

O LETRAMENTO EM AVALIAÇÃO NO ENSINO DE INGLÊS AERONÁUTICO: RELEVÂNCIA, NECESSIDADES E POSSIBILIDADES

Language Assessment Literacy in the Teaching of Aeronautical English

Paula Ribeiro e SOUZA
Instituto de Controle do Espaço Aéreo
paulaprs1@fab.mil.br
<https://orcid.org/000-0001-7202-531X>

RESUMO: O objetivo deste artigo é apresentar os resultados de uma investigação sobre as necessidades de letramento em avaliação de línguas (LAL) de professores de inglês aeronáutico e seus níveis percebidos de conhecimento sobre a área. Para isso, um questionário foi elaborado e enviado aos professores de inglês aeronáutico que atuavam nos cursos de línguas dos programas de formação em controle de tráfego aéreo no Brasil. Trinta e oito profissionais responderam ao questionário. Os dados foram analisados por meio da estatística descritiva a fim de identificar lacunas e necessidades prioritárias. Testes não-paramétricos também foram utilizados para levantar diferenças nas respostas de dois perfis distintos de professores: profissionais da carreira de magistério em língua inglesa e controladores de tráfego aéreo com formação em ensino de inglês aeronáutico. Os resultados desta análise visam informar o desenvolvimento de um programa de desenvolvimento profissional em LAL, voltado para professores de inglês aeronáutico, quanto às áreas que devem ser priorizadas, considerando as especificidades das práticas educacionais locais.

PALAVRAS-CHAVE: Letramento em avaliação de línguas; Inglês aeronáutico, Análise de necessidades.

ABSTRACT: The aim of this paper is to show the results of an investigation conducted to identify the language assessment literacy (LAL) needs of aeronautical English teachers, as well as their perceived level of knowledge in this area. To that end, a questionnaire was designed and delivered to the aeronautical English teachers who worked at pre-service air traffic control training programs in Brazil. Thirty-eight professionals answered the questionnaire, and their responses were first analyzed through descriptive statistics to identify lacks and urgent LAL training needs. Then, non-parametric tests were conducted to check the differences in the way two different professionals' profiles answered the questions: regular English teachers and air traffic controllers who received language teaching training. The findings of this study will inform the development of a language assessment training program for aeronautical English teachers in terms of urgent areas to be covered, considering the specificities of the local educational practices.

KEYWORDS: Language assessment literacy; Aeronautical English, Teacher training programs.

INTRODUÇÃO

O uso exponencial de exames nas últimas décadas, principalmente, como forma de elevar os padrões educacionais, com consequências graves para os envolvidos (Spolsky, 2008), tem demandado um entendimento mais profundo, por parte de professores e outros envolvidos no processo de avaliação de línguas, das questões teóricas, práticas e sociais que permeiam o desenvolvimento e uso de testes (Fulcher, 2012; Taylor, 2009, 2013; Scaramucci, 2016; Quevedo-Camargo; Scaramucci, 2018). Nesse sentido, Fulcher (2012) é incisivo quando diz que os professores devem tomar consciência do poder investido nos testes e sobre como eles afetam suas práticas. Scaramucci (2016) reforça esse posicionamento ao afirmar que:

para explorar o potencial que exames externos poderiam trazer para as práticas na escola, o professor necessita estar capacitado para analisar profundamente esses exames e avaliar os resultados de seus alunos, o que pressupõe entendimento das matrizes de referência em que se fundamentam e da relação que elas têm com o currículo e programas de ensino (Scaramucci, 2016, p.157).

No ensino de inglês para fins ocupacionais como é o caso do inglês aeronáutico para controladores de tráfego aéreo, a introdução do Exame de Proficiência em Inglês Aeronáutico do Sistema de Controle do Espaço Aéreo (EPLIS), ocorrida em 2014, nos cursos de formação profissional, também se desenhou como uma tentativa, por parte das autoridades do Controle do Espaço Aéreo Brasileiro, de aumentar a relevância do exame nesses contextos e provocar, assim, mudanças educacionais (Souza, 2019, 2020).

O estudo de Souza (2020), que analisou os efeitos do EPLIS nos professores e alunos de uma escola de formação de especialistas em controle de tráfego aéreo, demonstrou que, de fato, situações relacionadas aos professores de inglês aeronáutico, como a falta de uma compreensão mais profunda acerca do EPLIS e, mais especificamente, do seu propósito, conteúdo, características e tarefas, fizeram com que a influência do exame sobre as práticas pedagógicas fosse limitada e, por vezes, negativa.

Diante desse cenário, é bastante válida a crítica que Fulcher (2012) e Scaramucci (2016) fazem à visão dos responsáveis pelas políticas de que mudanças curriculares são facilmente implementadas com a adoção de um exame externo, sem que haja, para isso, investimento em programas de desenvolvimento profissional para os professores.

Este artigo tem como propósito apresentar uma análise das necessidades de

letramento em avaliação, conduzida com os professores de inglês aeronáutico que atuam na formação de controladores de tráfego aéreo, a fim de oferecer subsídios para o desenvolvimento de um treinamento em avaliação para esses profissionais.

LETRAMENTO EM AVALIAÇÃO: DEFINIÇÕES E ESTUDOS ANTERIORES

Em 1991, frente a uma crescente obsessão da sociedade americana por testes padronizados para medir e certificar o desempenho de escolas e alunos, Richard Stiggins escreve sobre a necessidade latente de “letramento em avaliação” tanto para aqueles que geram dados baseados em avaliações, ou seja, os elaboradores de exames, quanto para aqueles que consomem esses dados: professores, pais, alunos, diretores, coordenadores e educadores profissionais (Stiggins, 1991; Quevedo-Camargo; Scaramucci, 2018). Para Stiggins (1991), o cerne do “letramento em avaliação” estaria em promover uma compreensão profunda da avaliação para que esses profissionais pudessem julgar a qualidade de um teste, interpretar seus resultados, bem como propor melhorias no processo avaliativo.

Assim como Stiggins (1991), Inbar-Lourie (2008a, p. 389) caracteriza o “letramento em avaliação” como a “habilidade de compreender, analisar e aplicar informações sobre o desempenho do aluno para a melhoria da instrução”¹ ou para a tomada de decisões de forma informada e consciente de suas consequências. Por sua vez, o “letramento em avaliação em contextos de línguas” ou “letramento em avaliação de línguas” engloba, além das habilidades requeridas para o “letramento em avaliação”, um componente linguístico e cultural, que inclui teorias sobre aquisição de línguas, ensino e aprendizagem, uso da língua em diversos contextos (Inbar-Lourie, 2008a, 2013), e discute construtos como visão de língua, proficiência, e língua estrangeira (Scaramucci, 2016).

O baixo nível de letramento em avaliação parece não ser uma característica de países como o Brasil, e atinge também nações mais ricas (Taylor, 2013; Tsagari; Vogt, 2017). Para Taylor (2013), a pouca disseminação da área se deve ao fato de que muitos educadores percebem a avaliação de línguas como altamente técnica e especializada. No contexto brasileiro, Scaramucci (2016) pontua que o nível de letramento em avaliação,

¹ No original: “the ability to understand, analyze and apply information on student performance to improve instruction”.

em geral e em contextos de línguas, pode ser considerado pouco ou inexistente. Até mesmo no campo da pesquisa em Linguística Aplicada, Scaramucci (2006) encontra deficiências no embasamento teórico de estudos que discutem questões pertinentes à avaliação, demonstrando desconhecimento por parte de seus autores sobre a mudança de paradigma pela qual a área da avaliação passou nos últimos anos, vista, atualmente como uma prática social, permeada de valores sociais e culturais (Shohamy, 1993). Para Scaramucci (2016), o problema reside também na pouca importância dada à avaliação nos cursos de formação de professores em geral e de Letras.

As consequências da falta ou do pouco letramento em avaliação de línguas podem ser facilmente percebidas nas salas de aulas, onde a avaliação é essencialmente “somativa, classificatória, de produto, mecanicista, autoritária e excludente (Scaramucci, 2016, p. 154), e ignora o seu papel de fornecer informações sobre a aprendizagem a fim de identificar falhas e corrigi-las.

