

The washback of EPLIS on teachers' perceptions and actions: Implications for reviewing ICAO language policy in Brazil

O efeito retroativo do EPLIS nas percepções e ações de professores: Implicações para uma revisão da política linguística da OACI no Brasil

Paula Ribeiro e SOUZA (UNICAMP/ICEA)¹

ABSTRACT

This study investigated the washback effect of the Aeronautical English Language Proficiency Exam for the Brazilian Airspace Control System (EPLIS) on teachers' perceptions and actions in an Air Traffic Control Initial Training Program. EPLIS has been administered to Brazilian in-service air traffic controllers and aeronautical station operators every year since 2007, in compliance with the International Civil Aviation Organization (ICAO) language proficiency requirements, published in 2003. These requirements set that all professionals involved in international flight operations shall demonstrate a minimum level of proficiency in the English language. In 2014, Brazilian air traffic control authorities decided to extend EPLIS application to pre-service air traffic controllers to supposedly improve learners' proficiency in the training program. However, the consequences of this decision, whether intended or unintended, positive or negative, had not been appraised yet. Quantitative and qualitative data were collected through questionnaires, interviews and class observations. The results showed that the decision to introduce EPLIS in that educational context actually increased its impact. However, some teachers' deficiencies in understanding the exam and its demands, along with an underrepresentation of air traffic control tasks in the test, compromised to some extent the intended effects and need to be handled so that ICAO language policy in Brazil can be entirely implemented.

keywords: Washback effect, Proficiency test, Aviation English, Language Policy

RESUMO

Este estudo investigou o efeito retroativo do EPLIS (Exame de Proficiência em Inglês Aeronáutico do Sistema de Controle do Espaço Aéreo Brasileiro) nas percepções e ações de professores em um curso de formação em Controle de Tráfego Aéreo. Desde 2007, o EPLIS tem sido aplicado, anualmente, a controladores de tráfego aéreo e operadores de estação aeronáutica em serviço, em cumprimento aos requisitos de proficiência linguística estabelecidos pela OACI (Organização Internacional da Aviação Civil), nos quais os profissionais envolvidos em operações de voo internacionais devem demonstrar um nível mínimo de proficiência na língua inglesa. Em 2014, supostamente, com o objetivo de aumentar o nível de relevância do exame neste contexto educacional, decidiu-se estender a aplicação do EPLIS aos profissionais ainda em processo de formação, alavancando, assim, mudanças curriculares que visassem

¹ University of Campinas (UNICAMP), Campinas, São Paulo, Brazil and Airspace Control Institute (ICEA), São José dos Campos, São Paulo, Brazil. ORCID: <https://orcid.org/0000-0001-7202-531X>; paulalettas@gmail.com

à elevação do nível de proficiência desses alunos. Todavia, as consequências dessa decisão, sejam elas pretendidas, não pretendidas, positivas ou negativas não haviam sido, até então, analisadas. Dados quantitativos e qualitativos foram coletados por meio de questionários, entrevistas e observações de aulas. Os resultados demonstraram que a decisão de inserir o exame no contexto escolar realmente favoreceu o aumento de seu impacto, conforme previsto. Entretanto, deficiências no entendimento do teste e de suas exigências pelos professores, acrescidas de uma subrepresentação de tarefas pertencentes ao controle de tráfego aéreo, comprometeram de alguma forma os efeitos desejados, e necessitam serem revistas para que a política linguística da OACI seja totalmente implementada.

Palavras-Chave: *Efeito retroativo, Exame de proficiência, Inglês aeronáutico, Política linguística*

1. ICAO language policy in Brazil

English has officially been set as the common language in aeronautical communications since 1944, when the Treaty on International Civil Aviation, widely known as Chicago Convention, was signed by 52 States. However, in 1998, the International Civil Aviation Organization (ICAO)² decided to strengthen the existing provisions related to English language proficiency and urged the Air Navigation Commission to consider the matter with a high level of priority. This decision was based on reports of several aeronautical accident and incident investigations that pointed out the insufficient English language proficiency on the part of pilots and air traffic control personnel as having played a causal or contributing role (ICAO, 2010).

Then, in 2003, ICAO introduced a set of language proficiency requirements (LPRs) to ensure that the pilots, air traffic controllers and aeronautical station operators, involved in international flight operations, are proficient in the use of English in order to handle complicated situations. A language proficiency rating scale with six levels (from elementary to expert) was published. Level 4 (Operational level) was considered the minimum language proficiency for international operations. Moreover, intervals for retesting were also established according to the level obtained.

Each ICAO Member State³ should take over the responsibility for the development and administration of their own testing programs or for acquiring an external test. In order to support the countries to comply with the new provisions for language proficiency, in 2004 ICAO published the first edition of the Document 9835/AN453 - The Manual on the Implementation of the ICAO language proficiency requirements.

In Brazil, the Airspace Control Institute (ICEA) was commissioned by the Airspace Control Department (DECEA) to plan, develop, and administer teaching and testing programs regarding the ICAO language proficiency requirements for Brazilian air traffic controllers and aeronautical station

² ICAO is a United Nations regulatory agency created in 1944 to manage the administration of the Convention on International Civil Aviation.

