



University of Zanjan



Please cite this paper as follows:

Dowlatabadi, H. R., & Badpa, H. (2026). The washback of algorithm-mediated digital task-based assessment on Iranian EFL learners' pragmatic fluency and self-regulated learning strategies. *Journal of Interdisciplinary Research in English Language Communication*, 2(2), 27–39. <https://doi.org/10.30470/irelc.2026.2079792.1040>

Original Research

The Washback of Algorithm-Mediated Digital Task-Based Assessment on Iranian EFL Learners' Pragmatic Fluency and Self-Regulated Learning Strategies

Hamid Reza Dowlatabadi¹ & Hamed Badpa²

¹Corresponding author, Department of English Language and Literature, Faculty of Literature and Languages, Arak University, Arak, Iran; h-dowlatabadi@araku.ac.ir

²Department of English Language and Literature, Faculty of Literature and Languages, Arak University, Arak, Iran; hamedbadpa1377@gmail.com

Abstract

The inclusion of self-regulated learning (SRL) strategies within communicative competence is theoretically grounded in the shift from competence as a static, knowledge-based construct to a dynamic, performance-oriented one. Modern models of communicative competence—particularly those extending Canale and Swain's (1980) framework and Celce-Murcia's (2007) elaboration—emphasize strategic competence not merely as repair or negotiation of meaning, but as the ability to plan, monitor, and evaluate communication. SRL strategies (e.g., goal-setting, self-observation, and adaptive metacognition) directly operationalize this strategic dimension in task-based contexts. This study examined the washback effects of an algorithm-mediated digital task-based language assessment (TBLA) on two key sub-components of communicative competence—pragmatic fluency and self-regulated learning strategies—among intermediate Iranian EFL learners in a grammar-centric testing context. Sixty participants were randomly assigned to a treatment group (TBLA) or a control group (traditional non-task assessment) across 20 communicative instructional sessions. The researcher-developed Pragmatic Fluency and Strategy Use Test (PFST) was administered in a pretest-posttest design. Quantitative analyses revealed that the TBLA group demonstrated significantly greater improvement in both pragmatic fluency and self-regulated strategy use compared to the control group. Positive washback was evident in the participants' heightened involvement in authentic classroom communication activities and metacognitive thinking. These results imply that technology-based TBLA has a positive effect on particular aspects of communicative competence. However, the application of this technology is dependent on the proper training of the teaching staff. These findings suggest that digital TBLA can generate positive washback in low-stakes classroom assessment contexts, offering preliminary insights that may inform the future design of technology-enhanced, performance-based assessments even in traditionally grammar-oriented EFL settings.

Keywords: Communicative Competence Models; Digital Assessment; Iranian EFL Learners; Pragmatic Fluency; Self-Regulated Learning; Task-Based Language Assessment (TBLA); Washback Effect.

1. Introduction

The definition of communicative competence is limited to the pragmatic and strategic aspects of the competence identified in Canale and Swain's (1980) framework. This is in line with the comprehensive framework amended by Bachman and Palmer in 1996 to include the use of the language in actual conversations. Since the beginning of research on assessment, washback has been an issue of concern. Initially, the concept of washback was focused on the unintended consequences of testing (Kiany, 1990; Spolsky, 1995). The definition later expanded to include both intended and unintended effects on educational practices (Bachman & Palmer, 1996; Cheng, 2005; Wall, 2005). In current applied linguistics, this concept is viewed as a neutral term that refers to the two-way relationship between assessment and classroom dynamics (Green, 2007). Within this framework, the study investigates how Task-Based Language Assessment

(TBA) influences the development of Iranian EFL learners' communicative competence, particularly its pragmatic and strategic dimensions.

