

Teaching English Language Journal

ISSN: 2538-5488 – E-ISSN: 2538-547X – <http://teljournal.org>

© 2022 – Published by Teaching English Language and Literature Society of Iran

TELL



Please cite this paper as follows:

Ahmadi, A., & Montasseri, Z. (2022). Washback effects of an interactional competence checklist. *Teaching English Language*, 16(2), 285-314.

<https://doi.org/10.22132/TEL.2021.140223>

Research Paper

Washback Effects of an Interactional Competence Checklist

Zahra Montasseri

PhD in applied Linguistics, Department of Foreign Languages and Linguistics, Shiraz University

Alireza Ahmadi¹

Professor, Department of Foreign Languages and Linguistics, Shiraz University

Abstract

The concept of interactional competence and its assessment as a complex language ability has long been a concern for research. A more recent approach to testing this construct has called for an investigation into the commonly-used concept of washback which refers to the effect of testing on teaching and learning. This study aimed to investigate the washback effect of an interactional competence checklist on EFL teaching and learning. The study combined qualitative and quantitative research methods within a comparative design between an experimental and a control group of 27 advanced English learners. Data was collected through students' oral performance, teacher and student surveys, the IC checklist, and classroom observations. The data were analyzed using mixed ANOVA and content analysis using coding schemes and themes to compare the scores obtained before and after the treatment. The study revealed a statistically significant difference in learners' interactive oral performance before and after the implementation of the checklist. Results indicated positive washback in aspects including familiarity with the checklist, test quality, interpretation and use, motivation, professional development, and learners' studying habits.

¹ Corresponding author: arahmadi@shirazu.ac.ir

The implications for the classroom are that language program administrators and curriculum designers can introduce modifications in developing materials that foreground a focus on interactional competence.

Keywords: Washback Effects, Low-Stakes Tests, Interactional Competence, Classroom-Based Assessment

Received: December 5, 2021

Accepted: July 27, 2022



1. Introduction

The research literature abounds with the common argument that testing has an influence on teaching and learning, or what is known to be washback (Alderson & Wall, 1993). The majority of washback studies concentrate on the beneficial or harmful effects of tests on curriculum, methodology, and materials to assist learners to get prepared for an exam. In case the test is high-stakes, the results are then used for more significant educational decisions, namely graduation, admissions, promotions, and employment; thus, the results affect people's future (Muñoz & Álvarez, 2010). A test is believed to have a positive washback when a test encourages good teaching and learning practice (Taylor, 2005). On the other hand, negative washback occurs when the definition of linguistic competence becomes restricted since it may encourage a reductionist view of language proficiency, and hence teaching and learning contents may become narrow in scope (Bokiev & Abd Samad, 2021).

The number of empirical studies focused on washback effects has grown over the last few decades ever since Alderson and Wall (1993) put forward the 15 Washback hypotheses; generally, they have accentuated the complexity of this phenomenon along with important factors involved in testing that influence teaching and learning. This highlights the necessity of

careful investigations of how certain testing procedures operate within certain contexts (Tsagari & Cheng, 2017). Such being the case, a review of pertinent literature has revealed that unlike the majority of studies on large-scale standardized assessments, few studies have been conducted on the washback effects of classroom-based and teacher-led assessment on instructional and learning practices (Bokiev & Abd Samad, 2021). Classroom assessment can be distinguished from standardized tests in that the latter incorporates a larger number of test users and can be used in a larger scope; hence the degree of academic and professional formality and purposes for which they are used is more comprehensive. In contrast, low-stake tests have less power to modify teacher and learner behavior and, therefore, are not at the core of decision-making processes and have fewer consequences (Cheng, Sun, & Ma, 2015). Altogether, it is argued that any type of assessment, be it high-stake or low-stake, can introduce and be a source of educational change (Muñoz & Álvarez, 2010).

Moreover, even though some models of washback take into consideration multiple stakeholders of washback mechanisms (e.g., Alderson & Wall, 1993), the majority of previous studies focused on the effects of assessment on instruction; hence, other test users, particularly learners, their learning processes, outcomes, and perceptions have been relatively neglected (Tsagari & Cheng, 2017). On the other hand, despite devoted attention to teaching and testing speaking ability, little research to date has been conducted solely on oral assessment systems (e.g., Hirai & Koizumi, 2009; Muñoz & Álvarez, 2010); that is to say, washback studies mostly approach oral assessment within an integrated assessment system (Umashankar, 2017).

One facet of oral assessment is the spoken interaction in pairs or groups of interlocutors referred to as interactional competence (henceforth IC). IC refers to the ability to make use of interactional resources, such as speech

acts, turn-taking, and repair to articulate communicative intentions through available linguistic resources in real-life contexts (Young, 2011). The notion of IC originated from the focus on interaction which challenged the concepts of communicative competence and communicative language ability (Bachman & Palmer 1996). It was argued that such conceptualization needs to be extended by considering language competence from an individualistic to a more social view in which communication is co-constructed in a joint interactional context (Ducasse & Brown, 2009). Kramsch (1986) put forward convincingly that participants build upon effective communication mutually and reciprocally and the responsibility of interaction is not assigned to a single individual. A popular means of assessing performance and by extension IC is using scales, checklists, and rubrics (Lynch, 2003). May et al. (2020) developed a detailed checklist for IC assessment for teachers to provide learners with feedback. Such being the case, previous studies have failed to address how the implementation of this checklist impacts teachers and learners.