³ Countries that signed the Convention on International Civil Aviation.

operators (TOSQUI-LUCKS et al., 2016). In compliance with the ICAO language policy, ICEA designed the Aeronautical English Proficiency Exam for the Brazilian Airspace Control System (EPLIS) in 2007 and, since then, has administered it annually to in-service professionals. However, in 2014, the test application was extended to pre-service air traffic controllers in order to supposedly increase the stakes of the test at initial training contexts and, consequently, improve learners' proficiency.

Aviation language tests, such as EPLIS, are considered very high stakes since their results inform important decisions that bring about serious consequences not only for test takers, but also for the organizations involved and society at large. At the individual level, ICAO (2010) mentions that pilots and air traffic controllers may be denied a license to operate internationally if they do not show compliance with the LPRs. In the State level, besides the fact that organizations cannot afford to lose competent staff, considerable amounts of money have already been allocated for language training and testing. In addition, Aragão (2018) points out that language proficiency has become one of the items assessed in the ICAO's Universal Safety Oversight Audit Program⁴, in which non-conformity may result in political and financial losses to a country. For society, an increase in language proficiency reduces the risk of miscommunication, which positively impacts on flight safety.

DECEA has adopted an incentive plan for air traffic controllers and aeronautical station operators who achieve level 4 or above in EPLIS. The benefits include flexibility in shift swaps and reduced working hours. However, it is expected that language proficiency becomes a restrictive factor in issuing and renewing the professional license of these professionals. DECEA has also taken risk mitigation measures to ensure the excellence in the international air traffic control delivery service. In the operations management area, for instance, a minimum number of professionals who have level 4 or above in EPLIS has been kept in each shift and shift swaps have been closely monitored to guarantee that this condition remains.

The introduction of a test by a State is a conscious decision that it is capable of triggering and accelerating changes in teaching and learning. Shohamy (1993, p.1) argues that "few devices are as powerful, or are capable of dictating as many decisions, as tests", especially when their stakes are high. ICAO language policy is not an exception. Thus, aeronautical English tests are used not only as indicators of proficiency, but also as "instruments capable of dictating what will be studied, learned and taught by students and teachers" (SHOHAMY, 2006, p. 103). The relationship between testing and teaching and learning is known as washback effect.

⁴ This program assesses whether the State has effectively and consistently implemented ICAO safety-related Standards and Recommended Practices.

This study investigated whether the decision to introduce EPLIS in an initial training course for Brazilian air traffic controllers promoted the intended washback or whether there were unexpected or even negative effects associated with the test. The results informed how effective the ICAO language policy has been in Brazil and suggested initiatives that should be taken for its continuous improvement.

2. The washback of EPLIS on an initial training program

2.1 The definition of washback

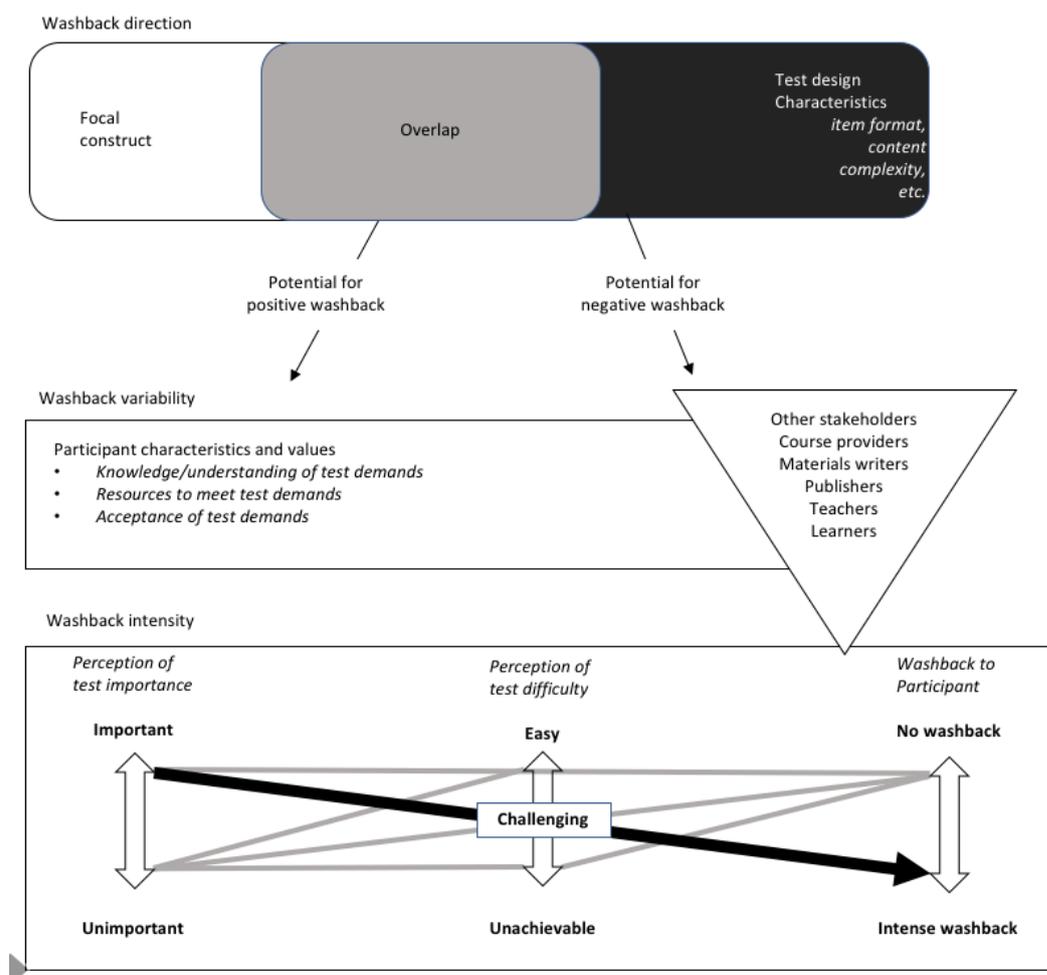
Up to the 1990s, washback was understood as having a deterministic nature. In other words, it was believed that a bad test would bring about negative effects whereas a good test would produce beneficial effects. Alderson and Wall (1993) argue that this is a very naive view as it does not take account of the numerous factors that may influence events in a classroom. In this regard, Messick (1996, p. 242) says that “a poor test may be associated with positive effects and a good test with negative effects because of other things that are done or not done in the educational system”. Studies about washback such as Alderson and Wall (1993), Cheng (1999), Scaramucci (1999) and Watanabe (2004) confirmed the complex and multifaceted nature of the phenomenon and in order to explain its process Green (2007) proposed the following model (Fig. 1), in which washback direction, variability and intensity are integrated.