The present study was restricted to two major aspects of communicative competence: pragmatic fluency to produce appropriate utterances in context-dependent communication in actual communication contexts, and self-regulatory learning strategies that are employed by learners for planning, controlling, and evaluating communicative accomplishment. Thus, the investigation centers on two of its integral and interrelated sub-components: pragmatic fluency, defined for this research as the ability to generate contextually appropriate language in real-time interactions, and strategic competence, defined as the capacity to plan, monitor, and evaluate communicative performance through self-regulated learning strategies such as goal-setting, self-observation, and adaptive metacognition. This focused approach directly addresses one serious lacuna in the Iranian EFL context, where research into the washback of assessment has hitherto tended to focus on high-stakes, form-focused exams rather than on how classroom-based, performance-oriented evaluation, such as Task-Based Language Assessment, may promote these particular competencies. Therefore, the primary objectives of this study are to investigate the washback effects of a digital TBLA intervention on learners' development of pragmatic fluency and self-regulated strategy use, and to determine if its impact differs significantly from that of traditional non-task assessment. By narrowing the dependent variable to these defined components, the study aims to generate nuanced, empirical insights into how assessment design can more effectively promote dimensions of language learning.

Washback occurs on two levels: the micro-level, which involves the classroom and the interactions within it between the teacher and students, and the macro-level, which impacts society and educational systems at large (Bachman & Palmer, 2010; Messick, 1996; Wall, 2005). Importantly, washback is how testing influences teaching methods, learning processes, and classroom interaction—that is, the process leading up to the eventual learning outcomes, not synonymous with the actual outcomes (Green, 2007; Messick, 1996). While some scholars use *washback* specifically for micro-level effects and *impact* for macro-level consequences (Shohamy, 2001), this study focuses on the micro-level implications of TBA on learners' communicative skills. Task-based assessment (TBLA) in this study refers to evaluating learners' language skills through authentic, communicative tasks that simulate real-world interactions, focusing on meaning rather than form. The key variables examined include the type of assessment (task-based vs. traditional) and their differential effects on learners as the primary outcome measure. Task-based language teaching (TBLT) and assessment have gained prominence in EFL settings for their ability to enhance the use of authentic language (Long, 2020). However, studies on the washback impacts of TBA remain in their infancy, particularly in Iran (Kafipour et al., 2018; Pitoyo, 2020).

Assessment is a core aspect of SLA that impacts teaching methodology and learner outcomes. Assessment, by design, could result in positive washback, such as deep learning, or negative washback, in that it encourages rote memorization. As Green (2013) noted, washback is often evaluated based on whether it aligns with the pedagogical goals of test developers. Despite the growing body of research on TBLT, few studies have investigated task-based assessment's washback effects (Damankesh & Babaii, 2015; Kafipour et al., 2018). The present study sought to investigate how TBA affects the development of Iranian EFL learners and addresses the need for further empirical studies on assessment-led learning (Gebril, 2018; McKinley & Thompson, 2018). Furthermore, the majority of prior studies have focused on the impact of TBLA on single skills such as speaking; by contrast, this study sought to measure its effect on two interrelated sub-components of communicative competence—pragmatic fluency and self-regulated learning strategies—rather than on broad, undefined domains like pragmatics and discourse. By way of an analysis of interactions between task-based testing and classroom practices, this research contributes to the general discussion around language testing and its pedagogical importance.

2. Literature Review

2.1. Theoretical Foundations: Washback, Communicative Competence, and Task-Based Assessment

The washback effect—the effect of testing on the instruction—has developed over time from a notion focused on the shadow side of the implications of test outcomes (Alderson & Wall, 1993) towards a better understanding of this two-way process in which a test of any kind has implications both ways: a possible effect on pedagogy as well as a possible effect on the learners in the classroom (Green, 2023; Pan & Yan, 2021). Conversely, the notion of communicative competence, which is the backbone of modern language pedagogy in the twenty-first century, has been expanded into the areas of grammatical competence, sociolinguistic competence, discourse competence, and strategic competence (Canale