Para que a “cultura da testagem”, na qual a aprendizagem é vista como o acúmulo de conhecimentos e a avaliação tem o objetivo exclusivo de testar o que foi aprendido, dê lugar à “cultura da avaliação”, na qual avaliação e a aprendizagem estão interligadas e avaliar é uma forma não apenas de monitorar a aprendizagem, mas de promovê-la, precisamos oferecer aos envolvidos, principalmente, aos professores, as ferramentas necessárias na área de letramento em avaliação (Scaramucci, 2006; Inbar-Lourie, 2008b; Scarino, 2013).

Considerando que o “letramento em avaliação” e o “letramento em avaliação de línguas” são conceitos bastante recentes e em construção, ainda não se estabeleceu um consenso sobre quais seriam os conhecimentos, habilidades e competências necessários para os diversos envolvidos (protagonistas) nas práticas de avaliação e se haveria componentes essenciais ou uma base comum que todos deveriam possuir. Por outro lado, Inbar-Lourie (2013) assegura já haver certa concordância de que o “letramento em avaliação em contextos de línguas” é um conceito multidimensional, que envolve diversas camadas e níveis de competência, fazendo com que sua profundidade e extensão dependam das circunstâncias contextuais e das funções e responsabilidades dos protagonistas envolvidos.

Em um esforço para mapear as necessidades de professores de línguas em relação à área de avaliação para a elaboração de cursos e materiais didáticos, Fulcher (2012)

analisou as respostas de 278 professores de várias regiões do mundo a um questionário. Os resultados apontaram a necessidade de integrar princípios e conceitos a uma abordagem procedimental (como fazer) na elaboração, desenvolvimento, manutenção e aplicação de exames em larga escala e de sala de aula, incluindo a promoção da conscientização sobre suas origens, razões e impactos.

Taylor (2013), baseando-se nos artigos publicados em um volume do periódico *Language Testing* dedicado a discutir o “letramento em avaliação de línguas”, propôs oito principais dimensões para descrever esse construto: 1) conhecimento teórico, 2) habilidades técnicas, 3) princípios e conceitos, 4) ensino de línguas, 5) valores socioculturais, 6) práticas locais, 7) crenças e atitudes individuais, e 8) resultados e tomada de decisões. Segundo Taylor (2013), haveria diferentes níveis e perfis para atender às necessidades específicas de “letramento em avaliação de línguas” dos envolvidos. No caso dos professores, foco deste artigo, Taylor (2013) recomendou um nível avançado de conhecimento sobre ensino de línguas, um nível intermediário de habilidades técnicas e de conhecimento das práticas locais, valores socioculturais, crenças e atitudes individuais. Por sua vez, conhecimento teórico, princípios e conceitos, bem como resultados e tomada de decisão não precisariam ser explorados a fundo para esse grupo.

Com o propósito de investigar a usabilidade do modelo de Taylor (2013), Kremmel e Harding (2020) conduziram um estudo com mais de 1000 participantes, contemplando diversos grupos de protagonistas envolvidos: professores de línguas, desenvolvedores de exames, pesquisadores, examinadores, examinandos, usuários de resultados, tomadores de decisões e pais ou responsáveis legais. Na percepção dos professores participantes, suas necessidades de treinamento em avaliação seriam atendidas com um bom nível de conhecimento acerca dos seguintes assuntos: teorias sobre a estrutura, uso e desenvolvimento da língua/linguagem; princípios da avaliação e interpretação de resultados; avaliação no ensino de línguas; impacto da avaliação e preparação dos alunos; valores e crenças individuais; e atribuição de notas. Aspectos como elaboração e administração de exames, estatística e métodos de pesquisa, políticas de avaliação, e práticas locais receberam grau menor de relevância, porém eram ainda vistos como necessários.

Partindo, então, do pressuposto de que as necessidades de letramento e os níveis

de competência variam conforme a função e responsabilidade dos protagonistas envolvidos e são dependentes das características dos contextos, o presente estudo busca analisar a percepção dos professores de inglês aeronáutico que atuam nos cursos de formação de controladores de tráfego aéreo sobre seu nível de letramento em avaliação e suas necessidades específicas de treinamento. Duas perguntas de pesquisa guiaram a coleta e análise dos dados:

- 1) Como os professores de inglês aeronáutico percebem seu nível de letramento em avaliação de línguas e suas necessidades de treinamento?
- 2) Há diferenças nas percepções desses professores considerando seus perfis profissionais?

O CONTEXTO DA PESQUISA

O uso do inglês como língua comum na comunicação aeronáutica data de 1944, quando foi firmada a Convenção da Aviação Civil Internacional. No entanto, somente a partir de 2003, exames de proficiência em língua inglesa passaram a ser aplicados a pilotos, controladores de tráfego aéreo (controladores, doravante) e operadores de estação aeronáutica, em diversos países, para atender às novas provisões da Organização da Aviação Civil Internacional² (OACI) quanto ao nível mínimo necessário para a operação de voos civis internacionais.

Em 2007, o Instituto de Controle do Espaço Aéreo (ICEA) desenvolveu o Exame de Proficiência em Inglês Aeronáutico do Sistema de Controle do Espaço Aéreo (EPLIS), para avaliar o nível de proficiência dos controladores e operadores de estação aeronáutica em serviço no Brasil. Em 2014, o público-alvo do EPLIS passou a englobar os controladores em pré-serviço, ou seja, que ainda estavam em processo de capacitação inicial. A implementação desse exame no contexto escolar é vista, segundo Souza (2019), como uma tentativa de elevar os níveis de proficiência desses futuros profissionais, ao estimular práticas pedagógicas mais condizentes com as necessidades de uso real da língua pelo futuro profissional.

Considerado um exame de desempenho com propósitos específicos, uma vez que

² A Organização da Aviação Civil Internacional, ICAO ou OACI em português, é uma agência regulatória da ONU, responsável por estabelecer os padrões e regulamentos necessários para a segurança, regularidade e eficiência da aviação civil mundial. Atualmente, é formada por 193 países.

os conhecimentos da área de controle de tráfego aéreo interagem com os conhecimentos linguísticos em uma situação comunicacional (SCARAMUCCI *et al.*, 2010), o EPLIS é composto por duas fases, a primeira fase é realizada a distância e tem como objetivo avaliar a compreensão oral do aluno em situações autênticas de gerenciamento do tráfego aéreo, por meio de 30 questões de múltipla escolha. A segunda fase consiste em uma entrevista oral, com foco na avaliação das habilidades de compreensão e produção de forma integrada, e está dividida em quatro partes. Na primeira parte, são colocadas ao examinando quatro perguntas sobre rotina profissional, experiência e carreira. Na segunda parte, são apresentadas 10 situações de tráfego aéreo incomuns, inesperadas ou emergenciais. É solicitado ao examinando que explique cada situação e sugira uma solução. Na parte 3, o examinando deve responder a quatro perguntas sobre controle de tráfego aéreo e aviação, emitindo sua opinião e justificando-a. Na última parte, uma imagem é apresentada ao examinando que deve, primeiramente, descrevê-la e, em seguida, contar uma história sobre a situação ilustrada. Para atribuição de notas, é utilizada uma escala com seis níveis de proficiência publicada pela OACI.

O curso básico de controle de tráfego aéreo tem duração de 2 anos e, no final do último ano, é aplicado o EPLIS a todos os alunos obrigatoriamente. Souza (2019, 2020) explica que os resultados obtidos pelos alunos são considerados oficiais e passam a constar em seus registros profissionais, assim que concluem o curso de formação. Porém, o desempenho insuficiente no exame não os impede de se formarem, nem de prosseguirem para o estágio *in loco* nos órgãos de controle de tráfego aéreo. No entanto, existe um estudo em andamento sobre a possibilidade de que o desempenho no EPLIS seja considerado como parte integrante da nota na disciplina de língua inglesa ministrada no 4º semestre do curso, o que aumentaria a relevância do exame no contexto escolar.