In this model, washback direction refers to what extent the test reflects the construct it intends to measure. Authenticity and directness are characteristics that make a test increase its chance of producing beneficial effects (FREDERICKSEN; COLLINS, 1989; MESSICK, 1996). Authenticity means that the tasks in the test are representative of the target language use. Being authentic ensures that nothing relevant to the construct is excluded of the assessment process, reducing the risk of construct underrepresentation. Direct tests, on the other hand, allow that nothing irrelevant be added to the test, avoiding interference and contamination of the construct, known as construct-irrelevant variance. In Green’s proposal, it can be observed that the greater is the overlap between the construct to be assessed and the test characteristics (items, content, skills), the greater is the power of the test to produce positive effects.

Washback variability is defined as the characteristics and values of the stakeholders involved in the washback process. Green (2007) underpins three different variables, previously mentioned by Hughes (1993) and Alderson and Hamp-Lyons (1996): the extent to which participants understand test demands, the extent to which they accept test demands and the resources available to meet them. Washback intensity refers to the extent to which participants adjust their actions to the test demands (WATANABE, 2004; CHENG, 2005). Participants’ perceptions about test importance will also

determine the degree of engagement in preparatory activities. Thus, the higher is the test importance to test users, the higher is its influence. Another factor included in the model is test difficulty. If participants perceive the test as too easy, they will not pay any attention to it. The test demands are also ignored if they are seen as intangible by teachers and students.

Figure 1. Washback model



Source: GREEN, 2007, p. 24

2.2 EPLIS characteristics

The purpose of EPLIS is to assess the use of plain English in unexpected, unusual and emergency situations that may happen in air traffic control (ATC) delivery service. The test assesses speaking and listening skills and its content includes themes, topics and functions related to aeronautical communications. As stated in Scaramucci et al. (2010), EPLIS can be considered a performance test with specific purposes as operational knowledge required in ATC interacts with language knowledge in communicative situations.

The test comprises two papers. Paper 1 is a listening comprehension computer-based test. There are 30 multiple choice items and each item refers to a pilot-controller transmission. The test taker must score 21 in order to take Paper 2. Paper 2 assesses integrated listening and speaking skills in an interview format. It is divided into four parts. In part 1, the interlocutor asks the test taker four questions related to ATC career. In part 2, ten complicated ATC situations are read aloud. For each situation, the test taker is asked to explain it and then to make a suggestion to solve the problem. In part 3, four aviation-related questions are asked. In the last part, the test taker describes a picture and tells a story about it.

Paper 2 is conducted with one test-taker at a time and lasts about 15 to 30 minutes. Test takers' performance is assessed by two raters: the interlocutor who gives a holistic score to it and a second rater who receives the interview audio file and assesses it analytically according to ICAO Rating Scale.

2.3 Expected washback according to ICAO

ICAO document 9835 AN/453 (2010) states that test designers have the responsibility to develop a test that fosters positive washback such as encouraging training programs to focus on building learners' communicative competence, instead of triggering memorization techniques and test preparation practices. However, as washback is a phenomenon influenced by several factors, unexpected and undesirable effects may occur. In this regard, studies that investigate not only whether the expected washback is being produced, but also the existence of negative effects so that they can be controlled and minimized are essential.

In order to assist countries with the LPRs, ICAO (2010) presents a few parameters to guide the design of aviation aeronautical English tests: a) the test should focus on listening and speaking; b) the test should assess plain language⁵ only; c) test content should be work related; d) test tasks should resemble real life operational activities and focus on unexpected, unusual and emergency situations; and e) intelligibility in an international community of users should be emphasized by all means. Furthermore, ICAO suggests that performance testing would be the most appropriate method to assess language proficiency in aeronautical communications.