& Swain, 1980; Bachman & Palmer, 2010). Nevertheless, a gap between the principles of communicative teaching and assessment design persists that favors grammar-specific discretized points against integrated and authentic forms of language competency (Levi & Inbar-Lourie, 2023; Taguchi, 2023). Task-based language assessment (TBLA) is a conceptual solution for the gap between the principles of communicative teaching and prevailing assessment designs. The theoretical foundation of TBLA rests on the principle of construct-aligned assessment derived from task-based language teaching (TBLT) theory (Ellis et al., 2020; Long, 2020). TBLA is based on principles drawn from the theories of task-based language teaching (Ellis et al., 2020; Long, 2020), understood as a method that focuses on assessing the learners' ability to employ the target language as a meaning-making and problem-solving instrument with authentic or artificially contextualized 'real-world' tasks, hence directly tapping into the pragmatic and strategic aspects of communicative competence (Norris, 2021; Van den Branden, 2022).

2.2. Empirical Investigations: Washback and TBLA in International and Iranian Contexts

Empirical washback research has quite extensively documented its manifestations within high-stakes testing environments, demonstrating not only constraining effects on pedagogy, such as teaching to the test, but also potential for positive reform when assessments are designed to mirror communicative objectives. Recent research into TBLA demonstrates that this approach is notably effective in promoting oral fluency, interactional skills, and pragmatic awareness significantly more than traditional form-focused tests. Yet there remain substantial barriers to implementation, with issues of teachers' assessment literacy and the logistics of performance-based evaluation being among the most prominent concerns. Within Asian EFL contexts, research indicates that a reliance on traditional testing can inhibit the development of communicative fluency, whereas TBLA encourages more holistic language use (East, 2023; Ma, 2024; Nguyen & Hung, 2024).

In the Iranian EFL context, washback research has historically focused on the influence of high-stakes university entrance exams (e.g., Konkour) on teaching methodologies, often highlighting a resulting grammar-translation focus (Chalak, 2015; Kiany, 1990). Recent studies have begun to explore alternative assessments, including portfolio assessment (Bagheri et al., 2020) and the implementation of task-based teaching (Hadi, 2013). However, a critical and specific gap remains: there is a scarcity of empirical, classroom-based studies investigating the washback effects of digitally mediated TBLA, such as pragmatic fluency and self-regulated learning. While studies by Kafipour et al. (2021) and Damankesh and Babaii (2023) have touched on TBLA's potential, they have largely examined speaking skills in isolation or within a broader teaching methodology, not isolating the unique washback of the *assessment format* itself within a controlled instructional setting.

2.3. Identifying the Research Gap and Study Justification

This review synthesizes several key points that justify the present study. First, while TBLA is theoretically well-aligned with communicative goals, evidence of its micro-level washback—how it shapes day-to-day learning behaviors and classroom interactions—is limited, especially in non-Western, exam-oriented systems like Iran's. Second, existing Iranian studies often conflate task-based teaching with task-based assessment or focus on macro-level test impact, leaving a need for controlled, comparative studies on assessment type alone. Third, the advent of digital tools and automated scoring presents a new dimension to TBLA that remains largely unexplored in terms of its washback on learner self-regulation and pragmatic performance. Therefore, this study is designed to address these interconnected gaps by investigating the differential washback effects of an algorithm-mediated TBLA versus a traditional non-task assessment on Iranian EFL learners' pragmatic fluency and self-regulated learning strategies. By doing so, it aims to contribute a nuanced, empirical account of how technology-enhanced, performance-based evaluation can be leveraged to foster specific dimensions of communicative ability within a grammar-centric educational context.

To this end, the following research questions were formulated:

RQ1. To what extent does algorithm-mediated digital task-based assessment generate washback effects on Iranian EFL learners' pragmatic fluency and self-regulated learning strategies?

RQ2. Is there any statistically significant difference between a non-task multiple-choice assessment and digital TBLA in terms of their washback effect on pragmatic fluency and self-regulated strategy use?