O ensino de inglês aeronáutico é parte do currículo do curso de formação, com 4 disciplinas, uma a cada semestre, correspondendo a uma carga horária total de cerca de 400 horas-aula. Os professores que ministram essas disciplinas podem ser divididos em duas categorias: a) profissionais da carreira de Magistério em Língua Inglesa que fazem parte do corpo docente permanente do curso, e b) controladores com alta proficiência em inglês e que possuem, no mínimo, certificado de ensino de inglês aeronáutico, equivalente a um *TEFL* (*Teaching English as a Foreign Language*). Esse segundo grupo de profissionais, neste estudo denominados controladores/instrutores, pertencem ao efetivo

de outras instituições, mas atuam no curso de formação para complementar o quadro docente. As duas categorias de professores de inglês aeronáutico foram convidadas a participar desta pesquisa.

Além de ter que lidar com o impacto que uma avaliação externa como o EPLIS gera nas práticas de sala de aula, os professores de inglês aeronáutico também se envolvem com atividades avaliativas somativas escritas como a elaboração e aplicação de exames de rendimento escritos, avaliações formativas como a elaboração e aplicação de provas de produção oral, além de avaliações informais. É inegável, portanto, a necessidade de promover o letramento em avaliação de línguas desses profissionais, dado que desempenham duplo papel como facilitadores e avaliadores do processo de ensino e aprendizagem (Scaramucci, 2006; Inbar-Lourie, 2008b; Scarino, 2013). Como argumentam Quevedo-Camargo e Scaramucci (2018, p.239), “a proposta de tornar o professor de línguas letrado em avaliação tem como meta principal fazer dele um consumidor e produtor crítico de avaliações”, capaz de transformar a cultura de avaliar no seu contexto. Infelizmente, tanto no contexto pesquisado (Souza, 2020) quanto no Brasil como um todo (Scaramucci, 2006), ainda é predominante a visão de avaliação como mera verificação do que foi aprendido, que ignora o seu poder de promoção da aprendizagem.

METODOLOGIA

Para avaliar o letramento em avaliação de línguas dos professores de inglês aeronáutico, optou-se por elaborar um questionário utilizando o instrumento de Fulcher (2012) como referência, por esse ter focado somente nos professores como grupo de protagonistas envolvidos. O questionário de Fulcher (2012) apresentava questões fechadas e abertas. Visando, porém, em um primeiro momento, coletar dados somente quantitativos, optou-se por um questionário com questões fechadas, ou seja, em que uma ou mais respostas são selecionadas. Outra fonte de referência para a elaboração do questionário foi o livro de Bachman e Damböck (2017) sobre avaliação na sala de aula.

O questionário elaborado foi composto por 86 itens, divididos em 5 partes (Apêndice A). O quadro 01 apresenta os temas explorados em cada parte do questionário, tipo de informações coletadas, e formato dos itens.

Na primeira parte do questionário (10 perguntas), foram coletados dados profissionais dos participantes relevantes para o propósito do estudo, e.g. perfil profissional, experiência profissional, tempo de atuação como professor de inglês aeronáutico, experiência como examinador de testes de proficiência, experiência como elaborador de itens de prova, e grau de escolaridade referente ao ensino de línguas.

A segunda parte do questionário, ‘Conhecimentos em Avaliação de Línguas’, focou em identificar o nível atual de letramento em avaliação dos participantes. Foi perguntado, primeiramente, se eles haviam recebido treinamento em avaliação nos cursos de ensino de línguas dos quais participaram e, então, eles foram instruídos a avaliar o nível de conhecimento geral que possuíam na área de avaliação, utilizando, para isso, uma escala *likert* de 5 pontos, sendo 1 “não tenho conhecimento”, 2 “pouco”, 3 “razoável”, 4 “bom” e 5 “excelente”. Em seguida, foi utilizada a mesma escala para que os participantes avaliassem o nível atual de conhecimento em 26 diferentes aspectos da avaliação de línguas, como, por exemplo, avaliação de proficiência, avaliação da compreensão oral, avaliação da produção oral, exames de alta-relevância, impacto e efeito retroativo, entre outros.

Quadro 1 – A estrutura do questionário

Temas	Conteúdos	Formato dos itens
Parte 1 – Dados profissionais	Experiência no ensino de inglês aeronáutico; experiência como controlador; experiência como examinador EPLIS; experiência como elaborador de itens de prova; nível de formação na área de ensino de línguas.	Itens de sim ou não, itens de múltipla escolha ordenado em anos e em níveis de formação.
Parte 2 – Conhecimentos em avaliação de línguas	Disciplinas com foco em avaliação de línguas; nível geral de conhecimento em avaliação de línguas; e nível de conhecimento em assuntos específicos e áreas a serem exploradas: avaliação de nivelamento, diagnóstica, rendimento, proficiência, somativa, formativa e alternativa; avaliação das habilidades de compreensão oral, produção oral, leitura, produção escrita, gramática e vocabulário; habilidades integradas; autoavaliação e avaliação em pares; validade, confiabilidade, impacto e efeito retroativo; ética e justiça; elaboração de testes; escalas de proficiência; e resultados de exames e estatística para testes.	Itens de sim ou não, itens de múltipla escolha em escala <i>likert</i> de nível de conhecimento e em escala <i>likert</i> de grau de interesse, ambas as escalas com 5 pontos.
Parte 3 – Métodos	Tipos de tarefas avaliativas presentes nos	Itens de múltipla escolha

de avaliação de línguas	contextos de ensino: múltipla escolha, perguntas com respostas escritas curtas, verdadeiro ou falso, preenchimento de lacunas, ordenação, entrevista individual ou em grupos, simulações, descrição oral de figuras, discussão oral, apresentação oral.	em escala <i>likert</i> de frequência com 5 pontos.
Parte 4 - Usos e propósitos da avaliação de línguas	Aperfeiçoamento da instrução e aprendizagem; implementação de mudanças curriculares; classificação dos alunos em turmas ou níveis; identificação de áreas a serem priorizadas; identificação do quanto foi aprendido; necessidade de instrução complementar; oferta de <i>feedback</i> sobre o progresso; necessidade de retomada dos objetivos de aprendizagem; seleção de alunos que seguirão para o próximo nível; e emissão de certificado de proficiência.	Itens de múltipla escolha em escala <i>likert</i> de frequência com 5 pontos.

Fonte a autora (2024)

Ainda na parte 2 do questionário, os respondentes avaliaram as áreas específicas de avaliação que gostariam que fossem exploradas em um possível programa de desenvolvimento profissional. Para isso, foram utilizados os mesmos 26 aspectos tratados anteriormente em uma escala de 1 a 5, sendo 1 “definitivamente não”, 2 “provavelmente não”, 3 “indiferente”, 4 “provavelmente sim” e 5 “definitivamente sim”.

A parte 3 (12 questões) do questionário teve como foco identificar a frequência com que diversos métodos de avaliação eram utilizados no contexto de ensino desses profissionais, tais como: múltipla escolha, perguntas com respostas escritas curtas, verdadeiro ou falso, entrevista oral, entre outros. Já, na parte 4 (10 questões), o objetivo foi identificar a frequência com que diversos propósitos de avaliação eram atendidos. Por exemplo, foi pedido aos participantes que indicassem a frequência com que as avaliações que conduziam tinham como propósito aperfeiçoar a instrução e a aprendizagem. Para ambas as partes, foi utilizada uma escala *likert*, com 5 pontos, sendo 1 “nunca”, 2 “raramente”, 3 “ocasionalmente”, 4 “frequentemente” e 5 “sempre”.

Embora o questionário de Fulcher (2012) estivesse em inglês, o questionário deste estudo foi formulado em português, a fim de reduzir a chance de que a língua pudesse interferir no entendimento dos itens. A primeira versão do questionário passou por uma revisão de quatro especialistas em avaliação, dos quais um possuía doutorado na área e os demais, mestrado. Foi pedido a eles que respondessem ao questionário e fizessem sugestões para que os itens ficassem mais claros e pertinentes. Os termos traduzidos do inglês foram revistos. Discutiu-se também sobre a possibilidade de o respondente não ter

conhecimento da metalinguagem utilizada no questionário. Considerou-se que a metalinguagem é parte inerente da profissão e que o seu desconhecimento indica, na verdade, nível de letramento insuficiente para uma visão crítica e informada do assunto.