Based on these test characteristics, it is expected that training programs will adopt a communicative approach to meet learners' needs and help them deal appropriately with the language constraints in aeronautical communications. The emphasis will be on the development of oral skills to

⁵ ICAO uses the term plain language to differentiate it from phraseology. Whereas phraseology is a set of pre-coded phrases to be used in routine situations, plain language refers to the spontaneous use of a language, although constrained by the functions and topics found in aeronautical communications.

perform language tasks demanded in problematic situations. Interaction will be further explored in class, as well as communicative strategies to enable learners to solve potential misunderstandings.

EPLIS intended effects are not directly mentioned by its designers on the test official website. However, as EPLIS is claimed to have been developed according to ICAO standards, it can be assumed that it will promote the washback described in the document 9835/AN453 and discussed previously.

3. Methodology

3.1 Research scenario

The study was carried out in an ATC Initial Training Program, at a Brazilian Air Force Organization, in 2017, three years after EPLIS had been implemented into it. The program consisted of a two-year intensive course, divided into four terms, and its syllabus contained about 500 hours of English lessons. At the end of grade four, students commonly took EPLIS. The test was mandatory and its results were officially accepted by national air traffic control authorities. This means that learners' scores were published onto their professional operational licenses. However, failing the test did not prevent students from graduating and doing their internship.

3.2 Participants

Sixteen out of 21 English teachers answered the questionnaire. Twelve held at least a major in English and 4 were air traffic controllers who had taken a TEFL course. Four teachers agreed to participate in interviews and have their classes observed. There were 407 students taking the ATC course at the time of the study. Students' age ranged from 17 to 24 years.

3.3 Instruments and analyses

The study comprised two phases. The first phase focused on gathering quantitative data. To this end, a 65-item questionnaire was designed. Most items were written on a five-point Likert scale and addressed five different areas: general information; knowledge about EPLIS; test impact on teaching; test preparation; as well as views and perceptions of the test impact. Descriptive statistics was used to identify possible trends and patterns in the data.

In the second phase, qualitative data was obtained through class observations and interviews. The aim was to explain, expand and corroborate the initial results. An observation sheet was designed to guide the researcher's attention to possible areas affected by the test and follow-up interviews were conducted as soon as a set of four classes had been observed. The topics discussed in the interviews included the issues covered by the questionnaire in phase 1 and the events that had been taken note of in

the observation sheet. The interviews were conducted in Portuguese, audio recorded and translated into English. Each interview took from 15 minutes to 1 hour and 42 minutes with the average of 44 minutes. The interview data were thematically analyzed and combined with the questionnaire and the class observations in a triangulation approach. In total, there were 16 responses to the questionnaire, 12 interviews and 52 hours of class observation.

4. Results

4.1 Test washback on teaching

In the first phase of the study, teachers were asked to analyze the influence of EPLIS on the ATC training program as well as on their teaching. Their answers ranged from 1 to 5, in which 1 meant very low influence, 2 low influence, 3 medium influence, 4 high influence and 5 very high influence. The results are shown in Table 1.

Table 1. Aspects of washback on teaching

	1	2	3	4	5
EPLIS influence on teacher's methodology	0.06	0.19	0.00	0.62	0.12
EPLIS influence on content knowledge	0.00	0.19	0.06	0.62	0.12
EPLIS influence on the use of textbooks	0.00	0.25	0.12	0.56	0.06
EPLIS influence on classroom activities	0.00	0.00	0.06	0.75	0.19
EPLIS influence on the course achievement tests	0.06	0.12	0.25	0.56	0.00
EPLIS influence on the course projects	0.00	0.06	0.06	0.69	0.19

The figures suggest that the great majority of teachers agreed that EPLIS has had a significant impact on what is taught and assessed at the ATC course as well as on what happens inside the classroom. These findings can be contrasted with the ones found in a baseline study carried out in 2014 (SOUZA, 2017), right before learners started sitting the test. Back then, the answers of the teachers to a similar but shorter questionnaire showed that the test had little or no impact on their teaching practice. Three years later, only 6% of teachers disagreed on the test influence upon their classroom methodology.

In relation to classroom activities, no one showed disagreement about the influence of EPLIS on the tasks carried out in class with students. In 2014, around 84% of teachers described this impact as low or very low. The present study also indicates a great influence of the test on the training projects with 94% of teachers answering high or very high to this item. In relation to the achievement tests, 56% of them defined the test impact as high and 25% as medium. A similar question in the baseline study

showed that 50% of teachers did not see any washback of EPLIS on the assessment system.

These results indicate that the test has increasingly changed various course and teaching aspects since its implementation in the ATC training program, such as the teachers' methodology, the content and textbooks used in the classroom, as well as the tasks, the projects and the assessment system. However, the data still show cases of disagreement on the influence of EPLIS, suggesting that the test has impacted some teachers more than others.

