

O nível de proficiência como fator mediador no processo de efeito retroativo do EPLIS

DOI: <http://dx.doi.org/10.21165/el.v48i1.2151>

Paula Ribeiro e Souza¹

Resumo

Este estudo tem como objetivo investigar como o nível de proficiência de alunos de um curso de formação em controle de tráfego aéreo interfere nas suas percepções e atitudes acerca de um exame externo de inglês aeronáutico, ao qual são submetidos ao final do programa. Para isso, foi elaborado e aplicado um questionário aos 407 alunos do curso. Os dados foram submetidos à análise de componentes principais e a testes de hipóteses para investigar relações de significância entre os construtos fundamentais presentes no questionário e o nível de proficiência declarado pelos participantes. Os resultados mostraram que os alunos que se consideram mais proficientes atribuem níveis mais altos aos conhecimentos que possuem sobre o exame, à frequência com que se engajam em práticas preparatórias, à importância atribuída ao sucesso no teste, bem como ao alinhamento entre ensino e avaliação externa.

Palavras-chave: efeito retroativo; ensino e aprendizagem de inglês aeronáutico; proficiência.

¹ Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brasil; paulaletras@gmail.com; <http://orcid.org/0000-0001-7202-531X>

The proficiency level as a mediator factor in the washback effect process in EPLIS

Abstract

This paper aims at investigating how the learners' proficiency level in an air traffic control initial training program affects their perceptions and attitudes towards an external aviation English test, administered to them at the end of the course. To this end, a questionnaire was designed and delivered to 407 students enrolled in the program. Principal component analyses and hypothesis tests were conducted on the data to plot the significance of the relationships between the fundamental constructs of the questionnaire and the learners' perceived level of proficiency. Results showed that learners with a higher perceived proficiency rate their knowledge about the exam higher, as well as the frequency with which they engage in preparation activities, the importance of being successful in the test and the alignment between teaching and external examination.

Keywords: washback effect; aviation English teaching and learning; proficiency.

Exames de alta-relevância no contexto aeronáutico

Exames externos são considerados instrumentos de poder, uma vez que possuem potencial para direcionar ações nos contextos de ensino e aprendizagem que os precedem. Shohamy (2006, p.103, tradução nossa) pontua que "exames são capazes de impor o que será estudado, aprendido e ensinado por professores e alunos"², principalmente e mais facilmente, se a relevância atribuída a eles for considerada alta. McNamara e Roever (2006) definem exames de alta-relevância como aqueles cujos resultados acarretam consequências graves para os indivíduos e instituições.

No contexto da aviação civil internacional, exames de proficiência em língua inglesa têm sido aplicados a pilotos, controladores de tráfego aéreo (controladores, doravante) e operadores de estação aeronáutica desde 2003, em cumprimento aos padrões e práticas recomendadas pela OACI³ que buscam elevar o nível de proficiência desses profissionais no gerenciamento de situações inesperadas, incomuns e emergenciais de voo (OACI, 2010).

2 No original: "Tests are capable of dictating what will be studied, learned and taught by students and teachers".

3 A Organização da Aviação Civil Internacional, ICAO ou OACI em português, é uma agência regulatória da ONU, responsável por promover o desenvolvimento seguro e ordenado da aviação civil internacional. É formada atualmente por 191 países-membros.

Como membro da OACI, o Brasil, representado pelo Departamento de Controle do Espaço Aéreo (DECEA), desenvolveu, em 2007, o Exame de Proficiência em Inglês Aeronáutico do SISCEAB⁴ (EPLIS), para avaliar o nível de inglês dos controladores e operadores de estação aeronáutica brasileiros⁵. Desde então, o EPLIS tem sido aplicado anualmente a esses profissionais que se encontram em pleno exercício da profissão e envolvidos diretamente com operações internacionais (TOSQUI-LUCKS *et al.*, 2016). Contudo, em 2014, o DECEA estendeu o EPLIS aos alunos do curso de formação em controle de tráfego aéreo, aparentemente, com o propósito de elevar o nível de relevância do exame nesse contexto.

Os resultados do EPLIS informam decisões importantes tanto para os indivíduos que se submetem a ele, com a geração de benefícios aos aprovados e sanções aos reprovados, quanto para o DECEA, que necessita implementar ações de mitigação para garantir a segurança de voo, bem como investir em cursos de pós-formação para manutenção e elevação do nível de proficiência desses profissionais. Nesse sentido, o EPLIS é considerado de alta-relevância e sua introdução no cenário escolar o torna um mecanismo capaz de nortear os eventos em sala de aula.

O impacto ou efeito retroativo que exames de alta-relevância exercem nos contextos de ensino e aprendizagem tem sido investigado por diversos pesquisadores na área da avaliação educacional e de línguas (ALDERSON; WALL, 1993; ALDERSON; HAMP-LYONS, 1996; CHENG, 1999; SCARAMUCCI, 1999; WATANABE, 2004). Os resultados desses estudos mostram que mudanças atreladas à adoção de determinado teste nem sempre são alcançadas conforme esperado, devido à natureza complexa, multifacetada e variável do efeito retroativo. Em outras palavras, diversos fatores, além do teste em si, interferem na determinação desse impacto, tais como, as crenças, os valores e os conhecimentos de alunos e professores, as exigências das escolas em que esses exames estão inseridos, bem como a percepção de relevância do teste, sua proximidade e o nível de proficiência dos alunos. Estudos sobre efeito retroativo são considerados de suma importância, pois, além de investigar a extensão de implementação dos efeitos pretendidos, informam ainda a presença e a natureza de impactos considerados não-pretendidos ou até mesmo indesejáveis.