O questionário foi disponibilizado, então, na plataforma *Google Forms*, e uma mensagem eletrônica foi enviada aos 115 professores de inglês aeronáutico que atuam nos cursos de formação e pós-formação de controle de tráfego aéreo, cadastrados junto ao Instituto de Controle do Espaço Aéreo, para convidá-los a participar do estudo. O *link* para o questionário ficou ativo por dois meses e foram recebidas 38 respostas. No entanto, dois participantes indicaram que ainda não ministravam aulas de inglês para controladores e suas respostas foram, portanto, desconsideradas. Por fim, foram analisadas as respostas de 36 participantes, o que representou uma taxa de retorno de 31%.

Para identificar as tendências gerais de letramento em avaliação dos participantes e possíveis diferenças entre diferentes perfis de profissionais foram utilizadas a estatística descritiva e testes estatísticos não-paramétricos, uma vez que a amostra de participantes era razoavelmente pequena ($n=36$). Foi também calculada a consistência interna dos itens redigidos em escala *likert* das partes 2 a 5, e o questionário foi considerado altamente confiável, apresentando Alfa de Cronbach em 0,95.

ANÁLISE E DISCUSSÃO DOS DADOS

Perfil profissional

Sobre o perfil profissional dos participantes, 25 são controladores (68%) e 11 são da carreira de Magistério de Língua Inglesa (32%). Essa distribuição é condizente com a situação atual, dado que os controladores que recebem capacitação no ensino de inglês aeronáutico são mais numerosos. Esses profissionais têm como principal função capacitar seus colegas de profissão nos locais em que atuam, mas também são, regularmente, chamados a complementar o quadro docente do curso de formação, em períodos específicos do calendário escolar. Já, os profissionais do Magistério de Língua Inglesa que participaram desta pesquisa atuam, com uma única exceção, exclusivamente no curso de formação de controle de tráfego aéreo.

Em termos de experiência no ensino de inglês aeronáutico, somente oito profissionais (22%) afirmaram ter dois ou menos anos de atuação. Porém, desses profissionais menos experientes, seis (75%) são da carreira de magistério de língua

inglesa e atuam somente no curso de formação inicial. Isso significa que metade do corpo docente permanente é composto por profissionais iniciantes no ensino de inglês aeronáutico. Por outro lado, seis dos 11 profissionais da carreira de magistério (56%) possuem nível de escolaridade acima da graduação, sendo dois com especialização, três com mestrado; e um com doutorado. No grupo dos controladores/instrutores, 17 (68%) possuem apenas o certificado de ensino de inglês aeronáutico. Dos oito restantes (32%), quatro possuem graduação na área, três têm especialização e um concluiu o mestrado.

Foi levantado, também, se existiam no grupo de participantes profissionais que atuavam como examinadores do EPLIS, e nove responderam positivamente, ou seja, (25%). Desses, nenhum pertencia ao quadro de docentes permanentes do curso de formação. Quanto à pergunta sobre elaboração de itens de provas ou simulados, 26 dos 36 participantes (72%) sinalizaram que realizavam essa atividade. Em relação ao grupo de professores da carreira de magistério, 9 de 11 (81%) responderam atuar como elaboradores de questões. Essa proporção foi um pouco menor no grupo dos controladores/instrutores, com 17 dos 25 respondendo afirmativamente (68%).

Podemos dizer que o perfil dos dois grupos se distingue razoavelmente. Os profissionais da carreira de magistério parecem menos experientes no ensino de inglês aeronáutico, porém possuem maior nível de escolaridade na área de ensino de línguas e reportam estar mais envolvidos com elaboração de itens. Os controladores/instrutores, por outro lado, são mais experientes no ensino de inglês, porém possuem nível menor de formação na área de ensino de línguas e estão menos envolvidos com elaboração de questões de prova, embora alguns sejam aplicadores do EPLIS.

Pelo fato de o primeiro grupo atuar majoritariamente (90%) somente no curso de formação de controladores, enquanto o segundo grupo atua também em contextos de pós-formação, tomou-se a decisão de identificar as necessidades em avaliação dos participantes não apenas como um todo, mas também comparativamente entre essas duas categorias de profissionais.

Nível de conhecimento em avaliação de línguas

Os participantes da pesquisa, de modo geral, avaliaram o seu nível de letramento em avaliação de línguas (LAL) conforme mostra a Tabela 1. Podemos observar que a média das respostas foi de 3,17, dentro de uma faixa de 1 a 5, ou seja, os participantes, de

modo geral, percebem o seu nível de conhecimento em avaliação próximo do critério “razoável” do questionário. Porém, quando perguntados se tiveram disciplinas voltadas para a avaliação nos cursos de ensino de línguas realizados, mais da metade (52,8%) respondeu negativamente. A princípio, esse dado sugere que o conhecimento em avaliação de línguas para muitos profissionais é adquirido no exercício de sua profissão, provavelmente, na interação com outros colegas de profissão ou coordenadores pedagógicos, ou ainda, por meio da iniciativa individual de buscar informações sobre o assunto.

Vogt e Tsagari (2014) chegaram a uma percepção semelhante ao entrevistar professores de línguas no contexto europeu. Ao serem questionados por Vogt e Tsagari sobre programas de desenvolvimento profissional, os participantes desse estudo disseram ter recebido pouco ou nenhum treinamento em avaliação nesses programas e para compensar essa lacuna adotavam práticas de avaliação de colegas ou mentores educacionais.

Tabela 1 – Nível percebido de LAL dos professores de inglês aeronáutico

	N	Média	Desvio Padrão
Nível percebido de LAL	36	3,17	1,056

Fonte: autora (2024)

Ainda sobre a Tabela 1, quando observamos o desvio padrão - índice de dispersão dos dados em relação a média, vemos que esse valor ficou acima de 1 ponto (DP=1,056), indicando que as respostas dos participantes não foram muito homogêneas. Na Tabela 2, ao comparar as médias dos dois perfis profissionais, carreira de magistério (Grupo 1) e controladores/instrutores (Grupo 2), percebemos que as médias se diferem em mais de 1 ponto, com os professores do magistério atribuindo um nível maior de conhecimento na área de avaliação de línguas (M=3,91), em relação aos controladores/instrutores (M=2,84).

Tabela 2 – Nível percebido de LAL por perfil profissional

		N	Média	Desvio Padrão
Nível percebido de LAL	Grupo 1	11	3,91	0,701
	Grupo 2	25	2,84	1,028

Fonte: autora (2024)

Em relação ao conhecimento de aspectos específicos da avaliação de línguas, a média geral das respostas foi de 3,15, bastante semelhante à percepção dos participantes sobre seu nível de conhecimento em avaliação de línguas. No entanto, as médias dos grupos se diferem mais uma vez, com o grupo de professores do magistério apresentando média geral de 3,54, enquanto o grupo de controladores/instrutores atingiu média geral de 2,50, indicando novamente uma diferença em como esses dois perfis profissionais percebem o seu nível de letramento em avaliação em assuntos específicos.

Na Tabela 3, propusemos uma comparação das médias dos dois grupos nos 26 aspectos de avaliação de línguas pesquisados. Utilizamos o teste não paramétrico U de Mann Whitney para amostras independentes, com significância de 0,05 ($p < 0,05$), a fim de identificar se a diferença entre essas médias apresentava significância estatística, ou seja, se essa diferença era realmente relevante em termos estatísticos ou não.