In an interview with teacher and pedagogical coordinator Pilar⁶, she said that, in her meetings with the teachers, the importance of developing learners' listening and speaking skills was always emphasized as they are essential for ATC delivery service. Another interesting point made by Pilar was the need to maximize learners' talking time. Promoting students' engagement in interactive activities either in pairs or in groups is a principle of the communicative approach as well as an expected effect of ICAO language policy. Here is an extract of Pilar's speech:

I always eh: at the beginning of every course I call a meeting with teachers to say what is expected from them (+) what they have to do (+) the types of tasks they have to propose, (+) we have to support oral production and listening activities, (+) it's necessary to do a speaking activity /.../ I ALWAYS focus on the importance of both skills (+) listening and speaking /.../ I always TELL them that we need to push students to speak (+) try to reduce teacher talking time (+) let students speak⁷ (Interview with Pilar, a teacher and pedagogical coordinator at the ATC course since 2005).

When asked about the influence of EPLIS on teaching, Pilar stated that the test had certainly impacted on her classes since students started taking it. She explained that some activities that she did not use to do very often with the students became more frequently performed in her practice, such as 'telling or retelling a story'. In her own words: "I ask students to: retell what they've heard (+) that's what I used to do less (+) after Eplis started to be taken here at school, I began doing that more often".

Teacher Eduardo also said that the influence of EPLIS on his teaching became higher. He attributed this increase to the knowledge he acquired about the test in the last couple of years and added that the test guided his classes, especially, in relation to the types of tasks carried out with students. Eduardo went on to say that the test was the reason why he considered doing picture description activities in the classroom: "I wouldn't normally do (+) when I entered here I wouldn't do 'describe the picture' (+) I do it today because I know it's part of Eplis". This type of behavior is what defines washback.

⁶ The names used in this article are pseudonyms in order to guarantee the anonymity of the participants.

⁷ The interview material was transcribed using the codes proposed by Marcuschi (2008).

Teacher Ana was also able to notice the impact of EPLIS on the activities she carried out with students. She explained that if it were not for the knowledge she had built about the test, she wouldn't feel obliged to explain to students how to describe a picture and tell a story.

It influences it (+) we feel concerned about / how to explain to students that part (+) how to describe pictures (+) and then how to tell a story (+) watch out for not mixing up the two things /.../ I wouldn't tell that student today to be careful (+) when describing a picture (+) just describe it (+) so, it is because I know a little bit about the test (+) about Eplis (+) that I'm able to guide her in that way (Interview with Ana, teacher at the ATC course since 2013).

In the teachers' reports, a great emphasis on picture description and storytelling activities was noticeable, which may suggest that these aspects of EPLIS are particularly more salient than others. This type of washback effect is seen as negative because it can lead to the narrowing of learning opportunities in the classroom.

Class observations revealed that, in relation to skills, oral comprehension was extensively explored in the classroom through pilot-controller communications, which shows that one of the expected effects has occurred. On the other hand, oral production activities were far less often. Reading tasks seemed to take on greater importance, contradicting what was intended.

In terms of oral production, picture description and storytelling were indeed overemphasized in the classroom, whereas reporting ATC situations and making suggestions (EPLIS Part 2 of Paper 1) were rarely mentioned to students. Furthermore, pilot-controller role play proposals, which were easily found in the course textbooks, were overlooked by teachers most of the time. It seemed that teachers considered this type of activity unimportant to students due to the fact it was not covered in EPLIS. Teacher Ana clearly explained why she focused less on role plays compared to picture description and storytelling:

It is not further explored (role plays) because we don't see the importance of doing so (+) neither in Eplis nor in the course achievement tests (+) thus' (+) when there's time left (+) you do it (+) but only when there's time available (Interview with Ana, teacher at the ATC course since 2013).

However, Ana seemed to be aware that a role play was a more adequate task to carry out with students as it resembled air traffic control activities: "we tend to focus more on describing and telling the story (+) because of Eplis (+) because we know that Eplis doesn't cover roleplays (+) even though it is real life".

Besides being one of intended effects of ICAO language policy, the management of pilot-controller dialogue, plus the development of negotiation strategies are a vital feature of aeronautical communications and should be well represented in the test. The failure of representing these important aspects of the construct in the test jeopardizes its validity as well as the test washback. The overemphasis placed on certain parts of the test at the expense of others is another negative aspect as it may narrow down the curriculum and the teaching methods, as pointed out earlier.

4.2 Washback variability

According to Green (2007), washback variability is related to participants' knowledge and values towards the examination. In the questionnaire, teachers were asked to rate how much they knew about EPLIS, its characteristics and ICAO language policy. Their answers are placed in Table 2, in which N means the number of respondents, the mean ranges from 1 to 5 and SD stands for standard deviation.

Table 2. Knowledge of the test

	N	Mean	SD
General knowledge	16	4.000	0.966
ICAO Doc 9835	16	3.187	1.167
Test purpose	16	4.187	0.834
Test content	16	4.065	0.853
Tasks in Paper 1	16	4.065	0.928
Tasks in Paper 2	16	4.065	0.997
ICAO Rating Scale	16	4.125	0.885
Operational Level 4	16	4.062	0.853

The figures show that teachers rated positively their knowledge about EPLIS. All items reached an average higher than 4, with the exception of the item 'ICAO Document 9835'. In percentage, teachers' answers to their general knowledge of the test were concentrated in the high level (51%) and very high level (31%), showing a quite different scenario from that of 2014, when 58% of teachers stated that they had little or no knowledge of the test. These findings suggest that the degree of participants' familiarity with EPLIS has increased, probably because of its implementation in the training program.