Nessas pesquisas, os professores têm ganhado destaque como foco de investigação, devido ao seu importante papel na determinação do impacto do teste em sala de aula. Bailey (1996) descreve esses participantes como a ponte entre o exame e o ensino e a aprendizagem. Os alunos, por sua vez, mesmo sendo considerados uns dos principais afetados pelas consequências do uso de testes, tendem a ser menos estudados. Cheng

4 Sigla utilizada para designar o Sistema de Controle do Espaço Aéreo Brasileiro.

5 A avaliação de proficiência dos pilotos civis brasileiros está a cargo da Agência Nacional de Aviação Civil (ANAC).

et al. (2010) afirmam que pouco ainda se sabe, de fato, sobre como o efeito retroativo opera nos alunos, fazendo-se, portanto, mister maiores investigações sobre como os alunos percebem e agem em relação à influência de um exame (ALDERSON; BANERJEE, 2001; SCARAMUCCI, 2004; STONEMAN, 2006; CHENG *et al.*, 2010).

Considerando, portanto, a importância de estudos sobre efeito retroativo de exames de alta-relevância, como o EPLIS, somada à necessidade de investigações que foquem o aluno⁶, este trabalho propõe-se a analisar a relação entre o nível de proficiência declarado pelos alunos do curso de formação em controle de tráfego aéreo e suas percepções e atitudes para com o exame, a fim de identificar se o fator proficiência interage no processo de efeito retroativo do exame. Para a coleta dos dados, foi aplicado um questionário a 407 alunos, com uma taxa de retorno de 95%. As respostas foram analisadas empregando-se a técnica estatística da análise de componentes principais e testes de hipóteses. Os resultados apresentados aqui são um recorte de uma pesquisa de doutorado em fase final de desenvolvimento.

Este artigo está organizado em seis seções. Nesta primeira seção, apresentamos um panorama sobre a introdução de exames de proficiência no contexto aeronáutico e o impacto no ensino e aprendizagem. Na próxima seção, discutimos o conceito de efeito retroativo e os fatores envolvidos na determinação desse fenômeno. Na seção 3, descrevemos os procedimentos metodológicos adotados na coleta e na análise de dados, bem como o cenário investigado e os participantes da pesquisa. Na quarta seção, analisamos os dados obtidos pelas técnicas estatísticas adotadas. Os resultados alcançados são, então, discutidos na seção 5. Por fim, encerramos o artigo tecendo algumas considerações sobre o estudo.

A complexidade do efeito retroativo

Alderson e Wall (1993, p.117, tradução nossa), em seu artigo seminal sobre a existência do efeito retroativo, definem o fenômeno como sendo o grau com que um teste influencia “professores e alunos de línguas a fazerem coisas que eles não necessariamente fariam se não fosse pela existência do teste”⁷. Eles explicam que, até os anos 90, esse impacto era compreendido como tendo um caráter determinista, ou seja, acreditava-se que um teste ruim resultaria em efeitos negativos enquanto um bom teste traria efeitos benéficos. No entanto, essa visão é ingênua, uma vez que não considera as diversas forças que podem influenciar o ensino e aprendizagem. Como bem coloca Messick (1996, p. 242, tradução

6 Souza (2017) investigou o efeito retroativo do EPLIS nas percepções e atitudes de professores do curso de formação em controle de tráfego aéreo, antes da implementação do exame no cenário escolar.

7 No original: “teachers and learners do things they would not necessarily otherwise do because of the test”.

nossa), “um teste ruim pode estar associado a efeitos positivos e um bom teste com efeitos negativos devido a outras coisas que são feitas ou não no sistema educacional”.

Dessa forma, embora elaborar um bom exame seja uma condição necessária para a promoção de melhorias no ensino e aprendizagem, segundo Scaramucci (2011), essa condição não é suficiente devido à multiplicidade de fatores que interferem na determinação do efeito retroativo. Os estudos de Alderson e Wall (1993), Cheng (1999), Scaramucci (1999) e Watanabe (2004) sobre efeito retroativo são alguns exemplos de muitos que confirmam a complexidade do fenômeno, cujo processo é de natureza multifacetada e mediado por variáveis que podem interferir de forma positiva ou negativa no impacto de um teste.

Dentre os diversos fatores constituintes do efeito retroativo, Alderson e Wall (1993) argumentam que a relevância de um teste influencia positivamente a intensidade de seu efeito. Assim, quanto maior for a percepção de relevância de um exame, maior o seu potencial de impacto. Além da importância dada à aprovação por alunos e professores, Hughes (1993) sustenta que esses participantes devem estar familiarizados com as exigências do teste e devem possuir os conhecimentos exigidos ou uma forma de alcançá-los por meio da disponibilidade de recursos necessários. Alderson e Hamp-Lyons (1996) acrescentam outros três elementos intermediadores no processo de efeito retroativo: a) até que ponto o exame é contrário às práticas em vigor; b) a opinião de professores e autores de livros didáticos sobre os métodos mais apropriados na preparação para o teste; e c) a disposição e capacidade de professores e autores de livros didáticos de inovar. Watanabe (2004) e Burrows (2004) confirmam que fatores pessoais ligados aos professores, tais como, conhecimentos, crenças e valores, interferem na determinação do efeito retroativo. Ademais, fatores micro e macro contextuais, por exemplo, as exigências das escolas onde o exame está inserido e da sociedade (WATANABE, 2004) também são considerados mediadores no processo de efeito retroativo, bem como a percepção de dificuldade do exame. Green (2007) explica que se os participantes percebem o exame como fácil demais, eles não dispensarão qualquer atenção ao teste. As demandas do exame também serão ignoradas se elas forem entendidas como intangíveis por professores e alunos.