Tabela 3 – Comparação do nível de conhecimento em aspectos do LAL
Teste U de Mann Whitney para amostras independentes

Aspectos da AL	Médias	Sig.	Decisão
Avaliação de nivelamento	3,18 - 2,40	115,00	Reter Ho
Avaliação diagnóstica	3,64 - 2,64	38,000	Rejeitar Ho
Avaliação de rendimento	3,64 - 2,44	8,000	Rejeitar Ho
Avaliação de proficiência	3,91 - 2,84	7,000	Rejeitar Ho
Avaliação somativa	3,64 - 2,64	63,000	Reter Ho
Avaliação formativa	3,55 - 2,56	45,000	Rejeitar Ho
Avaliação alternativa	3,09 - 1,88	8,000	Rejeitar Ho
Avaliação de desempenho	3,55 - 2,68	74,000	Reter Ho
Avaliação de compreensão oral	4,00 - 3,04	24,000	Rejeitar Ho
Avaliação de produção oral	4,00 - 2,92	13,000	Rejeitar Ho
Avaliação de leitura	4,00 - 2,60	1,000	Rejeitar Ho
Avaliação de produção escrita	3,91 - 2,40	1,000	Rejeitar Ho
Avaliação de gramática e vocabulário	4,00 - 2,76	1,000	Rejeitar Ho
Avaliação de habilidades integradas	3,64 - 2,56	12,000	Rejeitar Ho
Autoavaliação e avaliação em pares	3,45 - 2,36	11,000	Rejeitar Ho
Validade	3,27 - 2,20	22,000	Rejeitar Ho
Confiabilidade	3,45 - 2,44	29,000	Rejeitar Ho
Impacto e efeito retroativo	3,00 - 1,96	15,000	Rejeitar Ho
Ética e justiça na avaliação	3,73 - 2,68	38,000	Rejeitar Ho
Exames de alta-relevância	3,36 - 2,04	1,000	Rejeitar Ho
Especificações de exames	3,18 - 1,88	1,000	Rejeitar Ho
Métodos de avaliação	3,55 - 2,68	11,000	Rejeitar Ho
Elaboração de testes e provas	3,73 - 2,72	22,000	Rejeitar Ho
Escala de níveis de proficiência	3,55 - 3,36	612,000	Reter Ho
Interpretação de resultados de provas	3,18 - 2,48	93,000	Reter Ho
Estatística para avaliação	2,82 - 1,96	35,000	Rejeitar Ho

Fonte: autora (2024)

Podemos observar, na coluna 2, as médias dos dois grupos para cada aspecto do LAL listado, sendo a da esquerda referente ao grupo da carreira de magistério e a da direita, ao grupo de controladores/instrutores. Nota-se que as médias do primeiro grupo são maiores do que as do segundo em todas as categorias. Com exceção do aspecto ‘Estatística para Avaliação’, as médias dos professores variaram entre $M=3,18$ e $M=4,00$, ou seja, entre os níveis “razoável” e “bom” do questionário. Já os controladores tiveram a maior média calculada em 3,04 e a menor em 1,88, ou seja, entre os níveis “razoável” e “nenhum”. Essas diferenças fazem sentido ao considerarmos que o grupo dos profissionais do magistério possui maior nível de escolaridade na área de ensino de línguas e estão mais envolvidos com práticas avaliativas no contexto em que atuam.

Sobre isso, percebe-se que o primeiro grupo julga ter maior conhecimento na área de avaliação das habilidades de compreensão oral, produção oral, leitura e de gramática e vocabulário, enquanto as maiores médias dos controladores/instrutores são em escalas de níveis de proficiência, avaliação da compreensão oral e produção oral. Esse dado é relevante do ponto de vista da influência do contexto no letramento em avaliação desses profissionais. No caso dos professores, que atuam permanentemente e exclusivamente no contexto de formação, a avaliação das habilidades orais, de leitura, e de gramática e vocabulário é uma prática cotidiana, uma vez que as avaliações de rendimento focam nessas áreas. Os controladores/instrutores, por sua vez, atuam também em contextos de pós-formação nos quais o foco é a preparação para o EPLIS, um exame que utiliza uma escala de níveis de proficiência específica e avalia somente o desempenho dos examinandos nas habilidades orais.

No caso dos aspectos que apresentaram menores médias, os dois grupos afirmaram possuir menor nível de conhecimento em estatística para testes, avaliação alternativa, impacto e efeito retroativo de exames, tendo o grupo dos controladores/instrutores atribuído valores menores de conhecimento.

Ainda sobre a Tabela 3, na última coluna, é possível identificar o resultado do teste U de Mann Whitney de comparação de médias. A abreviação “ H_0 ” refere-se à “hipótese nula” e significa que não há diferença significativa entre as médias. A “ H_0 ” foi rejeitada em apenas cinco dos aspectos analisados: avaliação de nivelamento, avaliação somativa, avaliação de desempenho, escalas de níveis de proficiência, e interpretação de resultados de testes. Isso significa dizer que os dois grupos veem seus níveis de

conhecimento nessas áreas da mesma forma, ou seja, a diferença entre as médias dos dois grupos não é significativa do ponto de vista da estatística.

Foi também calculado o tamanho de efeito (“ r ”) para identificar o grau em que a H_0 é falsa, ou seja, rejeitada. Considera-se, geralmente, um tamanho de efeito pequeno para $r > 0,20$, médio para $r > 0,50$, e alto para $r > 0,80$. Neste estudo, para todas as situações em que a H_0 foi rejeitada, o tamanho de efeito foi considerado entre os níveis baixo e médio. Somente em cinco situações o “ r ” foi maior do que 0,50: avaliação das habilidades de leitura, da escrita, gramática e vocabulário, exames de alta-relevância e especificações de exames. Interpretamos que, embora, haja diferença significativa entre os níveis de conhecimento percebidos dos dois grupos em 21 dos 26 aspectos da avaliação de línguas, essa diferença se mostrou pequena, com exceção dos casos citados anteriormente.

Necessidades de letramento em avaliação de línguas

Para entender as necessidades de treinamento dos professores de inglês aeronáutico em relação à avaliação de línguas, foi pedido a eles que avaliassem novamente os 26 aspectos, tendo em mente as áreas que gostariam que fossem exploradas em um possível programa de desenvolvimento profissional.

A média geral alcançou 4,22, com desvio padrão calculado em 0,934. Na comparação entre os grupos, as respostas alcançaram médias gerais muito semelhantes: $M=4,30$ no grupo do magistério e $M=4,18$ no grupo de controladores/instrutores. Percebe-se que são médias altas, entre os critérios “provavelmente sim” e “definitivamente sim” do questionário. Na comparação entre as médias dos dois grupos nos 26 aspectos avaliados, o teste U de Mann Whitney apontou nenhuma diferença significativa nas respostas, ou seja, ambos os grupos avaliam da mesma forma os assuntos que gostariam que fossem tratados em um treinamento em avaliação de línguas, indicando necessidades muito parecidas. Médias altas e pouca variação nas respostas entre e dentro dos grupos também ocorreu no estudo de Fulcher (2012) e, na opinião do autor, devem-se ao fato de que os participantes, por terem mostrado interesse e se voluntariado para participar da pesquisa, tendem a ver todos os assuntos como pertinentes e relevantes para um curso de letramento em avaliação de línguas.

Contudo, se olharmos a distribuição das respostas dos participantes por porcentagem (Tabela 4), percebe-se uma concentração alta de respostas (acima de 60%),

no critério 5 “definitivamente sim” do questionário, nos seguintes aspectos: avaliação diagnóstica, avaliação de compreensão e produção orais, avaliação de habilidades integradas, elaboração de itens e provas, escala de níveis de proficiência, e interpretação de resultados de exames.