This study set out to help fill these gaps in the literature by investigating the washback effects of the mentioned recent checklist to assess learners' IC (May et al., 2020) in a classroom-based assessment context from teachers and learners' perspectives. More specifically, this study sought to explore how the use of an IC checklist influences teacher and learner practices.

2. Literature Review

2.1 Washback studies

Some scholars highlighted the complexity of washback in that the process of washback is generated under certain mediating factors (Shohamy et al, 1996; Wall & Alderson, 1993), such as test factors and participants' factors. Having reviewed the interplay between various elements, Umashankra (2017) proposed two major factors namely micro-context and macro-context factors.

Micro-context factors include teachers, students, and assessment-related factors. Macro-context factors, on the other hand, refer to factors that influence washback within a larger scope, such as the educational system and the society where the test is administered (Watanabe, 2003).

The majority of washback-intended investigations have focused on the positive and negative effects of high-stakes tests on teachers' and students' attitudes, course materials and content, and methodology (Kutlu & Demiroglari, 2020). The findings indicated that these studies have had a greater impact on course content rather than on teachers' methodology. Cheng (1997), for example, studied changes in teaching and learning introduced by the Hong Kong Certificate of Education Examination (HKCEE) in English and found that washback resulted in changes in teaching materials and not in the teaching approaches. Similarly, Qi (2004) and Burrows (2004) conducted studies on washback effects on local educational systems and conclusively claimed that the exams had negligible or no influence on teaching methods.

The studies conducted on the washback effects of an instrument are of two categories: selection tests and their rubrics and proficiency tests and the scales attached to them. Selection tests include NMET (National Matriculation English Test), O-level English Exam, and HKCEE (Hong Kong Certificate of Education Examination in English). For instance, Qi (2004) investigated the washback effects of NMET and concluded that high expectations from the learners and the design of the test failed to render positive effects on teaching and learning. In another study, Cheng (2005) examined the washback effect of HKCEE on in a three-level investigation and found that although the teaching content changed positively, teacher's attitudes remained unchanged towards the test. Similarly, Wall (2005) explored the washback effects of the O-level English Exam and concluded

that the instrument changed *what* teachers teach rather than *how* teachers teach.

In terms of proficiency tests, in one validation study, Hawkey (2006) inquired into the IELTS scoring rubric and reported the exam as a reasonable, authentic, and fair tool that boosts motivation and causes moderate anxiety. The author carried out another investigation on PL2000 (Progetto Lingue 2000) along with Cambridge ESOL (English for Speakers of Other Languages) and surmised that PL2000 positively affected motivation, teaching, and school management. Moreover, Gu (2007) discussed the effects of CET (Cambridge Education & Training) on college English learning and teaching where the findings were mainly positive; the majority of stakeholders reported desirable washback effects of the test and attributed minor negative effects to misuse of the instrument.

Moreover, as Bailey (1999) put it, studies on the effects of assessment on students' learning seem to lack a clear understanding of the washback process. Although both negative and positive test effects have been conducted before, few studies have reported verifiable gains in students' learning. Some studies have investigated the effects of coaching for assessment on the gain score (Robb & Ercanbrack, 1999); yet, it is uncertain whether or not the gains in these studies indicate test-taking skills or language ability tests.

Seen in its entirety, bringing significant changes by merely introducing the modifications is unlikely to happen (Larsson & Olin-Scheller, 2020). To make a meaningful institutional change, various elements beyond the test itself need to be taken into account. As Wall (1996) asserted following up a study on the Sri Lankan educational system, different factors play a role in preventing the assessment from providing positive washback namely the exam content, resistance to change, teachers and learners' lack of

understanding of the exam, lack of well-trained teachers, and the gap between teachers and test designers.

To overcome obstacles of the positive washback, some scholars suggest improving the assessment system by ensuring the task authenticity (Dong, 2020), teachers' and learners' understanding of the test (Shirzadi & Amerian, 2020; Zhan & Andrews, 2014), detailed score reporting (Liu & Yu, 2021), conformity between curriculum and exams (Sato, 2019), and self-assessment (Li, 2018). Other authors have also referred to providing meaningful feedback (Azadi & Gholami, 2013), and a variety of test tasks and formats (Kilickaya, 2016). Therefore, there is a need to study the interrelationship between assessment and language competence in the classroom and the effects of classroom low-stakes assessment on the degree and depth of learning (Green, 2020).

2.2 Interactional competence

Since the emergence of IC, a solid body of academic research regarding the theoretical debates toward a comprehensive definition of the construct has been carried out. Several studies have explored specific practical examples of IC in candidates' performance (e.g., Galaczi, 2014) while others have discussed the broader direction of the IC construct (e.g., Galaczi & Taylor, 2018; May, 2011; Nakatsuhara, 2013). While these investigations have contributed to a deeper understanding of the definition and conceptualization of IC, a need was felt for its operationalization and was partly administered, such as the Kanda English Proficiency Test (KEPT), Cambridge English General English tests, Trinity's Integrated Skills of English (ISE) Speaking and Listening test, and the Test of English for Academic Purposes (TEAP). However, the aim of the aforementioned tests is not to assess IC directly, and they address the issue only to an insufficient extent. Given their inadequacy to fully assess the IC and the challenge of how to assign separate scores to

test-takers (Chalhoub-Deville & Deville, 2005), an attempt was made to create an assessment tool solely dedicated to the operationalization of IC.