O nível de proficiência de alunos também tem sido investigado como um dos fatores capazes de afetar a promoção e a intensidade do efeito retroativo (LANZONI, 2004; CHENG *et al.*, 2010; STENLUND; EKLÖF; LYREN, 2017). Nesse sentido, a percepção do nível de proficiência do aluno relaciona-se com o nível de dificuldade atribuído ao teste. Assim, alunos mais proficientes, ao avaliar as demandas do exame, podem considerá-las desafiadoras, porém, tangíveis, e se mobilizam na preparação, o que faz aumentar a intensidade do seu efeito para esse grupo. Por outro lado, se esses alunos percebem o exame como fácil demais, o efeito retroativo pode ser baixo ou até mesmo nulo. O mesmo pode ocorrer com os que se consideram menos proficientes e avaliam o exame como muito difícil.

A metodologia do estudo

Para atingir o objetivo proposto neste trabalho, foi realizado um estudo quantitativo de caráter exploratório. Para a coleta de dados, foi elaborado e aplicado um questionário do tipo *survey* aos alunos de um curso de formação em controle de tráfego aéreo. Os dados foram analisados estatisticamente por meio do método de análise de componentes principais e de testes de hipóteses.

O cenário do estudo e os participantes

Este estudo foi conduzido em um curso de formação em controle de tráfego aéreo de uma escola militar pertencente ao Comando da Aeronáutica. O ingresso para a realização desse curso é feito por meio de concurso público e os candidatos devem possuir o 2º grau de escolaridade completo e ter entre 17 e 24 anos de idade. Após a conclusão do curso, esses profissionais são promovidos à Graduação de Terceiro Sargento da Aeronáutica e designados a atuar nos diversos órgãos de controle de tráfego aéreo espalhados no território nacional, conforme necessidade do serviço e classificação no curso de formação. Na época da geração de dados, o curso era composto por quatro semestres letivos e havia 407 alunos distribuídos em quatro séries. O currículo escolar previa quatro disciplinas de língua inglesa, uma por semestre, totalizando uma carga horária de 500 tempos de aula. Do 2º semestre em diante, os alunos eram distribuídos em turmas menores de acordo com o nível de proficiência obtido em um teste de nivelamento aplicado antes do início das aulas.

O EPLIS no contexto escolar

O EPLIS pode ser considerado um exame de desempenho com propósitos específicos, uma vez que os conhecimentos da área de controle de tráfego aéreo interagem com os conhecimentos linguísticos em uma situação comunicacional (SCARAMUCCI *et al.*, 2010). Foi desenvolvido com base nos princípios descritos no Manual de Implementação dos Requisitos de Proficiência Linguística da OACI (OACI, 2010) e 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.

Desde 2014, esse exame tem sido aplicado aos alunos do último semestre do curso de formação em controle de tráfego aéreo. Os resultados obtidos são considerados oficiais e constam no registro do aluno junto ao DECEA. Contudo, embora seja obrigatório, o EPLIS não é considerado ainda um exame de saída. Ou seja, a reprovação não impede o aluno de se formar no curso e nem de prosseguir para o estágio operacional nos órgãos de controle de tráfego aéreo espalhados pelo Brasil.

Como sustentam Bachman e Palmer (2010), a intenção dos elaboradores de um teste, bem como dos usuários de seus resultados, em geral, é contribuir para que efeitos benéficos sejam gerados tanto para os indivíduos quanto para os programas educacionais. Nesse sentido, a implementação do EPLIS no contexto escolar é vista como uma ação capaz de elevar os níveis de proficiência dos alunos, ao promover práticas mais condizentes com as necessidades de uso real da língua pelo futuro profissional.

O questionário

Os dados para análise foram coletados por meio de um questionário do tipo *survey*. Esse questionário era composto por 64 itens, divididos em quatro partes. A primeira parte abrangeu nove perguntas em escala nominal sobre o perfil dos alunos e o nível de proficiência declarado por eles. A segunda parte buscou acessar o nível de conhecimentos dos alunos sobre o EPLIS. A parte 3 focou na influência do exame no conteúdo, métodos, tipos de tarefas, habilidades e na dedicação dos alunos dentro e fora da sala de aula na preparação para o exame. A última parte dedicou-se a levantar informações sobre as visões e percepções do aluno em relação ao exame e sua influência. Os itens pertencentes às três últimas partes, 55 no total, foram elaborados em escala *likert* de 5 pontos, em termos de concordância, avaliação ou frequência e representam as variáveis que serão investigadas pela análise dos componentes principais. O quadro 1 apresenta os temas explorados no questionário, as variáveis a serem analisadas e o número de questões presentes em cada seção.

Quadro 1. A estrutura do questionário

Temas	Conteúdo/Variáveis	Itens
Parte I – Informações Gerais	Informações sobre sexo, série, turma de inglês, tempo de estudo na língua inglesa; nível de proficiência geral na língua e nível de proficiência nas quatro habilidades.	9
Parte II – Conhecimento sobre o EPLIS	Nível de conhecimento sobre o exame em geral, objetivo do exame, Doc 9835 da OACI, conteúdo, tarefas da fase 1 e fase 2, tabela de níveis de proficiência, desempenho do nível 4 de proficiência e manual do candidato; frequência com que os professores mencionam o EPLIS e discutem aspectos específicos do exame.	10
Parte III – Preparação dos alunos para o EPLIS	Frequência com que são trabalhadas atividades semelhantes às do exame em sala de aula; frequência com que as atividades em sala de aula estão relacionadas à aviação e controle de tráfego aéreo; frequência com que certas atividades são trabalhadas em sala de aula; preparação para o exame fora da sala de aula; tempo destinado à preparação para o exame.	32
Parte IV – Visões e Percepções sobre o EPLIS	Influência do exame na motivação do aluno e na elevação do nível de proficiência; pressão do exame no aluno; aumento na carga de estudos e no nível de estresse; importância do exame; relação entre exame e preparação.	13

Dos 407 alunos, 387 participaram do estudo com o preenchimento de um questionário. A tabela 1 contém informações sobre o número de alunos por série e a porcentagem de respondentes.