However, the standard deviation close to 1 point cannot go unnoticed as it indicates heterogeneity within the group of teachers. Looking closely at their answers, teachers who held a major in English rated their test knowledge between 3 (medium) and 4 (high), whereas air traffic controllers'

responses varied between 3.5 (medium to high) and 4.7 (close to very high). In addition, answers in the levels 'low' and 'medium' were found within the first group of teachers. This difference suggests that there may be gaps that need to be bridged in teachers' knowledge about the test. Regarding this matter, although EPLIS designers have improved the access to information about the test, only a two-day workshop had been offered to teachers. Teacher Eduardo, when asked about his knowledge of the test, said:

Much more than before because now I know MORE than before, (+) despite not knowing (+) as much as I'd like to (+) because I'm not an Eplis rater (+) I've never done a course (Interview with Eduardo, a teacher at the ATC course since 2010).

Pilar, teacher and pedagogical coordinator, explained that her role as a coordinator was more managerial than pedagogical and what she did in terms of guidance in the test was advise teachers, especially the ones who were teaching four graded students, to read the test taker's Manual and visit EPLIS website so that they could obtain further and accurate information about the examination.

The role of the coordinator at the Atc course is to manage the distribution of classes among teachers (+) the replacements (+) some changes in textbooks (+) management in general (+) how things are going to happen (+) the logistics (Interview with Pilar, a teacher and pedagogical coordinator at the Atc course since 2005).

It can be argued that what has been done in terms of pedagogical orientation is not enough compared to ICAO language policy ambitions regarding teaching and learning innovations. The lack of a more effective guidance possibly leads to a variation among teachers as they will have their own interpretation of the test, its purpose and demands. Moreover, deficiencies in teachers' knowledge of the test may be the reason why some test aspects are overemphasized in the classroom, certainly the ones teachers are more acquainted with.

In an interview with teacher Ana, conducted after having her classes observed, she confirmed that her knowledge of EPLIS derived from the test taker's Manual as well as from her interaction with more experienced workmates. At the end of her speech, though she claimed that she had a good knowledge of the test, she clarified that it did not come from specific training provided by official organizations, but rather from informal situations.

I know Eplis by reading the test taker's Manual that we have here (+) and by talking to the teachers who had already taught four graded students /.../as soon as I started working here she (teacher and pedagogical coordinator Pilar) seated with me and

carefully explained (+) what the test is like (+) how it works (+) so, it was nice (+) I think I got to know Eplis well (+) but'(+ through people who had already worked with it (Interview with Ana, teacher at the ATC course since 2013).

Later, when asked if she conducted other test-related activities in the classroom, apart from picture description and storytelling, Ana admitted that, despite having read the test taker's Manual, she might not fully know about the test demands. She also added that the recommendations given to the group of teachers aimed at developing learners' competence in describing pictures and telling stories, but she would feel willing to attempt something new if the pedagogical coordination instructed her to do so.

I only receive written instructions on how it should be (+) I EVEN read the test taker's manual (+) but it comes from the pedagogical coordination /.../ if they bring this part (EPLIS part 2 Paper 2) and present it as a suggestion to introduce in the classes (+) it would be very nice, (+) bring problematic situations to be solved (Interview with Ana, teacher at the ATC course since 2013).

Teacher João, who is an air traffic controller, also commented that his knowledge of EPLIS had been built on the test taker's Manual, on the information available on the test website as well as on sharing experiences with other teachers. It is clear in his speech that the changes to the website have positively impacted on his search for further information.

My knowledge of Eplis has improved (+) I began reading the test taker's Manual (+) which is available in the website (+) and Icea has also made more resources available lately (+) with more paper 1 mock tests (+) the manual has also been broken down (+) including more details (+) so I started to increase my knowledge (+) also by sharing experience with other teachers (Interview with João, a teacher at the Atc course since 2016).

Based on these reports, it can be inferred that a great deal of teachers' knowledge of EPLIS is due to an individual effort. Even though improvements have been made regarding the stakeholders' access to information about EPLIS, they are considered minor initiatives compared to the test's high stakes. As other washback studies have shown (ALDERSON; HAMP-LYONS, 1996; CHENG, 1999; SCARAMUCCI, 1999; WATANABE, 2004; QI, 2004; RETORTA, 2007), the test by itself is not able to entirely introduce the intended changes. Thus, a lack of continuous training courses to explain to the teachers the test purposes, its characteristics and demands may produce distorted views or even myths about the test, compromising its effects.

4.3 Washback intensity

In Green's model (2007), washback intensity is determined by the test importance and its level of difficulty. In terms of test importance to students, teachers were asked to rate some statements in an agreement scale from 1 to 5, in which 1 meant 'strongly disagree' and 5 'strongly agree'. The results are shown in Table 3.