Having reviewed a wider literature on assessing IC, Nakatsuhara et al. (2018) asserted that "the findings have not been fully realized in terms of informing the teaching and learning of interactional skills in a comprehensive and user-friendly way" (p. 8). Thus, they developed a checklist for assessing IC for learning-oriented assessment. Accordingly, this study intended to respond to the following research questions:

1. What is the intended washback of the IC checklist on teaching and learning English?
2. How do the teachers perceive the IC checklist and its washback on their English language teaching?
3. How do the students perceive the IC checklist and its washback on their English language learning?
4. Do learners who participate in the experimental group make more progress than the comparison group in terms of their IC?

3. Method

The study benefited from both qualitative and quantitative research methods, comprising a comparative investigation between a control and an experimental group. This mixed-methods design contributed to not only gathering numerical data but also seeking to explore and understand the washback effect from the test users' point of view. It paves a more flexible way that allows more spontaneity, greater in-depth analysis, and interaction between the researchers and the participants. Moreover, the sequential hybrid design was embedded, in that the qualitative data supported the quantitative phase, which adds to the credibility of the results. The rationale for embedded design is that more data is required to answer a single research question as a supplemental data source (Ary et al., 2018).

3.1 Sampling Method

Since there are different groups of participants in the study, the sampling method for each group was different. To select English language learners, the authors took an available sampling approach; however, the learners were randomly assigned into two experimental and control groups. English language teachers were chosen purposefully based on their experience and expertise in teaching advanced English courses. The raters were also chosen through purposive sampling because of their experience in rating and familiarity with oral performance assessment.

3.2 Participants

In line with the research objectives, three groups of participants were involved in this study: (1) English language instructors, (2) English language adult learners, and (3) professional raters. The sampling of the study was sequential, i.e., one sample was selected based on findings from the earlier phase. In this case, for the qualitative phase, samples were chosen based on the quantitative stage of the study. Despite the management awareness of the research purpose and the participation of the staff and students, the participants' anonymity was promised and maintained to obtain their confidence and larger degrees of openness to answer the survey questions (explained below). Furthermore, it was made clear that participating in the study was voluntary and that each participant could withdraw from the study at any time. All research participants confirmed their informed consent to participate in the study, including the use of video footage of their responses.

3.2.1 English language instructors

Two English language instructors who were teaching the two control and experimental classes at the language center participated in the study. The criteria for choosing these instructors were their teaching experience, academic qualifications, and willingness to participate. Hence, the two

teachers had more than 10 years of teaching experience and both were teaching adult advanced English learners. They both had formal academic qualifications from local universities in Iran, holding Master's degrees in English language teaching. The purpose of including language instructors was twofold: first, to ascertain their understanding of how the IC checklist must be used in class to boost learners' interactional strategies, and second, to investigate their viewpoints regarding the washback effects on their teaching. The experimental group instructor received training on how to use the IC checklist (explained below) in teaching and how to implement it in class for assessment. It is worth mentioning that both teachers followed the same teaching approach, i.e., communicative-based language teaching (ongoing institutional observations approved the similarity of approaches), and the only difference in the two classes was the introduction and practice of the IC checklist in the experimental group as explained below. The same amount of time (16 sessions) in the control group was allocated to general speaking ability including the skills required for effective communication and the ability to convey information verbally in the target language.

3.2.2 English language learners

Another group of participants consisted of 27 adult learners both male and female who were studying English as a foreign language at [*name of the university – removed for anonymity reasons*]. Their level of proficiency was advanced based on their previously taken courses and achievement records, and they were all in the same course level and they ranged in age from 18 to 32. Learners were taking English as an extracurricular activity and had different academic fields of study such as physics, law, biology, engineering, medicine, dentistry, and sociology. The participants had already passed 11 English courses and were assessed through the same testing system by the institute. They were randomly assigned to control and experimental groups.

Apart from the instructors and teaching or testing methods, all other conditions namely the textbook, the timing of the class, the midterm, and final examinations were all the same. The reason why these learners were chosen was due to the nature of the IC checklist, which requires intermediate to advanced language users who can take on interactional strategies. The role of language learners was fundamental in the study because their overall perception of the IC checklist was the basis for analysis.

3.2.3 IC raters

The third group of participants was two professional raters with fair knowledge and experience in rating oral performance through purposive sampling, mainly because the raters needed to be familiar with the rating procedure and rating instruments (rubrics, checklists, etc.) and above all, a knowledge of interactional competence. Since the present study is part of a larger project, these raters were already trained to use the IC checklist and the feedback attached to it and they had rated advanced learners based on the checklist. Briefly put, during the three two-hour online training sessions, the following phases took place: a general overview of the IC, familiarization with the IC checklist, introduction of interactive task types, rating practice, and discussion and negotiation. The raters were asked to assign scores to the learners' performances once before the study and once again at the end of the course for the purpose of comparison.

3.3 Data-gathering tools

Apart from the textbook used throughout the semester as the main material, some instruments were used to collect data from the participants as explained below.

3.3.1 The IC checklist

One of the instruments, which is of main concern in the present study, is the IC checklist developed by May et al. (2020). The checklist is designed in two

full and concise versions for providing the test takers with suitable feedback. To fulfill the objectives of this study, the latter is utilized, which consists of four main criteria, namely initiating new ideas, keeping the discussion, negotiating towards an outcome, and using body language appropriately. For each criterion, there are some interaction strategies that interlocutors are expected to follow to gain feedback from the instructor. Also, as the number of items on the checklist is 13, each speaker will gain a score ranging from 0 to 13.

