. V. R. Proficiência em LE: considerações terminológicas e conceituais. *Trabalhos em Lingüística Aplicada*, v. 36, p. 11-22, 2000.

SCHOFFEN, J. R. *Gêneros do discurso e parâmetros de avaliação de proficiência em português como língua estrangeira no exame Celpe-Bras*. 2009. Tese (Doutorado em Linguística Aplicada) – Universidade Federal do Rio Grande do Sul, Porto Alegre.

TARAS, M. Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*. v. 53. Sunderland: Blackwell Publishing, 2005.

TARONE, E. Interlanguage variation: implications for language testing. In: BACHMAN, L. F.; COHEN, A. D. *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press, 1998.

XAVIER, R. P. Avaliação diagnóstica e aprendizagem. *Contexturas*, São José do Rio Preto, v. 4, p. 115-124, 1998.

Recebido em: 14/04/2016

Aceito em: 21/08/2016