Table 3. Teachers' views about EPLIS

	1	2	3	4	5
The test motivates students	0.00	0.06	0.12	0.50	0.31
The test increases learners' language proficiency	0.00	0.00	0.12	0.38	0.50
The success in the test is important for the ATC job	0.00	0.00	0.25	0.50	0.25
The test application should be optional	0.25	0.56	0.00	0.19	0.00

According to the figures, most teachers saw the test as a beneficial tool, able to motivate students to study English and boost their proficiency level. In addition, they agreed that the test should continue being mandatory because of its positive effects. During the interviews, teachers expressed that EPLIS made students more aware of the importance of the language to their future profession. Teacher Isabel, for instance, commented that students were taking the English classes more seriously after they started sitting EPLIS. Teacher João also mentioned that the test made students dedicate themselves to studying the language:

the fact that Eplis is administered for four graded students (+) at least they have become aware that they need to study (+) they are studying something (+) something about aeronautical English they are studying. (Interview with João, a teacher at the ATC course since 2016).

Although teachers claimed that the test increased students' awareness of the importance of English in ATC, the achievement tests were considered higher stakes than EPLIS. Teacher and pedagogical coordinator Pilar explained that students' final average score, obtained in the tests taken during the course, determines their position in line to choose where they are going to work after graduating. Besides, the students who are not able to get the minimal score in any test may have their course enrollments cancelled. Differently, students' scores in EPLIS, though accepted by Brazilian ATC authorities as an official result, have not been demanded yet. It means that they can retake the test the

following year without facing any penalty. Thus, the course achievement tests have more serious consequences to students than EPLIS. Pilar summed up saying: “the course tests are FAR more important (+) /.../ they determine where the student is going to live (+) they determine LIFE”. EPLIS washback intensity undoubtedly decreases for students as its results have no immediate consequence to them.

The course achievement tests assess listening and reading skills. The listening test resembles EPLIS Paper 1, in which students have to listen to a pilot-controller conversation and answer a multiple-choice test item. However, EPLIS does not cover reading comprehension as aeronautical communications take place via radiotelephony. The fact that achievement tests assess students’ ability to read aviation-related texts, plus the higher relevance of its results to teachers and students, may explain the reason why reading activities are further explored in class, whereas speaking activities are less frequent.

When asked how oral production was assessed, teacher and pedagogical coordinator Pilar explained that they had developed a speaking test, but students’ performance on them was not scored and because of that it had low stakes for students compared to the pencil and paper tests.

The observations also showed that in higher proficient groups, teachers tended to mention the test more often, to develop oral production skills further, to promote more opportunities for interaction among learners and to explore activities similar to the ones found in the test. This scenario confirms that the test washback is more intense to those who see a real possibility of success.

In João’s opinion, students’ attitudes towards the test differed depending on their proficiency level. For him, students who were more proficient showed greater confidence, whereas the less proficient ones were demotivated and worried.

The less proficient ones, those who / those students who know about Eplis (+) who are aware they are going to do the test (+) of the test importance (+) the less proficient ones are demotivated (+) and worried (+) and the more proficient ones are SO confident that they want level 10 (the highest level) (Interview with João, a teacher at the ATC course since 2016).

João explained that less proficient students, even though they were aware of EPLIS importance, had to focus their energy, firstly, on the achievement tests so that they could pass the course. On the other hand, more proficient students did not need to expend too much effort in order to obtain a high score in the course tests. Souza (2019) investigated the relationship between proficiency and washback and found out that more proficient students indeed felt more motivated, had more positive views of the test and engaged more frequently in test preparation activities. Therefore, it can be inferred that due to

the fact that EPLIS results are not demanded, the test washback is more intense in higher proficient groups and reduced in lower proficient ones.

Implications for reviewing ICAO language policy in Brazil

The results of this study show that ICAO language policy in Brazil has had difficulties in promoting the intended effects only through EPLIS. Using Green's washback model, limitations in the communicative functions assessed in the exam had a negative impact in the classroom, therefore tasks that focused on the management of pilot-controller dialogue and on the use of negotiation strategies tended to be overlooked. Besides, the overemphasis placed on certain parts of the test at the expense of others, was considered a negative effect as it narrows down the teaching methods used in the classroom. Therefore, a review in the test is urgently needed so that the overlap area between aeronautical English and test characteristics can be broadened.

Fulcher (2012) reminds us of the power rested in tests, especially, in the external ones, to manipulate teachers' behaviors and hold them accountable for the achievement of goals set by language policies. For him,

policymakers frequently believe that changes can be implemented relatively quickly and cheaply without having to undertake curriculum development or change classroom practices through teacher education programs (FULCHER, 2012, p. 114).

However, there are many situations associated with teachers that can prevent the expected effects from occurring, such as lack of knowledge of the test, lack of knowledge of the test content, lack of the necessary pedagogical skills, as well as issues related to resistance and willingness to innovate (ALDERSON; HAMP-LYONS, 1996; BURROWS, 2004; GREEN, 2007).

Thus, it is clear in the study that more opportunities for teacher training courses on the target language use as well as on the test demands should be promoted. Given that the course achievement tests are perceived as higher stakes than EPLIS, they need to be thoroughly reviewed so that they can better represent the professional needs of learners. To do so, it is necessary to increase teachers' level of assessment literacy as they are the ones responsible for developing these tests.

At last, Brazilian ATC authorities should analyze the possibility of considering students' results in EPLIS when distributing work locations among them. The fact that the test results are accepted, but not demanded has lowered its stakes in students' perceptions. So, considering students' performance in EPLIS fully or partially will probably increase the test stakes in the ATC initial training program and consequently the intensity of its washback.