The first criterion, initiating new ideas, deals with the speaker taking the initiative to offer new ideas, judging the appropriate time to articulate opinions after the previous idea is sufficiently debated, and the ability to use proper choices of language to initiate his/her idea. The next criterion, keeping the discussion going over several turns, is concerned with six strategies: 1. developing one's idea (how each speaker extends his/her idea further); 2. developing a partner's idea (how each speaker extends the proposed idea of his/her partner by adding adequate ideas instead of just a single response); 3. inviting (how the speaker gets actively engaged in the conversation by encouraging the partner to maintain the interaction by asking questions, completing a sentence, etc.); 4. listening (showing involvement while listening using short responses or backchanneling); 5. being collaborative (attempts to keep the interaction natural by avoiding long pauses or interruptions); and 6. language (using convenient language through showing agreement, disagreement, explanation, justification, examples, etc.).

The third criterion, negotiating towards an outcome, pertains to two strategies, the first of which is a joint decision. This strategy, as Nakatsuhara et al. (2018, p. 67) put it, is defined as the extent to which the interlocutors "proactively work towards making a joint decision (e.g., inviting the partner to make a choice, showing a willingness to compromise)". The second

interactive policy is the language which refers to using the appropriate language including comparison, evaluation, summary, and prioritization of points discussed. The last criterion, using body language appropriately, accounts for non-verbal aspects of language significant in interaction, that is to say, body language (smiling, nodding to show interest) and eye contact.

For each strategy, the rater is provided with two levels of feedback: *Well done* and *needs more work* as well as a column to add comments. It is noteworthy that a detailed description of the level of performance at any level for all strategies is given so that learners gain self-awareness of their weaknesses and strengths in this regard. The results obtained from this checklist were used for analysis. Also, as the number of items on the checklist is 13, each speaker gained a score ranging from 0 to 13.

3.3.2 A students' survey

To collect data from students' perceptions of the IC checklist, a survey consisting of 10 closed and open-ended questions was adapted from Loumbourdi (2016, pp. 171-172). It consisted of the objective of the survey, a description of how to fulfill it, the consent form, nine open-ended questions, and a Likert-type question with seven sub-topics. The survey asked students questions about the course objectives and instructional practices, the assessment formats, their familiarity with the checklist criteria, the feedback received on their performances, and the probable effects it had on their learning. The survey was analyzed by an expert in the field before implementation. The open-ended questions were analyzed through coding to come up with themes.

3.3.3 A teacher's survey

This instrument was adapted from Loumbourdi (2016, pp. 163-164) and aimed at collecting data from the teacher regarding her teaching and assessment practices, the test format, the IC checklist criteria, the feedback

procedure, and the possible effects on teaching. The survey included 19 open-ended questions and closed items as well as two recommendation questions. The survey was analyzed by an expert in the field before implementation. The open-ended questions were analyzed through coding to come up with themes.

3.3.4 Class observations

To check the extent to which the IC checklist was used in the classroom during the course, eight observations out of 16 sessions were conducted by one of the authors in the experimental group. The observations were conducted only during sessions and when the IC checklist was practiced. The purposes of class observations were to witness how the instructor introduced the checklist, explained the features of IC and the criteria related to it, its significance in interactions, kept track of its implementation during the course, and monitored the learners' practice. The observer filled out observation sheets that required the following information: class objectives, instructional tasks, practices, and feedback procedures given by the teacher to learners.

3.3.5 Oral interactive task

To obtain data from the students, an interactive task was given to them. The task consisted of a controversial topic to engage students' interaction while they were asked to express and support their opinions in English on the issue in the form of paired conversations. During the task, learners refuted each other's ideas and challenged one another. The reason for choosing a disputable topic was to increase learners' engagement in the interaction and keep the discussion going for a longer time. Each paired performance took between 10-15 minutes and was videotaped with the learners' permission. The same task was given once to the learners before the study and once again at the end of the course for comparison.

3.4 Data Collection and Analysis Procedure

At the beginning of the course, an oral task was given to all the participants in both classes and the teaching approach in both classes was communicative-based and the focus of the coursebook, Summit 2A (Ascher & Saslow, 2016), was on oral skills. The role of the checklist was to introduce the features of the IC, its significance in language communications, and interactional skills. To do this, the instructor allocated 8 sessions to the introduction of the IC features, explaining the IC checklist and its sub-skills, giving feedback to learners accordingly, and raising awareness towards the interactional strategies. The instructor kept the students engaged in the IC checklist by asking them to comment on their classmates' peer performances. The sessions in which the IC checklist was practiced were observed to make sure the treatment was fully completed. The time spent on speaking interactions was the same based on the institute schedule; however, in the experimental group, the speaking time was divided into two parts, practicing general speaking ability by doing the oral tasks of the book (making conversations, sharing opinions, giving summaries) and learners' reflection time on interactional features, while in the control group only speaking skills were practiced.

At the end of the course, once again the same oral task was given to all the learners to make a comparison between the control and experimental group performances in terms of IC development. Meanwhile, the experimental group's teacher and learners answered survey questions about the possible effects of the IC checklist on their teaching and learning.

To analyze the quantitative data, SPSS 23 was used to run a series of mixed ANOVA on the learners' test score data to make a comparison between experimental and control groups' performances before and after the intervention. Moreover, survey answers were analyzed after developing a coding scheme based on the preliminary readings to identify salient themes and patterns. To examine the agreement between the raters, the inter-rater reliability between the two raters was conducted using the Kappa statistic

through SPSS 23 and the results indicated 81% consistency (Kappa = 0.81, $p < 0.001$) which suggests an acceptable reliability index.