Tabela 4 – Necessidades percebidas em aspectos do LAL em porcentagem

Aspectos da AL	1	2	3	4	5
Avaliação de nivelamento	2,8	5,6	13,9	30,6	47,2
Avaliação diagnóstica	2,8	2,8	11,1	22,2	61,1
Avaliação de rendimento	5,6	2,8	11,1	30,6	50,0
Avaliação de proficiência	0	0	5,6	36,1	58,3
Avaliação somativa	2,8	2,8	19,4	36,1	38,9
Avaliação formativa	2,8	2,8	19,4	36,1	38,9
Avaliação alternativa	5,6	5,6	11,1	47,2	30,6
Avaliação de desempenho	2,8	0	8,3	38,9	50
Avaliação de compreensão oral	2,8	0	2,8	27,8	66,7
Avaliação de produção oral	2,8	0	2,8	27,8	66,7
Avaliação de leitura	2,8	11,1	16,7	33,3	36,1
Avaliação de produção escrita	2,8	19,4	11,1	33,3	33,3
Avaliação de gramática e vocabulário	0	5,6	16,7	30,6	47,2
Avaliação de habilidades integradas	2,8	2,8	8,3	25,0	61,1
Autoavaliação e avaliação em pares	2,8	5,6	16,7	33,3	41,7
Validade	2,8	2,8	30,6	27,8	36,1
Confiabilidade	2,8	0	33,3	30,6	33,3
Impacto e efeito retroativo	2,8	0	27,8	19,4	50,0
Ética e justiça na avaliação	0	0	19,4	33,3	47,2
Exames de alta-relevância	2,8	0	30,6	25,0	41,7
Especificações de exames	2,8	0	27,8	30,6	38,9
Métodos de avaliação	0	0	11,1	30,6	58,3
Elaboração de itens e provas	2,8	0	5,6	25,0	66,7
Escala de níveis de proficiência	0	0	11,1	27,8	61,1
Interpretação de resultados de provas	2,8	0	5,6	19,4	72,2
Estatística para avaliação	2,8	0	13,9	33,3	50,0

Fonte: autora (2024)

Esses aspectos, considerados pela grande parte dos participantes como essenciais para um curso de letramento em avaliação de línguas, podem ser vistos como áreas prioritárias ou urgentes a serem tratadas em um possível programa de desenvolvimento. Além de apresentarem uma natureza bastante prática, os assuntos selecionados pelos participantes parecem estar relacionados com o contexto de ensino e a avaliação de inglês aeronáutico, que foca no uso das habilidades orais de forma integrada.

Para compreender melhor as necessidades dos participantes, também foi pedido a eles que avaliassem a frequência com que certos tipos de tarefas eram utilizados nas atividades avaliativas nos contextos em que atuavam. A Tabela 5 mostra as médias de

respostas, comparando os dois perfis profissionais em 12 tipos de tarefas.

Nota-se que as médias dos professores do magistério (Grupo 1) alcançaram valores acima de 4 em atividades envolvendo múltipla escolha, entrevista oral individual, *role-plays* e descrição de figuras, ao passo que, no grupo dos controladores/instrutores (Grupo 2), somente o item questões de múltipla-escolha atingiu média superior a 4. Esse dado sugere que o grupo do magistério lida com uma variedade maior de atividades avaliativas formais e isso pode ser explicado pelo fato de estarem inseridos, exclusivamente, no contexto de formação, no qual, além do EPLIS, existem provas de rendimento escritas e orais.

No teste U de Mann Whitney para comparação das médias, foram identificadas diferenças estatisticamente significativas para tarefas de múltipla escolha, entrevista oral individual e leitura em voz alta. O tamanho de efeito também foi calculado e se mostrou entre os níveis “baixo” e “médio” para os três casos. Embora tenha havido diferença nas respostas dos grupos em relação à múltipla escolha, esse tipo de tarefa é o mais trabalhado por ambos, além de estar presente tanto na fase 1 do EPLIS para avaliação isolada da compreensão oral, quanto nas provas de rendimento do curso de formação para avaliação das habilidades de leitura e compreensão oral. Devido a isso, a elaboração de itens de múltipla escolha deve ser explorada em um possível programa de letramento para esse público, principalmente, como um dos métodos possíveis para avaliar a compreensão oral, dado que a avaliação da compreensão oral é também uma necessidade prioritária na visão dos participantes. Ademais, considerando que ambos os grupos colocaram como áreas prioritárias a avaliação de habilidades orais e integradas, atividades que envolvem a compreensão oral e a produção oral em conjunto devem também estar contempladas no programa.

Tabela 5 – Frequência de tarefas avaliativas nos contextos de ensino

		N	Média	Desvio Padrão
Questões de múltipla escolha	Grupo 1	11	4,91	0,302
	Grupo 2	25	4,16	1,214
Perguntas com respostas escritas curtas	Grupo 1	11	2,73	1,618
	Grupo 2	25	2,64	1,381
Questões de verdadeiro ou falso	Grupo 1	11	3,00	1,549
	Grupo 2	25	3,48	1,262
Questões com preenchimento de lacunas	Grupo 1	11	3,64	1,206
	Grupo 2	25	3,28	1,308
Questões com ordenação de eventos	Grupo 1	11	3,18	1,401

Entrevista oral individual*	Grupo 2	25	2,92	1,222
	Grupo 1	11	4,45	0,820
Entrevista oral em par ou em grupo	Grupo 2	25	3,36	1,186
	Grupo 1	11	3,36	1,502
Atividades orais de simulação (role-play)	Grupo 2	25	2,60	1,118
	Grupo 1	11	4,00	1,183
Leitura em voz alta*	Grupo 2	25	3,68	1,314
	Grupo 1	11	3,91	1,375
Descrição oral de figuras	Grupo 2	25	2,76	1,332
	Grupo 1	11	4,45	0,688
Discussão oral	Grupo 2	25	3,84	1,068
	Grupo 1	11	3,91	1,578
Apresentação oral	Grupo 2	25	3,76	1,052
	Grupo 1	11	3,82	1,328
	Grupo 2	25	3,48	1,122

Fonte: autora (2024)

Por fim, os participantes do estudo indicaram a frequência com que as avaliações nos contextos em que atuavam atendiam aos propósitos listados na Tabela 6. Percebe-se que os professores do magistério registraram médias mais altas de respostas, com variação entre $M=4,27$ e $M=3,27$, ou seja, entre os critérios “frequentemente” e “eventualmente”.

Esse dado pode ser explicado pelo fato de esses profissionais estarem mais envolvidos com as práticas avaliativas no curso de formação por, provavelmente, pertencerem ao quadro permanente de docentes. Nesse grupo, os propósitos avaliativos com maior média ($M>4,0$) foram o de aperfeiçoar a instrução e a aprendizagem, seguidos por classificar os alunos em turmas ou níveis e oferecer *feedback* sobre o progresso.

Tabela 6 – Propósitos das tarefas avaliativas nos contextos de ensino

		N	Média	Desvio Padrão
Aperfeiçoar a instrução e aprendizagem	Grupo 1	11	4,27	0,647
	Grupo 2	25	2,96	1,207
Implementar mudanças nos métodos de ensino	Grupo 1	11	3,36	0,924
	Grupo 2	25	2,36	1,114
Classificar os alunos em turmas ou níveis	Grupo 1	11	4,00	1,000
	Grupo 2	25	3,00	1,528
Decidir áreas que deverão ser priorizadas	Grupo 1	11	3,45	0,934
	Grupo 2	25	2,84	1,281
Definir o quanto os alunos aprenderam	Grupo 1	11	3,73	1,009
	Grupo 2	25	3,20	1,190
Decidir quais alunos receberão instrução complementar	Grupo 1	11	3,27	1,348
	Grupo 2	25	2,40	1,225
Oferecer <i>feedback</i> sobre o progresso	Grupo 1	11	4,00	1,104
	Grupo 2	25	3,40	1,225
Decidir sobre a retomada de objetivos de ensino	Grupo 1	11	3,36	1,120
	Grupo 2	25	2,48	1,447
Decidir se os alunos passarão para próxima série	Grupo 1	11	3,73	1,348

ou nível	Grupo 2	25	2,48	1,358
Certificar o nível de proficiência em uma escala	Grupo 1	11	3,82	1,168
	Grupo 2	25	3,16	1,214

Fonte: autora (2024)

É interessante pontuar, no entanto, que o propósito com maior índice de frequência nesse grupo se contrapõe à cultura da testagem, apontado por Souza (2020) como característica predominante do contexto do curso de formação. Utilizar resultados das avaliações com o objetivo de aperfeiçoar o ensino e aprendizagem denota uma visão transformadora da avaliação, na qual avaliar é uma forma de promover o ensino e a aprendizagem. Pode haver, nesse sentido, uma contradição entre o discurso e a prática desses professores, que precisa ser mais bem investigada.