Tabela 1. Alunos participantes do estudo

	N° alunos	Respostas recebidas	Taxa de retorno
1ª série	129	116	89,92%
2ª série	117	116	99,14%
3ª série	113	109	96,46%
4ª série	48	46	95,83%
TOTAL	407	387	95,08%

Os métodos estatísticos

Primeiramente, os dados coletados foram submetidos à análise fatorial de componentes principais com o objetivo de identificar e agrupar variáveis que se correlacionaram altamente no questionário, reduzindo, assim, as informações e gerando uma estrutura mais parcimoniosa, ou seja, formada por um número menor de variáveis independentes - os chamados fatores ou componentes. Esse tipo de técnica é bastante adequado para analisar fenômenos considerados multifacetados, como é o caso do efeito retroativo, pois permite identificar as dimensões mais latentes, ou seja, as facetas fundamentais do fenômeno dentro do cenário investigado.

Para obter a estrutura fatorial, os itens do questionário foram organizados em três construtos ou temas: 1) Conhecimento sobre o EPLIS; 2) Preparação para o EPLIS e 3) Visões e percepções sobre o EPLIS, representados pelas partes II, III e IV do questionário, respectivamente. Quanto ao levantamento do número ótimo de componentes a serem extraídos, utilizamos os critérios da raiz latente (ou critério Kaiser), o teste *screeplot* e a análise paralela de Montecarlo. Isso se faz necessário pelo fato de ser possível extrair tantos componentes quantos forem o número de variáveis ou itens presentes no questionário. No entanto, a partir de certo ponto, a variância compartilhada de uma variável nas correlações com as demais começa a ser dominada pela variância única, quando os componentes apresentam itens com pouca partilha de informações entre si.

Após a extração dos componentes principais, chegamos à fase de categorização. De acordo com Hair *et al.* (2009), nosso conhecimento sobre determinado conceito nos possibilita interpretar e descrever as relações existentes entre os itens específicos dentro de cada componente (dimensão), a fim de dar significado àquilo que eles representam conjuntamente. Para auxiliar na interpretação da matriz fatorial, dois processos foram

considerados: a rotação de fatores e significância de carga fatorial. O objetivo da rotação de fatores tipo varimax é fazer com que valores em cada coluna da matriz fatorial se mantenham entre +1 e -1 a fim de facilitar a interpretação. Correlações variável-fator mais próximas de +1 ou -1 indicam uma clara associação positiva ou negativa, enquanto correlações próximas de 0 apontam para falta de associação. A significância de carga fatorial refere-se ao grau de correlação de cada variável com o fator e varia de 0 a 1. Quanto maior a carga fatorial de determinada variável, maior a sua representação dentro daquele componente. Neste estudo, foram consideradas relevantes para a determinação dos componentes cargas fatoriais iguais ou maiores que $\pm 0,50$. Variáveis com cargas mais altas tiveram maior peso na interpretação do componente extraído. Para manter um bom nível de consistência dos dados, itens que apresentaram valores de Alfa Cronbach inferiores a 0,3 foram descartados.

Uma vez identificados e categorizados, aplicamos testes de hipóteses para investigar relações de significância entre esses componentes e o nível de proficiência declarado pelos alunos. Os testes de hipóteses nos permitem testar afirmações sobre as características de uma amostra, a fim de descrever tendências e padrões gerais a respeito da população. Neste estudo, foi utilizado o teste *t* de *student* para amostras independentes que compara as médias nas respostas de dois grupos independentes para identificarmos se a diferença entre elas é estatisticamente significativa. Para isso, utiliza-se o critério chamado 'p-valor' ou Sig 2 tailed que indica se a hipótese nula será rejeitada ou não. Ou seja, se p-valor (Sig) for igual ou menor que o ponto de corte estabelecido, a hipótese nula é rejeitada, existindo, assim, diferença significativa nas médias de respostas dos grupos analisados. Caso contrário, a hipótese nula não é rejeitada. Normalmente, a significância é constatada quando o p-valor é inferior ou igual a 0.05. Contudo, como submetemos os dados à análise fatorial, esse valor teve de ser reajustado.

Para a aplicação do teste *t* de *student* para amostras independentes, dividimos os alunos em dois grandes grupos: com alta e baixa proficiência linguística. No grupo com alta proficiência, estão aqueles que responderam possuir nível intermediário ou avançado na aprendizagem de língua inglesa. No grupo com baixa proficiência, encontram-se os que se consideraram iniciantes ou no nível básico. Foram descartados os participantes que reportaram nível pré-intermediário por se posicionarem entre as duas categorias.

Análise dos dados

A análise dos componentes principais possibilitou reduzir um fenômeno envolvendo 55 variáveis a 10 elementos, facilitando, assim, a aplicação e a interpretação de testes de hipóteses com o intuito de analisar a significância estatística na relação entre os componentes extraídos e o fator proficiência. A estrutura fatorial do questionário, com os valores de AlfaCronbach e o número de itens correlacionados em cada componente, está representada no quadro 2. A maior parte dos valores de AlfaCronbach ficou acima

de 0,7, o que demonstra um alto grau de confiabilidade de correlação entre os itens. Os componentes foram interpretados com base nos modelos teóricos apresentados⁸ (HUGHES, 1993; ALDERSON; WALL, 1993; ALDERSON; HAMP-LYONS, 1996; WATANABE, 2004; GREEN, 2007).