4. Results and Discussion

4.1 The quantitative findings

In the quantitative part of the study, a mixed, within-between-subjects design was conducted to assess the washback effect of the implementation of the IC checklist on participants' scores across two time periods (pretest, post-test).

Table 1 depicts the descriptive statistics of the results.

Table 1

Descriptive Statistics of IC checklist

	Group	Mean	SD	N
Pretest	Experimental	6.60	1.50	15
	Control	6.41	1.44	12
	Total	6.51	1.45	27
Posttest	Experimental	10.26	1.22	15
	Control	6.83	1.26	12
	Total	8.74	2.12	27

As shown in Table 1, the two groups seem to be similar at the beginning but different at the end of the study. The following table shows the results of within-subjects effects.

Table 2

Tests of Within-Subjects Effects

	Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.606	38.46	1.000	25.000	.000	.606
	Wilks' Lambda	.394	38.46	1.000	25.000	.000	.606
	Hotelling's Trace	1.539	38.46	1.000	25.000	.000	.606
	Roy's Largest Root	1.539	38.46	1.000	25.000	.000	.606
Time * Group	Pillai's Trace	.494	24.36	1.000	25.000	.000	.494
	Wilks' Lambda	.506	24.36	1.000	25.000	.000	.494
	Hotelling's Trace	.975	24.36	1.000	25.000	.000	.494
	Roy's Largest Root	.975	24.36	1.000	25.000	.000	.494

The results of the analysis indicated a significant interaction group-by-time interaction effect.

The findings suggest that the effect size for both time and interaction between group and time are very large. Considering the guidelines proposed by Cohen (1988, pp. 284–7): .01=small effect, .06=moderate effect, .14=large effect.

Table 3., tests of between-subject effects, presents the ANOVA results for the between-groups variable as indicated in Table 3.

Table 3
Tests of Between-Subjects Effects

Measure: IC						
Transformed Variable: Average						
Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	3023.379	1	3023.379	1326.237	.000	.981
Group	43.601	1	43.601	19.126	.000	.433
Error	56.992	25	2.280			

As seen in the above table, the significance value for the group is .000 ($F=43.6$, $p < .05$, partial eta square=.43), suggesting a significant difference in the effectiveness of the IC checklist. That is to say, the implementation of the IC checklist in the experimental group made a difference in their performances. The result of this performance is depicted in the interaction plot of IC in Figure 1.

Washback Effect ...

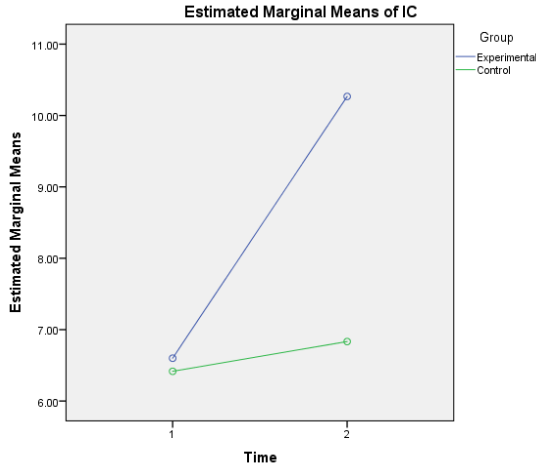


Fig. 1. Interaction plot of the IC

Figure 1 indicates the substantial effects of the IC checklist instruction and practice in the experimental group compared to the control group. It also shows the positive effect of time on both groups, especially for the experimental group where the participants significantly outperformed the control group in posttests. To locate the differences in group performances and make a comparison, a set of paired samples t-tests was administered and the results are given.

Table 4
Paired Samples Test of Groups

	Paired Differences			t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error			
Experimental Pre-Post	-3.66	2.19	.56	-6.4	1	.000
Control Pre-Post	-.41	.66	.19	-2.1	1	.054

As seen in the table, unlike the experimental group, there was no significant difference between the means of the pretests and posttests for participants in the control group. Despite this, a minor improvement can be observed in the control group's performance which can be attributed to the general speaking practice during the course. Moreover, to ascertain the

homogeneity of the groups before the study and their difference afterward, a set of independent samples t-tests were run. The findings are presented below.

Table 5
Independent Sample Tests of Groups

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2- tailed)	Mean Difference	Std. Error Difference
Pretest	Equal variances assumed	.057	.813	.321	25	.751	.183	.571
	Equal variances not assumed			.322	24.1	.750	.183	.569
Posttest	Equal variances assumed	.002	.961	7.134	25	.000	3.433	.481
	Equal variances not assumed			7.105	23.3	.000	3.433	.483

The table indicates the similarity of both groups before the study in terms of their IC and statistically significant results between their performances at the end. Therefore, to answer the third research question, this information yields the effectiveness of the IC checklist practice in the experimental class. That is to say, while at the beginning of the study, the groups were similar in terms of their IC, the experimental group statistically improved while the control group slightly increased in performance.

4.2 The Qualitative Findings

In the qualitative part of the study, the research findings from survey questions with teachers and learners aimed to find their respective evaluations of the washback effects of the checklist. Generally, the IC checklist was positively evaluated by the teacher who approved of its design, the sub-categories, the content, and the items. The teacher agreed that the checklist promotes the implementation of interaction strategies, evaluates higher-order thinking skills, aligns with their expectations about effective oral assessment tools, and provides feedback for learners, who can adjust

teaching and learning accordingly. She pointed out the easy-to-follow scoring system and report since it offers separate scores for each section. In other words, it facilitates teaching and learning interaction. The following themes were extracted from the survey questions collected from the teacher and learners.