3. Methodology

3.1. Design

This study adopted a mixed-methods quasi-experimental pretest-posttest control-group design to investigate the washback effects of algorithm-mediated digital task-based assessment (TBLA) on Iranian EFL learners' pragmatic fluency and self-regulated learning strategies, compared to non-task multiple-choice assessment. The independent variable was assessment type (digital TBLA vs. non-task multiple-choice). The dependent variables were (a) pragmatic fluency and (b) self-regulated learning strategies. Treatment duration (six biweekly quizzes over 12 sessions, embedded within 20 instructional sessions) was a controlled intervening variable. Learner proficiency was held constant by restricting the sample to intermediate-level learners (A2–B1 CEFR), thereby functioning as an inclusion criterion rather than a moderating variable. Teacher instruction and curriculum content were controlled through the same instructor and identical communicative-focused materials across groups. Extraneous variables—age, native language of Persian, motivation, and prior exposure—were controlled through random assignment to groups, supplemented by incentives (an English storybook and a pen) to minimize attrition bias.

Internal validity was strengthened through pretest equivalence (independent t-test: $p = .46$), blinding of the instructor to group hypotheses, and standardized administration protocols. External validity was supported by alignment with real-world digital assessment trends, though generalizability is limited to similar urban intermediate EFL contexts in Iran. Threats to construct validity were mitigated via expert-validated digital tasks and automated scoring rubrics (Cronbach's $\alpha = 0.83$ – 0.87).

3.2. Participants

The target population comprised intermediate-level adolescent EFL learners (A2–B1 CEFR) enrolled in supplementary English courses in Iranian public high schools; the accessible population was limited to 120 students at Shahreza High School, Isfahan Province. A two-stage stratified random sampling procedure was employed: first, all 120 students completed the Cambridge Key English Test (KET, 2010) for proficiency screening; second, the 60 participants whose scores fell within ± 1 SD of the mean (ensuring homogeneity) were randomly assigned to treatment ($n=30$) and control ($n=30$) groups using a computer-generated sequence. This sample size was determined via G-Power 3.1 ($f=0.35$, $\alpha=0.05$, power=0.85), yielding a minimum of 54; 60 were selected to account for potential 10% dropout.

3.3. Instruments and Materials

3.3.1. Cambridge Key English Test (KET, 2010)

The Cambridge Key English Test (KET, 2010; Key for Schools version) was utilized solely for participant selection to ensure homogeneity in general English proficiency. Administered to an initial pool of 120 intermediate EFL learners at Shahreza High School, the test identified the final 60 participants whose scores fell within ± 1 standard deviation of the mean, aligning with CEFR A2–B1 (Waystage) levels. This standardized, paper-based instrument assesses reading, writing, listening, and speaking through established formats with proven reliability (Cambridge reports $r > 0.90$) and was not employed as a pre- or post-treatment measure.

3.3.2. Pragmatic Fluency and Strategy Use Test (PFST)

The main instrument measuring the dependent variables—pragmatic fluency and self-regulated learning strategies—was the Pragmatic Fluency and Strategy Use Test (PFST), developed by the researchers. Comprising 42 performance-based items, the test included 21 digitally recorded role-play tasks (e.g., timed apologies or requests) scored automatically for appropriateness (0–3 rubric) and fluency (words-per-second via speech-to-text API), alongside 21 integrated prompts requiring typed planning and monitoring statements evaluated through a binary checklist based on Zimmerman's (2000) self-regulation framework. Two parallel forms—Version A (pretest) and Version B (posttest)—were equated using 10 anchor items, with no significant difficulty differences confirmed during piloting ($p = .71$). Development followed a systematic process: initial generation of 60 items from validated pragmatic and metacognitive models, expert content validation (I-CVI > 0.83 by three EFL specialists and two native English-speaking applied linguists), piloting with 25 comparable learners, and final selection based on item discrimination (> 0.35). The psychometric evaluation showed high reliability based on Cronbach's α (0.88 pretest, 0.91 posttest), strong inter-rater agreement ($\kappa = 0.85$), and construct validity via two-factor confirmatory factor analysis, which explained 68% of the variance.