Quadro 2. A estrutura fatorial do questionário

CONSTRUTOS	COMPONENTES	Itens
Conhecimento sobre o EPLIS ($\alpha=0.88$)	1. Conhecimento específico sobre o exame ($\alpha=0.88$)	7
	2. O exame em sala de aula ($\alpha=0.89$)	3
Preparação para o EPLIS ($\alpha=0.86$)	3. Atividades de compreensão oral e escrita e produção oral e escrita ($\alpha=0.93$)	6
	4. Atividades de produção oral relacionadas ao exame ($\alpha=0.76$)	8
	5. Impacto do exame no conteúdo – tópicos e habilidades ($\alpha=0.72$)	5
	6. Preparação individual focada no exame fora da sala de aula ($\alpha=0.73$)	3
	7. Preparação em grupo de natureza variada fora da sala de aula ($\alpha=0.70$)	4
Visões e Percepções sobre o EPLIS ($\alpha=0.74$)	8. A significância do exame ($\alpha=0.79$)	5
	9. Alinhamento entre ensino e avaliação externa ($\alpha=0.84$)	3
	10. Impacto negativo do exame ($\alpha=0.65$)	4

O primeiro componente, *Conhecimento específico sobre o exame*, correlacionou itens que envolvem as tarefas da fase I e II, o manual do candidato, a tabela de níveis de proficiência, além do objetivo e do conteúdo do exame. O segundo componente engloba itens sobre a frequência com que o exame é mencionado e discutido em sala de aula. No componente *Atividades de compreensão oral e escrita e produção oral e escrita*, atividades que abordam

8 Em Souza (2018), encontramos a descrição detalhada da análise e interpretação dos componentes principais do efeito retroativo do EPLIS nos alunos do curso de controle de tráfego aéreo.

temas do ambiente de trabalho se correlacionaram altamente. O quarto componente é composto por itens relacionados às tarefas presentes na fase II do teste e foi interpretado como *Atividades de produção oral relacionadas ao exame*. O quinto componente apresenta itens sobre a frequência com que as atividades em sala de aula se relacionam com o controle de tráfego aéreo e a aviação. O sexto e sétimo componentes correlacionaram atividades preparatórias fora da sala de aula, de natureza variada ou mais focadas no exame. O componente *A significância do exame* é composto por itens sobre a motivação para o estudo da língua e a percepção de relevância, autenticidade e confiabilidade do exame. O *Alinhamento entre ensino e avaliação externa* envolve itens sobre o grau de adequação da preparação recebida na EEAR. O último componente relaciona questões ligadas à ansiedade, ao estresse e à sobrecarga de estudos, provocadas pelo teste.

Na tabela 2, podemos comparar as médias de respostas dos alunos, divididos nos grupos de baixa e alta proficiência, às variáveis pertencentes a cada componente. N refere-se ao número total de respondentes em cada grupo. Na coluna 'M', estão as médias de respostas de cada subgrupo por componente e podem variar de 1 a 5, conforme escala *likert* adotada no questionário. SD é a abreviação em inglês de desvio padrão (*standard deviation*) e refere-se ao quanto as respostas se distanciam da média reportada. SEM é a abreviação em inglês para erro padrão da média (*standard error of the mean*).

Tabela 2. Média e desvio padrão das respostas dos subgrupos por proficiência

Construtos	Componentes	Grupos	N	M	SD	SEM	
Construto 1	Conhecimento específico sobre o exame	Baixa	113	2,413	0.838	0.078	
		Alta	155	2.748	0.791	0.063	
	O exame em sala de aula	Baixa	113	2.734	0.480	0.045	
		Alta	155	3.051	0.473	0.037	
Construto 2	Atividades de compreensão oral e escrita e produção oral e escrita	Baixa	113	2.991	1.004	0.094	
		Alta	155	3.348	1.035	0.083	
	Atividades de produção oral relacionadas ao exame	Baixa	113	2.991	1.004	0.094	
		Alta	155	3.348	1.035	0.083	
	Impacto do Exame no Conteúdo – tópicos e habilidades	Baixa	113	4.017	0.747	0.070	
		Alta	155	4.187	0.676	0.054	
	Preparação individual focada no exame fora da sala de aula	Baixa	113	1.512	0.785	0.073	
		Alta	155	1.511	0.893	0.071	
	Preparação em grupo de natureza variada fora da sala de aula	Baixa	113	2.316	0.816	0.076	
		Alta	155	2.416	0.892	0.071	
	Construto 3	Significância do exame	Baixa	113	3.157	0.907	0.085
			Alta	155	3.432	0.821	0.066
Alinhamento entre ensino e avaliação externa		Baixa	113	3.059	0.980	0.092	
		Alta	155	3.352	0.856	0.068	
Impacto negativo do exame		Baixa	113	3.404	0.818	0.077	
		Alta	155	3.190	0.880	0.070	

Percebemos, primeiramente, que as médias de resposta do grupo mais proficiente (Alta) tendem a ser maiores do que as médias do grupo julgado com baixa proficiência (Baixa) em grande parte dos componentes extraídos. Essa diferença é mais expressiva nos componentes *Conhecimento específico sobre o exame*, *O exame em sala de aula*, *Atividades de produção oral relacionadas ao exame*, *Significância do exame* e *Alinhamento entre ensino e avaliação externa*. Isso nos leva a supor que o grupo que declarou maior proficiência

conhece melhor o exame, realiza atividades preparatórias em sala de aula com mais frequência, considera o sucesso no exame mais importante, sente-se mais motivado e capaz de perceber melhor o alinhamento entre ensino e exame.

Por outro lado, o grupo menos proficiente atribui níveis mais baixos de conhecimentos sobre o exame, afirma realizar menos atividades de produção oral voltadas para o exame em sala de aula, não confere ao exame um alto nível de importância, além de reportar níveis mais altos de experiências negativas para com o teste quando comparado ao grupo mais proficiente.