4.2.1 Familiarity with the checklist

Regarding familiarity with the checklist, the teacher asserted that a higher familiarity rate would lead to better interpretation and use of the scale. The instructor reported the usefulness of pre-course preparation in acquaintance with the assessment requirements and criteria. The teacher suggested that interaction strategies must be implemented in the teaching syllabus, marking criteria, teacher training, and adequate time should be allotted to practice. Familiarity with the checklist aided the instructor in setting course objectives and lesson plans in line with the scale since the checklist incorporates the interactional requirements.

The language learners claimed that at the beginning of the course, they were not much familiar with the marking criteria; however, as time passed by and the teacher spent sessions explaining the checklist, they came to a deeper understanding of the IC features and checklist.

4.2.2 Checklist quality

One major theme extracted from the survey was the quality of the checklist, which was positively evaluated by the teacher. Although the syllabus did not directly address the interaction strategies, the checklist was largely aligned with it because of the communicative nature of the curriculum. Hence, the content and item types contributed to oral practice depth by developing a greater level of meta-cognition in the experimental group regarding their speaking (i.e., self-monitoring and self-regulation). Regarding the score allotment, the instructors believed that it was clear enough to assign a score

and provide learners with feedback; however, the score followed a more holistic approach than an analytic one. Comments by the teacher indicated that the IC checklist helped her monitor the learners' progress and provide appropriate feedback. The form of feedback included offering constructive comments to those who achieved certain IC features and suggestions on how to develop IC skills for those who need more practice. Also, the administration of the checklist was reported to be convenient (the guidelines were easy to follow and understand) and fair.

Test-takers agreed that the introduction of the checklist within the course raised their awareness about the significance of many aspects of a successful interaction otherwise neglected. They claimed to have focused on the interactional strategies even if they were not included in the final assessment of the course. Another positive aspect of the IC checklist was the feedback attached to the checklist which helped them identify their current level of performance and how to progress. Because the feedback contained useful phrases of functional language for interaction to repair the interactional breakdowns, the learners found it informative for their improvement. Similar to the teachers, the learners were better able to identify aspects of their IC skills that they could reflect on and improve. This is in line with Shackleton's (2020) findings in which the standardized tests developed learners' awareness of the construct being tested.

4.2.3 Interpretation and use

The survey data showed the extent to which the IC checklist influenced the interpretation and use of test results. According to the teacher, the implementation of the checklist was beneficial for her teaching quality to reflect on her practice, promote her knowledge of IC, and review her lesson plans. The instructor favored the interactive teaching approach that highlighted students' active engagement, task-based language teaching, and

students' expression of their opinions. Hence, the checklist encouraged communicative language teaching and encouraged student-centered instruction. It also reflected a higher level of language proficiency, contributed to better communicative outcomes, and increased their motivation. The teacher put forward that based on the results gained from this checklist, the students' future performance could be predicted. She also highlighted the changes she had made in the course design, schedule, teaching content, method, ideology, attitude, depth, and speed of instruction to fit the IC checklist in the class, and pointed out the overall positive effects of the IC checklist on teaching and learning.

The learners had mixed opinions about the checklist use and interpretation. Some students approved of some items of the checklist, because of its parameters which include having a short pause, listening well, and continuing the conversation, through which learners' speech and listening ability can be measured. They not only encourage learners to be more active in the class, but characterize a good relationship; i.e., all conversation items are considered to be in everyday use, and these tips can be used for both language learning and normal conversations. According to the learners, many of its items have been forgotten in the scoring of interactions by teachers and it seems that little attention had been paid to them. Moreover, the learners were asked whether they would take the test again as an integral part of the learning process, and the majority agreed to retake it, indicating no discouragement. Many students also found the paired oral test less intimidating than the usual monologic tasks.

4.2.4 Motivation and professional development

The comments made by the teacher suggested that the IC checklist had a positive impact on her motivation because it made her focus on the learning outcomes. She maintained that as a result of employing this scale in the

course, she made several innovations in teaching interactional skills and took pragmatic aspects into account. For example, the accompanying comments in the checklist encouraged the teacher to bring more functional fixed phrases and formulaic language to the class. In addition, she reported the positive impact of the checklist on her professional development by improving their assessment literacy, reflecting on their teaching, and exploring students' potential. According to the instructor, reflecting upon the aspects of the checklist deepened her understanding and expertise of the IC construct, which in turn encouraged students' achievement. The teacher believed that this assessment tool pushed students to take interaction seriously and take responsibility for their learning, (i.e., they became more aware of the parts of successful interaction, and were more capable to develop as speakers).

In a similar vein, most students regarded the IC checklist as beneficial in that it encouraged paired and group interactions, which they perceived as more entertaining and challenging in improving their communicative skills than solo performances. They felt that working in groups enabled them to work on social aspects of interaction by exchanging ideas, listening carefully, and getting passive students more engaged in groups. Moreover, apart from fostering language skills, some learners mentioned the usefulness of collaborative oral tasks in developing team spirit and leadership skills and enhancing mutual learning. The students also recognized the IC checklist as a tool for interaction outside the classroom and relevant to real-life situations, which is necessary for using language in various contexts. The majority of learners (70%) also believed that their confidence increased as a result of realizing their strengths/weaknesses.