No grupo de controladores/instrutores, as médias de respostas variam entre $M=3,40$ e $M=2,36$, ou seja, entre os critérios “eventualmente” e “raramente” do questionário. O propósito com maior média de respostas foi ‘oferecer *feedback* sobre o progresso’, porém ainda abaixo de 4,0. Esse dado sugere, mais uma vez, que o grupo de controladores/instrutores não tem tanto envolvimento com avaliações formais e institucionalizadas nos contextos em que atuam ou não participa ativamente dessas práticas e tampouco tem poder de decisão sobre elas. Na comparação entre as médias dos dois grupos, foram encontradas diferenças significativas somente em relação aos propósitos de ‘aperfeiçoar a instrução e a aprendizagem’ e ‘decidir se os alunos passarão para o próximo nível ou série’, o que reforça ainda mais a percepção anterior. Por outro lado, embora os controladores/instrutores não estejam envolvidos diretamente com práticas formais de avaliação, é importante que eles também saibam compreender as informações que elas veiculam para que possam agir mais ativamente e participar também da tomada de decisões que as envolvem.

CONCLUSÃO E IMPLICAÇÕES FUTURAS

Este estudo buscou responder duas perguntas de pesquisa, tendo a primeira o propósito de identificar o nível percebido de LAL de professores de inglês aeronáutico e suas necessidades na área, enquanto a segunda pergunta buscou verificar as diferenças com base no perfil profissional dos respondentes.

De forma a sintetizar as tendências gerais e destacar as questões mais salientes

levantadas a partir das respostas dos participantes no questionário, podemos dizer que houve diferença em como as duas categorias de professores de inglês aeronáutico percebem o seu nível de LAL. Os profissionais da carreira de magistério atribuíram um nível maior de conhecimento em relação aos controladores/instrutores e isso pode estar ligado ao fato de o primeiro grupo ter um nível de escolaridade mais alto na área de ensino de línguas e afirmarem se envolver com maior frequência com a elaboração de provas e itens. Porém, a maior parte dos respondentes disse não ter recebido treinamento específico em LAL nos cursos que realizou. Esses dados sugerem que pode haver carências ou até mesmo equívocos de entendimento em relação à avaliação de línguas. Mesmo havendo diferença no nível percebido de LAL dos grupos, essa diferença se mostrou relativamente pequena e não justificaria desenvolver treinamentos com níveis distintos de profundidade para cada grupo.

Ademais, em termos de necessidades, as análises mostraram que ambos os grupos percebem suas necessidades de forma bastante semelhante e com prioridades a serem tratadas: avaliação de compreensão oral e produção oral; avaliação das habilidades integradas; elaboração de testes e itens com ênfase em questões de múltipla escolha e tarefas que integrem as habilidades orais; escala de níveis de proficiência para avaliação de habilidades orais integradas; e interpretação de resultados de testes. Esses assuntos podem ser considerados como urgentes em um treinamento que pretende aumentar o nível de LAL desse público, unindo, assim, as habilidades técnicas e resultados aos conceitos e princípios que os subjazem, dentro de uma abordagem procedimental (Fulcher, 2012; Taylor, 2013; Kremmel; Harding, 2020).

Quanto aos propósitos de avaliação, houve diferença entre os grupos em como os resultados das avaliações são utilizados em seus contextos de atuação. Enquanto os professores parecem estar mais envolvidos com avaliações formais e institucionais, os controladores/instrutores disseram que as utilizam mais para oferecer *feedback* sobre o progresso dos alunos. Adicionado a isso, dado que os controladores/instrutores atuam em contextos de pós- formação nos quais, provavelmente, não há um programa de ensino instituído, seu nível de envolvimento com avaliações que visam, por exemplo, à implementação de mudanças metodológicas ou à retomada dos objetivos de aprendizagem, é relativamente baixo. Outro ponto seria o não envolvimento nas tarefas avaliativas devido ao fato de esses profissionais não pertencerem ao quadro permanente

de professores do curso de formação. De qualquer forma, entendemos que os propósitos e as decisões que são tomadas com base nos resultados de exames devem fazer parte do currículo de um treinamento em LAL para ambos os profissionais, bem como discussões sobre o papel social da avaliação na aprendizagem, como argumentado por Kremmel e Harding (2020). É de suma importância que os professores de inglês aeronáutico sejam capazes de compreender as informações sobre o desempenho de seus alunos nos testes e utilizá-las em benefício do processo de ensino e aprendizagem, bem como na tomada de decisões mais informadas e apropriadas.

O impacto e efeito retroativo de exames foi um aspecto do LAL que recebeu níveis mais baixos de conhecimento por parte dos participantes. Considerando que ambos os grupos atuam em contextos nos quais exames externos, como o EPLIS, são implementados para impulsionar mudanças, um programa de LAL para esse público deve explorar esses conceitos e suas implicações na preparação dos alunos, bem como promover uma análise crítica da implementação de políticas de avaliação e seus desdobramentos em contextos locais.

Por fim, pelo fato de este estudo ter focado somente na coleta e análise de dados quantitativos, não foi possível um entendimento mais profundo das necessidades e práticas avaliativas dos participantes em seus contextos de atuação. Sugere-se que sejam conduzidas entrevistas e grupos focais, com o intuito de explorar as características das avaliações nos contextos de formação e pós-formação de controladores de tráfego aéreo, os papéis dos professores da carreira de magistério e dos controladores/instrutores nessas avaliações, e os propósitos com que, de fato, são utilizadas.

REFERÊNCIAS

BACHMAN, L.; DAMBÖCK, B. *Language assessment for classroom teachers*. Oxford: OUP, 2017.

FULCHER, G. Assessment literacy for the language classroom. **Language Assessment Quarterly**, v. 9, n. 2, p. 113-132, 2012.

INBAR-LOURIE, O. Constructing a language assessment knowledge base: A focus on language assessment courses. **Language Testing**, v. 25, n. 3, p. 385-402, 2008a.

_____. Language assessment culture. In: SHOHAMY, E.; HORNBERGER, N. H. (Ed.). **Encyclopedia of language and education: Language testing and assessment**. Springer Science+Business Media LLC, 2008b, p. 285-299. v. 7.

_____. Guest editorial to the special issue on language assessment literacy. **Language Testing**, v. 30, n. 3, p. 301-307, 2013.

KREMMEL, B.; HARDING, L. Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. **Language Assessment Quarterly**, v. 17, n. 1, p. 100-120, 2020.

QUEVEDO-CAMARGO, G.; SCARAMUCCI, M.V.R. O conceito de letramento em avaliação de línguas: Origem de relevância para o contexto brasileiro. **Linguagem: Estudos e Pesquisa**, v. 22, n. 1, p. 225-245, 2018. Disponível em: <<https://www.revistas.ufg.br/lep/article/view/54474>>. Acesso em: 11 set. 2022.

SCARAMUCCI, M. V. R. O professor avaliador: sobre a importância da avaliação na formação do professor de língua estrangeira. In: Rottava, L. (Org.). **Ensino-aprendizagem de línguas: Língua Estrangeira**. 1. ed. Ijuí: Editora da UNIJUI, 2006. p. 49-64. v. 1.

_____. Letramento em avaliação (em contexto de línguas): Contribuições para a linguística aplicada, educação e sociedade. In: JORDÃO, C. M. (Org). **Linguística aplicada no Brasil: rumos e passagens**. Campinas: Pontes Editora, 2016. p. 141-165.

SCARAMUCCI, M.V.R. et al. **Validação do exame de proficiência em Língua Inglesa para profissionais prestadores de serviços de tráfego aéreo do SISCEAB**. 2010. 243f. Relatório final de pesquisa. Campinas: Unicamp, 2010.

SCARINO, A. Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. **Language Testing**, v. 30, n. 3, p. 309-327, 2013.

SHOHAMY, E. **The power of tests: the impact of language tests on teaching and learning**. NFLC Occasional Paper. Washington: National Foreign Language Center, 1993.

SOUZA, P. R. O nível de proficiência como fator mediador no processo de efeito retroativo do EPLIS. **Revista Estudos Linguísticos**, v. 48, n. 1, p. 540-560, 2019. DOI: 10.21165/el.v48i1.2151.

_____. The washback of EPLIS on teachers' perceptions and actions: Implications for reviewing ICAO language policy in Brazil. **The Specialist**, v. 41, n. 3, p. 1-18, 2020. DOI:10.23925/2318-7115.2020v41i3a7.