Quanto à *Preparação individual focada no exame fora da sala de aula* e à *Preparação em grupo de natureza variada fora da sala de aula*, praticamente não houve diferença nas médias de respostas dos grupos com alta e baixa proficiência, indicando que não há relação entre a proficiência declarada e a frequência com que os alunos se engajam em atividades extraclasse. No entanto, vemos que a média para a preparação extraclasse de natureza variada é maior do que a média atribuída à preparação voltada para o formato do exame, sugerindo que os alunos tendem a se engajar mais em atividades variadas do que em atividades similares àquelas encontradas no teste. De qualquer forma, as médias em ambos componentes são consideradas baixas (entre 1,5 e 2,5) e denotam que, independentemente do nível de proficiência, os alunos costumam não se engajar muito em práticas preparatórias extraclasse.

Em relação ao *Impacto negativo*, percebemos que o grupo menos proficiente apresenta uma média maior, sugerindo que esses alunos se sentem mais ansiosos e pressionados pelo exame e possuem visões mais negativas sobre a presença do teste no cenário escolar.

A tabela 3 resume os resultados obtidos no teste *t* de *student* para amostras independentes. Como o desvio padrão entre os dois grupos é bastante semelhante em todos os componentes, assumimos que as variâncias são iguais. O *t*-valor e o Sig. (*p*-valor) indicam se a diferença entre as respostas dos grupos pode ser considerada significativa ou não. Como regra, quanto maior o *t*-valor, menor será o *p*-valor. Pelo fato de termos aplicado a análise de componentes principais aos dados, o *p*-valor necessitou ser reajustado de acordo com o número de componentes extraídos. Assim, para os componentes do construto 1, utilizamos como índice de significância *p*-valor ≤ 0.025 ; para o construto 2, $p \leq 0.010$; e para o construto 3, $p \leq 0.016$. O sinal de asterisco (*) na coluna Sig. (*p*-valor) marca que a diferença entre as médias foi considerada estatisticamente significativa. O sinal de negativo (-) nos valores de *t* (*t*-valor) assinala que a média do primeiro grupo é menor do que a média do segundo grupo, ao passo que o valor positivo indica o contrário.

Tabela 3. Resultados do teste t para o nível de proficiência

Construtos	Componentes	Subgrupos	N	SD	t-test para igualdade de médias		
					t (t-valor)	df	Sig. (2tailed) (p-valor)
Construto 1 p<= 0.025	Conhecimento específico sobre o exame	Baixa	113	0.838	-3.338	266	0.001*
		Alta	155	0.791			
	O exame em sala de aula	Baixa	113	0.480	-5.381	266	0.000*
		Alta	155	0.473			
Construto 2 p<=0.010	Atividades de compreensão oral e escrita e produção oral e escrita	Baixa	113	1.004	-2.584	266	0.005*
		Alta	155	1.035			
	Atividades de produção oral relacionadas ao exame	Baixa	113	1.004	-2.710	266	0.007*
		Alta	155	1.035			
	Impacto do Exame no Conteúdo – tópicos e habilidades	Baixa	113	0.747	-1.936	266	0.54
		Alta	155	0.676			
	Preparação individual focada no exame fora da sala de aula	Baixa	113	0.785	0.13	266	.990
		Alta	155	0.893			
	Preparação em grupo de natureza variada fora da sala de aula	Baixa	113	0.816	-0.936	266	0.818
		Alta	155	0.892			
Construto 3 p<=0.016	Significância do exame	Baixa	113	0.907	-2.590	266	0.010*
		Alta	155	0.821			
	Alinhamento entre ensino e avaliação externa	Baixa	113	0.980	-2.605	266	0.010*
		Alta	155	0.856			
	Impacto negativo do exame	Baixa	113	0.818	2.028	266	0.044
		Alta	155	0.880			

A diferença entre médias é considerada estatisticamente significativa em seis componentes: *Conhecimento específico sobre o exame*, *O exame em sala de aula*, *Atividades de compreensão oral e escrita e produção oral e escrita*, *Atividades de produção oral relacionadas ao exame*, *Significância do exame* e *Alinhamento entre ensino e avaliação externa*, indicados

pelo símbolo (*) na última coluna da tabela. Esses resultados confirmam as hipóteses levantadas anteriormente e sugerem que os alunos mais proficientes possuem níveis mais elevados de conhecimentos sobre o exame, se engajam com mais frequência em atividades de produção oral voltadas para o exame em sala de aula, sentem-se mais motivados, consideram mais importante a aprovação no teste, apresentam atitudes mais positivas em relação ao exame e percebem melhor o alinhamento entre o ensino e a avaliação externa. No entanto, no componente Impacto negativo, apesar de haver uma diferença expressiva entre as médias dos grupos, essa diferença não foi considerada estatisticamente significativa. Além disso, a preparação fora da sala de aula não manteve relação com o nível de proficiência declarado.

Pelo fato de os alunos serem agrupados em turmas por nível de proficiência a partir do 2º semestre do curso, esses resultados também levantam hipóteses de que a abordagem de ensino do professor se modifica em função da proficiência dos alunos. Dessa forma, podemos supor que, em turmas mais proficientes, os professores mencionam e discutem o exame com mais frequência, o que faz com que os alunos adquiram mais conhecimentos sobre o exame. Além disso, os professores podem atribuir um número maior de atividades de produção oral e compreensão oral similares às encontradas no teste em turmas mais avançadas linguisticamente.