4.2.5 Impacts on studying habits

The students reported having spent more hours practicing interactional skills compared to the previous semesters. They stated that what had changed was

not only the amount of time dedicated to studying but also the method of studying they employed under the influence of the IC checklist. Almost all the students stressed that the focus of their study was not on learning grammar and vocabulary, but on other materials rather than the textbook, including phrasal verbs, formulaic language, watching movies, and other authentic materials. Altogether, the learners shifted to a more authentic communication approach which is in line with Ahmadi et al.'s (2021) findings on reciprocity in spoken communication.

4.2.6 Observations

To support the abovementioned results, classroom observations also yielded similar findings regarding the survey questions. During the observations, the classroom activities, atmosphere, content, and teaching method were analyzed. Regarding the teaching content, the analysis indicated that a wide range of activities was directly related to interactional skills within the allotted time. Besides widening students' scope of knowledge, the instructor attempted to raise their awareness of aspects of interaction otherwise neglected. Concerning the teaching method, the teacher fully covered various activities to boost the learners' IC, including teacher-student interaction, peer work, and group discussion. Feedback was frequently given to students on both form and meaning of their utterances which was particularly scrutinized for further improvement in using the suggested phrases. Concerning classroom atmosphere, the frequency of laughter is commonly suggestive of a friendly and warm classroom atmosphere (Gu, 2007), and several cases of laughing were observed in the classroom. Plus, almost no instances of reprimanding, anxiety or tension were found. Since laughing also occurred during the review of the IC checklist, it can be inferred that learners' passion and interest in language learning did not change negatively while examining the checklist.

Overall, to answer the first and second research questions, the implementation of the IC checklist resulted in a strong positive washback on the educational processes, instruction, and learning. There was an increased focus on the notion of IC from both the teacher and learners, which in turn resulted in an increase in time allotment for its development, employment of teaching and learning strategies toward test success, and raising motivation. This is in agreement with Estaji and Alikhani's (2020) findings about teachers and learners' perceptions of a test and its washback effect. Furthermore, the focus on the interactional strategies and the test format was also perceived as a positive washback since test familiarization and learning promotion led to the alleviation of test anxiety. This is in contrast with Hughes's (2021) results where teachers had mixed perceptions in this regard and did not consider the test to bring a mere positive impact.

To assure that negative washback had not occurred, careful investigation of the results suggested neither narrowing down the scope and content nor any increased pressure to cover the material and a high level of anxiety. Negative washback was not observed in materials memorization; rather, interactional skills were acquired and developed. However, students and teacher's consensus on insufficient descriptive levels in the checklist prevents teachers from providing more detailed feedback and does not allow learners to notice their improvements in the IC and, consequently, to obtain higher scores.

If important decisions are to be made based on the results obtained from this checklist, it would be necessary to revise the present reporting system to supply more information (e.g., a rubric) on the following: (1) scores for each sub-category, (2) correction rate for each item to identify student's strength and weaknesses, (3) student ranking in total scores and sub-category scores,

and (4) average performance of each language center (school, university, institution) in a certain section.

Any positive consequences of introducing a new testing system are linked to consequential validity. In this study, consequential validity is observed in that a positive washback effect occurred. The findings indicated that the IC checklist can promote learning and is beneficial to the educational system. The employment of the checklist had an impact on the teacher's methodology, the content of teaching, students' perceptions of learning and the exams, their motivation, and their self-confidence.

5. Conclusion

This study showed clear evidence of positive washback effects of the IC checklist in an advanced English course. It can be concluded that washback can be fostered by training the teachers, specifying the objectives, informing test-takers of the scoring scale and assessment procedures, and structuring assessment tasks. However, to aid teachers to understand and appropriately use the scale, it is necessary to provide constant help and systematic support over time. If learners are well-informed about assessment practices, they concentrate on specific goals, increase their language scores, and perform better. Moreover, positive washback can be further encouraged if both teachers and learners set up a connection between educational objectives and assessment. Finally, learners who are directly influenced by washback must be given chances to offer constructive feedback on both teaching and assessment to contribute to scale improvement.

An implication of this study for test designers is developing authentic tests requires an alignment between test objectives and teachers and learners' needs and expectations. Moreover, exam preparation and course instruction can be shaped by the teachers' understanding of the exam specifications; hence, teachers can make a proper balance between learners' needs, goals,

and wants and the test requirements. Another implication of the study is for syllabus designers and materials developers to consider teachers and learners' perceptions and implement them in materials and textbooks to be accurate reflections of test specifications.

Finally, it is noteworthy that no washback study is complete without investigating the macro levels of test influence on the educational system and policies, socio-political aspects, and the larger population outside the classroom. The study investigated the washback effects only partly (i.e., the washback effects of the IC checklist at the classroom scope on the immediate stakeholders directly linked with it). To achieve a more comprehensive impact, further research is required on socio-political levels regardless of much more time and space.

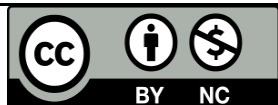
References

- Ahmadi, H., Behnam, B., & Seifoori, Z. (2021). The reciprocal questioning as a formative assessment strategy: EFL learners' reading comprehension and vocabulary learning. *Teaching English Language, 15*(2), 61–93.
- Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115–129.
- Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education*. U.S.: Cengage Learning.
- Azadi, G., & Gholami, R. (2013). Feedback on washback of EFL tests on ELT in L2 classroom. *Theory & Practice in Language Studies, 3*(8), 35–41.
- Bachman, L. & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey K (1999). Washback in language testing. *TOEFL monograph series*. Princeton, NJ: Educational Testing Service.
- Bokiev, U., & Abd Samad, A. (2021). Washback of an English language assessment system in a Malaysian University foundation program. *The Qualitative Report, 26*(2), 555–587.
- Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing* (pp. 113–128). Mahwah, NJ: Lawrence Erlbaum.