SPOLSKY, B. Language testing at 25: Maturity and responsibility? **Language Testing**, v. 25, n. 3, p. 297-305, 2008.

STIGGINS, R. Assessment literacy. **Phi Delta Kappan**, v. 72, p. 534-539, 1991.

TAYLOR, L. Developing assessment literacy. **Annual review of applied linguistics**, v. 29, p. 21-36, 2009.

_____. Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. **Language Testing**, v. 30, n. 3, p. 403-41, 2013. DOI:10.1177/0265532213480338.

TSAGARI, D.; VOGT, K. Assessment Literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects. **Papers in Language Testing and Assessment**, v. 6, n. 1, p. 41-62, 2017.

VOGT, K.; TSAGARI, D. Assessment literacy of foreign language teachers: Findings of a European study. **Language Assessment Quarterly**, v. 11, n. 4, p. 374-402, 2014. DOI:10.1080/15434303.2014.9600.

Recebido em: Fev. 2024.
Aceito em: Abr. 2024.

APÊNDICE A

NÍVEL PERCEBIDO E NECESSIDADES DE LETRAMENTO EM
AVALIAÇÃO DE LÍNGUAS

Este questionário visa coletar informações sobre as necessidades de professores e instrutores de inglês aeronáutico na área de avaliação de línguas, bem como os usos e propósitos da avaliação nos contextos de ensino em que atuam. O objetivo é analisar os dados para a elaboração de um curso que busque suprir as lacunas na formação desses profissionais, a fim de contribuir para um entendimento mais profundo sobre avaliação e o aperfeiçoamento das práticas avaliativas em sala de aula.

Parte 1 Dados profissionais

1a - Você é professor ou instrutor de inglês aeronáutico?

Sim ou Não

1b- Há quanto tempo atua como professor ou instrutor de inglês aeronáutico?

- a) Há 02 anos ou menos
- b) Entre 03 e 06 anos
- c) Entre 07 e 10 anos
- d) Entre 11 e 14 anos
- e) Há 15 anos ou mais

2a - Você é controlador de tráfego aéreo?

Sim ou Não

2b - Há quanto tempo atua como controlador de tráfego aéreo?

- a) Há 02 anos ou menos
- b) Entre 03 e 06 anos
- c) Entre 07 e 10 anos
- d) Entre 11 e 14 anos
- e) Há 15 anos ou mais

3a - Você é um examinador EPLIS?

Sim ou Não

3b - Há quanto tempo atua como examinador EPLIS?

- a) Há 02 anos ou menos
- b) Entre 03 e 06 anos
- c) Entre 07 e 10 anos
- d) Entre 11 e 14 anos
- e) Há 15 anos ou mais

4 - Você elabora itens de teste/prova ou simulado nos contextos de ensino em que atua?

Sim ou Não

5 - Com que frequência você elabora itens de teste/prova ou simulado nos contextos de ensino em que atua?

- a) Nunca
- b) Raramente
- c) Ocasionalmente
- d) Frequentemente
- e) Sempre

6 - Qual é o seu nível atual de formação na área de ensino de línguas?

- f) Cursos como CTP011, CELTA ou TESOL
- g) Graduação
- h) Especialização
- i) Mestrado
- j) Doutorado

7 - Qual é a sua faixa etária?

- a) 21-29
- b) 30-39
- c) 40-49
- d) 50-59
- e) 60 ou acima

Parte 2 Conhecimentos em avaliação de línguas

8 - Você teve disciplinas com foco em avaliação nos cursos que realizou?

Sim ou Não

9 - Como você avalia o seu nível de conhecimento na área de avaliação de línguas?

5 = excelente 4 = bom 3 = razoável 2 = pouco 1 = não tenho conhecimento

10 – Como você avalia o seu nível de conhecimento dos assuntos listados abaixo.

Utilize a seguinte escala:

5 = excelente; 4 = bom; 3 = razoável; 2 = pouco; e 1 = não tenho conhecimento

	5	4	3	2	1
Avaliação de classificação					
Avaliação diagnóstica					
Avaliação de rendimento					
Avaliação de proficiência					
Avaliação somativa					
Avaliação formativa					
Avaliação alternativa					
Avaliação de desempenho					
Avaliação de compreensão oral					
Avaliação de produção oral					
Avaliação de leitura					
Avaliação de produção escrita					
Avaliação de gramática e vocabulário					
Avaliação de habilidades integradas					
Autoavaliação e avaliação em pares					
Validade					

Confiabilidade					
Praticidade					
Impacto e efeito retroativo de exames					
Ética e justiça na avaliação					
Exames de alta relevância					
Especificações de exames					
Métodos de avaliação					
Elaboração de testes e itens de testes					
Escalas de níveis de proficiência (ex: ICAO Rating Scale)					
Interpretação dos resultados de testes					
Estatística para avaliação					

11 – Quais assuntos você gostaria que fossem explorados em um curso de avaliação de inglês aeronáutico? Utilize a seguinte escala:

5= definitivamente sim; 4= provavelmente sim; 3= indiferente; 2= provavelmente não; 1= definitivamente não

	5	4	3	2	1
Avaliação de classificação					
Avaliação diagnóstica					
Avaliação de rendimento					
Avaliação de proficiência					
Avaliação somativa					
Avaliação formativa					
Avaliação alternativa					
Avaliação de desempenho					
Avaliação de compreensão oral					
Avaliação de produção oral					
Avaliação de leitura					
Avaliação de produção escrita					
Avaliação de gramática e vocabulário					
Avaliação de habilidades integradas					
Autoavaliação e avaliação em pares					
Validade					
Confiabilidade					
Praticidade					
Impacto e efeito retroativo de exames					
Ética e justiça na avaliação					
Exames de alta relevância					
Especificações de exames					
Métodos de avaliação					
Elaboração de testes e itens de testes					
Escalas de níveis de proficiência (ex: ICAO Rating Scale)					
Interpretação dos resultados de testes					
Estatística para avaliação					

Parte 3 Métodos de avaliação de línguas

12 - Com que **frequência** as tarefas abaixo aparecem em exames utilizados nos contextos de ensino em que atua? Utilize a seguinte escala:

5= sempre; 4= frequentemente; 3= ocasionalmente; 2= raramente; 1= nunca

	5	4	3	2	1
Questões de múltipla escolha					
Perguntas com respostas escritas curtas					
Questões de verdadeiro ou falso					
Questões com preenchimento de lacunas					
Questões com ordenação de eventos					
Entrevista oral individual					
Entrevista oral em par ou em grupo					
Atividades orais de simulação (<i>role-play</i>)					
Leitura em voz alta					
Descrição oral de figuras					
Discussão oral sobre determinado tópico					
Apresentação oral sobre determinado tópico					

Parte 4 Usos e propósitos da avaliação de línguas

13 - Com que **frequência** as avaliações nos contextos de ensino em que você atua são utilizadas com os seguintes propósitos? Utilize a seguinte escala:

5= muito freqüente; 4= frequentemente; 3= eventualmente; 2= raramente; 1= nunca

	5	4	3	2	1
Aperfeiçoar a instrução e a aprendizagem.					
Implementar mudanças nos métodos de ensino e nas atividades de sala de aula, materiais e currículo escolar.					
Classificar os alunos em turmas ou em níveis de acordo com o programa de ensino.					
Decidir as áreas que deverão ser priorizadas na instrução.					
Definir o quanto os alunos aprenderam em determinado período.					
Decidir quais alunos deverão receber instrução complementar.					
Oferecer <i>feedback</i> aos alunos sobre seu progresso no curso.					
Decidir se será necessário retomar determinado objetivo de ensino.					
Decidir se os alunos passarão para a série seguinte ou para o próximo nível.					
Certificar o nível de proficiência dos alunos dentro de uma escala.					

Referências para elaboração do questionário:

BACHMAN, L.; DAMBÖCK, B. *Language assessment for classroom teachers*. Oxford: OUP, 2017.

GLENN, F. (2012). Assessment Literacy for the Language Classroom, *Language Assessment Quarterly*, v. 9, n. 2, p. 113-132, DOI: 10.1080/15434303.2011.642041.