Discussão dos dados

Os resultados da análise dos dados demonstraram que o fator “proficiência” aparenta exercer papel importante no processo de efeito retroativo, sugerindo que o impacto do teste é mais intenso para os alunos que se consideram mais proficientes e pertencem a turmas mais avançadas. Essa suposição pode estar ligada à percepção do nível de dificuldade do EPLIS por parte de alunos e professores, como sugere Green (2007). Assim, podemos presumir que os professores são mais propensos a discutir o exame e a engajar os alunos em práticas preparatórias com maior frequência em turmas mais proficientes, por acreditarem que existem chances reais de aprovação. Esses alunos, ao perceberem a possibilidade de serem bem-sucedidos, podem demandar de seus professores mais informações a respeito do teste, o que faz com que o seu nível de entendimento sobre o exame seja maior, bem como a percepção do alinhamento entre as aulas e as demandas do teste. A relevância atribuída ao exame também se eleva, juntamente com a motivação, e eles passam a apresentar atitudes mais positivas para com a avaliação externa, o que contribui, então, para a intensidade do seu efeito.

Por outro lado, em turmas menos proficientes, ao assumir que o nível do exame está muito além do nível dos alunos, os professores podem optar por não enfatizar, a fim, talvez, de evitar frustração. Esses alunos, por sua vez, tendem a se conformar com o fato de que o exame está acima de sua capacidade atual e acabam perdendo qualquer interesse, o que faz com que o efeito seja menor para esses participantes e nessas turmas.

Cheng *et al.* (2010) alcançaram resultados muito semelhantes a esses. Esses autores investigaram, estaticamente, o impacto de uma avaliação de produção oral em inglês, conduzida em sala de aula, nas percepções de alunos e pais. Essa avaliação tinha sido incorporada a um exame de alta-relevância no contexto educacional de Hong Kong, com um peso equivalente a 15%. A análise referente aos alunos apontou que eles eram afetados de maneiras diferentes dependendo do grau declarado de competência linguística, ou seja, alunos com níveis mais altos de proficiência na língua apresentavam maior motivação para o estudo da língua, realizavam mais atividades dentro e fora da sala de aula, possuíam visões mais positivas sobre exames externos e sentiam-se menos preocupados com os desafios impostos pelo exame quando comparados aos alunos com competência menor.

Stenlund, Eklöf e Lyren (2017) também analisaram as atitudes de alunos de diferentes níveis de proficiência em relação a um exame. Seus resultados evidenciaram que examinandos com baixo desempenho costumam reportar níveis mais altos de ansiedade e baixos níveis de motivação. O estudo também apontou que examinandos com alto desempenho tendem a se informar mais a respeito do teste e de seu conteúdo. Por outro lado, examinandos com baixo desempenho relataram, em maior medida, procurar por testes antigos a fim de conhecer o conteúdo cobrado.

As conclusões de Cheng *et al.* (2010) e Stenlund, Eklöf e Lyren (2017) se diferenciam das encontradas neste trabalho em relação às atividades preparatórias extraclasse e ao impacto negativo do teste. Não constatamos relação entre a proficiência e o grau com que os alunos se engajam em atividades preparatórias fora da sala de aula, tampouco no tocante à ansiedade e à pressão causadas pelo exame. Essa constatação pode indicar que, embora os alunos mais proficientes atribuam níveis mais altos à importância dada ao EPLIS, a percepção de relevância, de modo geral, pode ser ainda baixa no cenário investigado, o que faz com que o efeito fique restrito à sala de aula e a preocupação com o teste não seja tão alta para nenhum dos dois grupos. Essa percepção de baixa relevância, por sua vez, pode estar ligada ao fato de que, embora os resultados no EPLIS sejam validados e utilizados pelas autoridades do DECEA, eles não são ainda exigidos.

A pesquisa de Lanzoni (2004) também investigou a relação entre competência e efeito retroativo. No entanto, seus resultados contrastam em parte com os obtidos neste estudo, no sentido de que os alunos com alto desempenho disseram ser pouco ou nada afetados pelo exame, ao passo que os alunos com baixo desempenho relataram níveis mais altos de ansiedade. O autor explica que o teste analisado foi considerado muito fácil pelos alunos com alto desempenho, o que levou a um efeito retroativo fraco nesses participantes. Sobre isso, em relação ao EPLIS, podemos presumir que os mais proficientes consideram o exame em um nível desafiador, o que provocou mobilização, como pontuado por Green (2007).

Por fim, podemos concluir que o efeito retroativo é um fenômeno complexo mediado por inúmeras forças, como atestaram Alderson e Wall (1993), Hughes (1993), Alderson e Hamp-Lyons (1996), Watanabe (2004) e Green (2007). Dentre os fatores que atuam na determinação do efeito retroativo, esses autores enumeraram as características do teste, a percepção de sua relevância e dificuldade, sua proximidade, os conhecimentos adquiridos pelos envolvidos, suas crenças e valores. A essa lista, podemos acrescentar o nível de proficiência dos alunos como elemento constituinte do fenômeno e de sua intensidade.

Considerações finais

Este estudo demonstrou que o nível de proficiência de alunos de um curso de formação em controle de tráfego aéreo é um fator importante a ser considerado no processo de efeito retroativo do EPLIS. A intensidade do impacto do exame nas percepções e atitudes dos participantes variou significativamente de acordo com o nível de proficiência declarado por eles, sendo maior para os que se consideraram mais proficientes. Essa associação pode estar relacionada às características individuais dos alunos que, ao se autoavaliarem como mais proficientes, percebem a possibilidade de obter um bom resultado, apresentando, assim, atitudes mais positivas com o exame externo. Da mesma forma, a proficiência dos alunos pode influenciar as práticas dos professores, fazendo com que o impacto do exame seja mais intenso em turmas mais proficientes.

Embora essas inferências sejam embasadas por análises estatísticas rigorosas e corroborem estudos anteriores, elas precisam ser contrastadas com dados qualitativos, obtidos por meio de entrevistas com professores e alunos, bem como observações *in loco*. Estudos exclusivamente quantitativos, apesar de tecerem informações consistentes sobre as relações entre os fatores e o efeito retroativo, não oferecem uma compreensão mais detalhada sobre a complexidade do fenômeno, além de se apoiarem somente no que foi relatado pelos participantes.