- Chalhoub-Deville, M., & Deville, C. (2005). A look back at and forward to what language testers measure. In *Handbook of research in second language teaching and learning* (pp. 839–856). Routledge.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38–54.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge University Press.
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48(4), 436–470.
- Dong, M. (2020). Structural relationship between learners' perceptions of a test, learning practices, and learning outcomes: A study on the washback mechanism of a high-stakes test. *Studies in Educational Evaluation*, 64, 100824.
- Ducasse, A. & Brown, A. (2009). Assessing paired orals: Rater's orientation to interaction. *Language Testing*, 26, 423–443.
- Estaji, M., & Alikhani, H. (2020). The first certificate in English textbook washback Effect: A comparison of teachers' and learners' perceptions. *Teaching English Language*, 14(1), 119-152.
- Galaczi, E.D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired tests? *Applied Linguistics*, 35(5), 553–574.
- Galaczi, E. D., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, 15(3), 219–236.
- Green, A. (2020). Washback in language assessment. *International Journal of English Studies*, 13(2), 39–51.
- Gu, X. (2007). *Positive or negative—an empirical study of CET washback*. Chongqing: Chongqing University Press.
- Hawkey, R. (2006). *Impact theory and practice*. Cambridge University Press.
- Hirai, A. & Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story-retelling technique. *Language Assessment Quarterly*, 6, 151–167.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hughes, L. (2021). Washback and the assessment practices of ESL instructors at Japanese university. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, 5(1), 1–9.
- Kilickaya, F. (2016). Washback effects of a high-stakes exam on lower secondary school English teachers' practices in the classroom. *Lublin Studies in Modern Languages and Literature*, 40(1), 116–134.

- Kramsch, C. (1986). From language proficiency to interactional competence, *The Modern Language Journal*, 70(4), 366–372.
- Kutlu, M. O., & Demiroglari, G. (2020). Washback effect of higher education institution exam--foreign language test on university candidates' attitudes. *African Educational Research Journal*, 8, 121–130.
- Larsson, M. & Olin-Scheller, C. (2020). Adaptation and resistance: Washback effects of the national test on upper secondary Swedish teaching. *The Curriculum Journal*, 31(4), 687–703.
- Li, X. (2018). Self-assessment as 'assessment as learning' in translator and interpreter education: validity and washback. *The Interpreter and Translator Trainer*, 12(1), 48–67.
- Liu, X., & Yu, J. (2021). Relationships between learning motivations and practices as influenced by a high-stakes language test: The mechanism of washback on learning. *Studies in Educational Evaluation*, 68, 96–107.
- Lynch, B. K. (2003). *Language assessment and programme evaluation*. Minnesota: Capstone.
- May, L. (2011). *Interaction in a paired speaking test*. Frankfurt am Main: Peter Lang.
- May, L., Nakatsuhara, F., Lam, D., & Galaczi, E. (2020). Developing tools for learning-oriented assessment of interactional competence: Bridging theory and practice. *Language Testing*, 37(2), 165–188.
- Muñoz, A. P. & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33–49.
- Nakatsuhara, F. (2013). *The co-construction of conversation in group oral tests*. Frankfurt am Main: Peter Lang.
- Nakatsuhara, F., May, L., Lam, D., & Galaczi, E. (2018). *Learning-oriented feedback and interactional competence*. United Kingdom: Cambridge Assessment English.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng & Y. Watanabe (Eds), *Washback in language testing* (pp. 171–190). Mahwah, NJ: Lawrence Erlbaum.
- Robb, T. & Ercanbrack, J. (1999). A study of the effect of direct test preparation on the TOEIC scores of Japanese university students. *TESL-EJ*, 3(4), 1–22.
- Sato, T. (2019). An investigation of factors involved in Japanese students' English learning behavior during test preparation. *Papers in Language Testing and Assessment*, 8(1), 69–95.
- Shackleton, C. (2020). Planning for positive washback: The case of a listening proficiency test. In *Perspectives on Language Assessment Literacy* (pp. 220-240). Routledge.
- Shirzadi, D. & Amerian, M. (2020). Washback effects of multiple-choice, cloze and metalinguistic tests on EFL students writing. *Indonesian Journal of Applied Linguistics*, 9(3), 536–544.
- Shohamy E, Donitsa-Schmidt S. & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13, 298–317.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.

- Trinity College London (2015). *Integrated Skills in English (ISE) Guide for Teachers—ISE II (B2)*.
- Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In E. Shohamy, I.G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 359–372). Springer International Publishing AG.
- Umashankar, S. (2017). *Washback effects of speaking assessment of teaching English in Sri Lankan schools* [Unpublished doctoral dissertation]. The University of Bedfordshire.
- Wall D (1996). Introducing new tests into traditional systems: Insights from general education and innovation theory. *Language Testing*, 13, 231–354.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge University Press.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- Watanabe, Y. (2003). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp.16–36). Mahwah, New Jersey: Lawrence Erlbaum.
- Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self-theories. *Assessment in Education: Principles, Policy & Practice*, 21(1), 71–89.



2022 by the authors. Licensee Journal of Teaching English Language (TEL). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution–NonCommercial 4.0 International (CC BY-NC 4.0 license). (<http://creativecommons.org/licenses/by-nc/4.0>).