References

- ALDERSON, J. C.; D. WALL. 1993. Does washback exist? *Applied Linguistics*, 14.2: 115-129.
- ALDERSON, J. C.; L. HAMP-LYONS. 1996. TOEFL preparation courses: a study of washback. *Language Testing*, 13.3: 281-297.
- ARAGÃO, B. F. 2018. O uso de critérios autóctones no contexto aeronáutico: Contribuições para uma nova escala de proficiência para controladores de tráfego aéreo. In: SCARAMUCCI, M. V. R.; P. TOSQUI-LUCKS; S. M. DAMIÃO (Ed.). *Pesquisas sobre inglês aeronáutico no Brasil*. Campinas: Pontes. p. 243-269
- BURROWS, C. 2004. Washback in Classroom-Based Assessment: A study of the Washback Effect in Australian Adult Migrant English Program. In: CHENG, L.; Y. WATANABE; A. CURTIS (Eds.). *Washback in Language Testing: Research, Contexts and Methods*. New Jersey: Lawrence Erlbaum Associates. p. 113-128.
- CHENG, L. 1999. Changing assessment: washback on teacher's perspectives and actions. *Teaching and Teacher Education*, 15.3: 253-271.
- _____. 2005. *Changing language teaching through language testing: a washback study*. Cambridge: Cambridge.
- FREDERIKSEN, J. R.; A. COLLINS. 1989. A systems approach to educational testing. *Educational Researcher*, 18.9: 27-32.
- FULCHER, G. 2012. Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9.2: 113-132.
- GREEN, A. 2007. IELTS washback in context: Preparation for academic writing in higher education. *Studies in Language Testing* 25. Cambridge, UK: Cambridge University Press and Cambridge ESOL.
- HUGHES, A. 1993. Backwash and TOEFL 2000. Unpublished manuscript. Reading, UK: University of Reading.
- MARCUSCHI, L. A. 2008. *Análise da conversação*. 6. ed. São Paulo: Ática.
- INTERNATIONAL CIVIL AVIATION ORGANIZATION/ICAO. 2010. *Manual on the implementation of ICAO language proficiency requirements: Doc. 9835 AN/453*. 2. ed. Montreal.
- MESSICK, S. 1996. Validity and washback in language testing. *Language Testing*, 13.3: 241-255.
- QI, L. Has a High-stakes test produced the intended changes? In: CHENG, L.; Y. WATANABE; A. CURTIS (Eds.). *Washback in Language Testing: Research, Contexts and Methods*. New Jersey: Lawrence Erlbaum Associates. p. 171-190.
- RETORTA, M. S. 2007. Efeito retroativo do vestibular da Universidade Federal do Paraná no ensino de língua inglesa em nível médio no Paraná: uma investigação em escolas públicas, particulares e cursos pré-vestibulares. Doctoral Thesis, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas. Available at: <<http://repositorio.unicamp.br/handle/REPOSIP/269563>>. Access: 25 Jun. 2017.
- SCARAMUCCI, M. V. R. 1999. Vestibular e ensino de língua estrangeira (Inglês) na escola pública. *Trabalhos de Linguística Aplicada*, 34: 7-20.
- SCARAMUCCI, M. V. R.; S. R.B. GATTOLIN; C. H. ROCHA; E. T. N. SATO; N. C. AZEVEDO. 2010. 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.
- SOUZA, P. R. 2017. O efeito retroativo do EPLIS nas percepções de professores. *Brazilian English Language Teaching Journal*, 8.1: 5-29. Available at: <<http://revistaseletronicas.pucrs.br/faced/ojs/index.php/belt/article/view/26863/0>>. Access: 25 Jun. 2017.

_____, P. R. 2019. O nível de proficiência como fator mediador no processo de efeito retroativo do EPLIS. *Revista Estudos Linguísticos*, 48.1: 540-560. Available at: <<https://revistas.gel.org.br/estudos-linguisticos/article/view/2151>>. Access: 10 Apr. 2019.

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

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

TOSQUI-LUCKS, P. T.; P. R. SOUZA; N. A. RAYMUNDO; N. C. GUERREIRO; B. F. ARAGÃO.

2016. Ensino e avaliação de Língua Inglesa para controladores de tráfego aéreo como requisito de segurança em voo. *Revista Conexão SIPAER*, 7.1: 44-54. Available at:

<<http://conexaosipaer.cenipa.gov.br/index.php/sipaer/issue/view/19>>. Access: 25 Jun. 2017.

WATANABE, Y. 2004. Teacher factor mediating washback. In: CHENG, L.; Y. WATANABE; A.

CURTIS (Eds.). *Washback in Language Testing: Research, Contexts and Methods*. New Jersey:

Lawrence Erlbaum Associates. p. 129-146.

Paula Ribeiro e Souza is currently the supervisor of the Aeronautical English Assessment Sector of the Airspace Control Institute, being responsible for the development, maintenance and delivery of the Aeronautical English Language Proficiency Exam (EPLIS). She holds a PhD in Applied Linguistics from the Campinas State University. Her doctoral thesis on the washback of EPLIS has received a commendation by the Brazilian National Council for Scientific and Technological Development.