É necessário, portanto, verificar o que ocorre nas salas de aulas que faz com que alunos de níveis distintos de proficiência percebam o exame diferentemente e como os professores colaboram nesse sentido. Como dito anteriormente, esses dados fazem parte de uma pesquisa de doutorado, que contempla também dados qualitativos com o intuito de expandir, explicar mais profundamente e corroborar essas alegações iniciais, contribuindo, assim, para um melhor entendimento sobre como o efeito retroativo do EPLIS atua nas percepções, atitudes e ações de alunos.

REFERÊNCIAS

ALDERSON, J. C.; WALL, D. Does washback exist? *Applied Linguistics*, v. 14, n. 2, p. 115-129, 1993.

ALDERSON, J. C.; BANERJEE, J. Language testing and assessment – Part I. *Language Teaching*, v. 34, n. 4, p. 213-236, 2001.

ALDERSON, J. C.; HAMP-LYONS, L. TOEFL preparation courses: a study of washback. *Language Testing*, v. 13, n. 3, p. 281-297, 1996.

BACHMAN, L. F.; PALMER, A. S. *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press, 1996.

BAILEY, K. Working for washback: A review of the washback concept in language testing. *Language Testing*, v. 13, n. 3, p. 257-279, 1996.

BURROWS, C. Washback in Classroom-Based Assessment: A study of the Washback Effect in Australian Adult Migrant English Program. In: CHENG, L.; WATANABE, Y.; CURTIS, A. (ed.). *Washback in Language Testing: Research, Contexts and Methods*. New Jersey: Lawrence Erlbaum Associates, 2004. p. 113-128.

CHENG, L. Changing assessment: washback on teacher's perspectives and actions. *Teaching and Teacher Education*, v. 15, n. 3, p. 253-271, 1999.

CHENG, L.; ANDREWS, S.; YU, Y. Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, v. 28, n. 2, p. 221-249, 2010.

GREEN, A. *IELTS washback in context: Preparation for academic writing in higher education*. Studies in Language Testing 25. Cambridge, UK: Cambridge University Press and Cambridge ESOL, 2007.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. *Análise Multivariada de Dados*. 6. ed. Porto Alegre: Bookman, 2009.

HUGHES, A. *Backwash and TOEFL 2000*. Unpublished manuscript. Reading, UK: University of Reading, 1993.

LANZONI, H. DE P. *Exame de proficiência em leitura de textos acadêmicos em Inglês: um estudo sobre efeito retroativo*. 2004. Tese (Doutorado em Linguística Aplicada) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2004.

MCNAMARA, T.; ROEVER, T. *Language Testing: The social dimension*. Malden, MA: Blackwell, 2006.

MESSICK, S. Validity and washback in language testing. *Language Testing*, v. 13, n. 3, p. 241-255, 1996.

ORGANIZAÇÃO DA AVIAÇÃO CIVIL INTERNACIONAL/OACI. *Manual on the implementation of ICAO language proficiency requirements: Doc. 9835 AN/453*. 2. ed. Montreal, 2010.

SCARAMUCCI, M. V. R. Vestibular e ensino de língua estrangeira (Inglês) na escola pública. *Trabalhos de Linguística Aplicada*, v. 34, p. 7-20, 1999.

SCARAMUCCI, M. V. R. Efeito Retroativo da Avaliação no Ensino/Aprendizagem de línguas: o estado da Arte. *Trabalhos de Linguística Aplicada*, v. 43, n. 2, p. 203-226, 2004.

SCARAMUCCI, M. V. R. Validade e consequências sociais das avaliações em contextos de ensino de línguas. *Linguarum Arena*, v. 2, p. 103-120, 2011.

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. Relatório final de pesquisa. Campinas: UNICAMP, 2010.

SOUZA, P. R. O efeito retroativo do EPLIS nas percepções de professores. *Brazilian English Language Teaching Journal*, v. 8, n. 1, p. 5-29, 2005. Disponível em: <http://revistaseletronicas.pucrs.br/ojs/index.php/belt/issue/view/1150>. Acesso em: 25 jun. 2017.

SOUZA, P. R. Os componentes determinantes do efeito retroativo de um exame de proficiência em inglês aeronáutico. In: SCARAMUCCI, M. V. R.; TOSQUI-LUCKS, P.; DAMIÃO, S. M. *Pesquisas sobre Inglês Aeronáutico no Brasil*. Campinas: Pontes, 2018. p. 221-242.

SHOHAMY, E. *Language Policy: Hidden agendas and new approaches*. New York: Routledge, 2006.

STENLUND, T.; EKLÖF, H.; LYREN, P. Group differences in test-taking behavior: an example from a high-stakes testing program. *Assessment in Education*, v. 24, n. 1, p. 4-20, 2017.



STONEMAN, B. *The impact of an exit English test on Hong Kong undergraduates: a study investigation of the test' status on students' test preparation behaviours*. 2006. Unpublished PhD Dissertation. Hong Kong Polytechnic University: Hong Kong, China, 2006.

TOSQUI-LUCKS, P. T. *et al.* Ensino e avaliação de Língua Inglesa para controladores de tráfego aéreo como requisito de segurança em voo. *Revista Conexão SIPAER*, v. 7, n. 1, p. 44-54, 2016. Disponível em: <http://conexaosipaer.cenipa.gov.br/index.php/sipaer/issue/view/19>. Acesso em: 25 jun. 2017.

WATANABE, Y. Teacher factor mediating washback. *In*: CHENG, L.; WATANABE, Y.; CURTIS, A. (ed.). *Washback in Language Testing: Research, Contexts and Methods*. New Jersey: Lawrence Erlbaum Associates, 2004. p. 129-146.