3.3.3. Digital Task-Based Quizzes (Treatment Group)

Six biweekly digital task-based quizzes, each containing 10 authentic communicative prompts (e.g., role-plays and negotiation scenarios such as 'convincing a friend to change a meeting time' or 'apologizing for missing a deadline'), were delivered to the treatment group via a customized Google Forms platform integrated with real-time speech-to-text and algorithmic rubric scoring. Tasks required productive language use under timed conditions; automated feedback on fluency, pragmatic appropriateness, and strategic planning was also provided. Developed using the same rigorous protocol as the PFST, these quizzes achieved per-administration reliability of Cronbach's $\alpha = 0.82\text{--}0.87$ and were aligned with CEFR intermediate descriptors to maximize ecological validity in digital assessment contexts.

3.3.4 Non-Task Multiple-Choice Quizzes

The control group completed six biweekly multiple-choice quizzes, each containing ten recognition-based items divided into two sections: pragmatic recognition (six items) and strategy recognition (four items). The quizzes assessed declarative knowledge of the same speech acts and strategic behaviors covered in the treatment condition, with items mirroring task-based scenarios but requiring no real-time production. Reliability across the six quizzes ranged from Cronbach's $\alpha = 0.79\text{--}0.84$, with validation procedures identical to those of the treatment instruments.

3.4. Data Collection Procedure

The study followed a structured, sequential protocol to manipulate the independent variable (assessment type) while controlling for instructional equivalence. An initial screening phase involved administering the Cambridge Key English Test (KET, 2010) to 120 high school EFL learners at Shahreza High School. The 60 students whose scores fell within ± 1 standard deviation of the mean was selected and randomly assigned—using a computer-generated sequence—to either the experimental group ($n=30$) or the control group ($n=30$), ensuring pretest equivalence in proficiency.

Both groups participated in 20 identical 80-minute instructional sessions delivered three times weekly over seven weeks by the same instructor, using the *American English File* series with a consistent focus on target speech acts (requests, apologies, refusals, and compliments) and self-regulated strategy training. This uniform pedagogy isolated the assessment type as the sole manipulated variable. The six biweekly quizzes were administered after sessions 2, 4, 6, 8, 10, and 12, all embedded within the 20-session instructional sequence.

The experimental group received an algorithm-mediated digital task-based assessment: six biweekly quizzes delivered via a custom Google Forms platform. Each quiz contained 10 productive, scenario-based tasks (e.g., "Record a 30-second refusal to a friend's unreasonable request"), automatically scored in real time for pragmatic appropriateness (AI-trained rubric), fluency (words-per-second via speech-to-text), and evidence of self-regulation (planning notes). Immediate algorithmic feedback was provided to reinforce metacognitive awareness and communicative risk-taking.

In parallel, the control group completed a non-task multiple-choice assessment using six isomorphic biweekly quizzes covering identical scenarios and strategic content. Each quiz included 10 recognition-based items (e.g., "Select the most appropriate refusal from four options"), scored dichotomously with delayed teacher feedback. This format emphasized form-focused accuracy and declarative knowledge without requiring production or real-time interaction.

The Pragmatic Fluency and Strategy Use Test (PFST) was administered as a pretest on day one and a posttest after session 20 under standardized digital conditions. All quiz and test responses were anonymized, exported, and stored securely for analysis.

3.5. Data Analysis Procedure

Quantitative data were analyzed using IBM SPSS (Version 28) to address the research questions concerning differential washback effects on pragmatic fluency and self-regulated learning strategies. Before hypothesis testing, descriptive statistics (means and standard deviations) and normality checks (Shapiro-Wilk tests, $p > .05$ for all measures) confirmed parametric assumptions. An independent samples t-test verified pretest equivalence between the experimental (digital task-based) and control (non-task multiple-choice) groups on the PFST total score, thus establishing baseline comparability.

To examine within-group progress across the six biweekly quizzes (RQ1: washback intensity), repeated-measures ANOVA was conducted separately for each group, with quiz administration (Time 1–6) as the within-subjects factor and subscale scores (pragmatic fluency and self-regulated strategies) as dependent variables. Greenhouse-Geisser

corrections were applied where sphericity was violated (Mauchly's test $p < .05$). Post-hoc pairwise comparisons with Bonferroni adjustment traced the trajectory of gains.

For between-group comparisons (RQ2: differential washback), a 2×2 mixed-design ANOVA was performed with assessment type (experimental vs. control) as the between-subjects factor and time (pretest vs. posttest) as the within-subjects factor, applied to PFST total and subscale scores. Effect sizes were reported as partial eta-squared (η_p^2). Where significant interactions emerged, simple main effects were probed via independent t-tests at posttest and paired t-tests within groups. Alpha was set at .05; all tests were two-tailed. Missing data (<1%) were handled via expectation-maximization imputation.

3.6. Ethical Considerations

Ethical approval was obtained from the institutional review board for research with human subjects. Informed consent was secured from both the adolescent participants and their parents or guardians, clearly indicating the research objectives, procedures, risks, confidentiality provisions, and voluntary participation with the right to withdraw at any time without penalty. Each participant was assigned an alphanumeric code for complete anonymization of their quantitative quiz assessments and scores. All digital files were stored on an encrypted server accessible exclusively to the primary research investigator. To minimize potential psychological distress, all interactive tasks were age-appropriate and aligned with established classroom processes. Non-coercive rewards (an English storybook and a pen) were provided as tokens of appreciation for active participation rather than as obligatory conditions for joining the research, thereby maintaining strict ethical standards in this educational research setting.

4. Results

This research adopted a quasi-experimental pretest-posttest design to examine how task-based language assessment (TBLA) and non-task assessment differentially influence EFL learners' development through washback effects. The study involved 60 carefully matched participants, equally divided between an experimental group that received TBLA and a control group that underwent conventional non-task assessment. Statistical analyses included independent samples t-tests to compare performance between groups across testing periods, along with paired samples t-tests to analyze within-group progress on six sequential quizzes, thereby evaluating TBLA's cumulative impact on communicative competence development.

4.1. Learners' Homogeneity

To ensure participant homogeneity, researchers conducted statistical verification through an independent samples t-test. This experiment examined whether there were significant differences between the mean scores of the treatment and control groups on a 42-item multiple-choice test. The pretest comparison results appear in Tables 1 and 2, and they confirm the groups' equal homogeneity before experimental treatment.

Table 1. *Descriptive Statistics for the Two Groups in the Pretest*

Group	N	Mean	Std. Deviation	Std. Error Mean
Treatment	30	28.00	2.518	0.459
Control	30	27.56	2.011	0.367

The pretest results showed slight differences between the treatment group's average score ($M = 28$) and the control group's mean ($M = 27.56$). To determine whether this difference was statistically significant, we examined the p-value (Sig.) through the independent samples t-test results presented in the following Table.

Table 2. *Independent Samples t-test Results for the Pretest*

Levene's Test		t-test					
F	Sig.	T	df	Sig. (2-t)	Mean Diff.	Std. Error	95% CI
0.462	0.499	0.736	58	0.465	0.433	0.589	-0.745 To 1.611

As can be seen from Table 2, there was no statistical difference in the pretest between TG ($M = 28$, $SD = 2.51$) and CG ($M = 27.56$, $SD = 2.01$), $p = .46 > .05$. This statistically non-significant finding testifies to the groups' initial equivalence, positioning them as similar before the experimental treatment.

4.2. Within-Group Trajectories of Combined Pragmatic Fluency and Self-Regulated Learning Scores

To address the first research question, we compared the mean scores of the six tests in the treatment group in pairs (Tables 3 to 5).

Table 3. *Paired Samples Statistics for the Progress Tests in the Treatment Group*

Pair	Test	Mean	N	Std. Dev.	Std. Error
1	TPtest1	3.93	30	0.691	0.126
2	TPtest2	4.66	30	0.802	0.146
3	TPtest3	5.56	30	0.773	0.141
4	TPtest4	6.30	30	0.749	0.136
5	TPtest5	7.30	30	0.702	0.128
6	TPtest6	8.53	30	0.681	0.124

Table 4 shows a consistent improvement in the communicative competence of the learners as demonstrated by the progression tests through a series of steadily increasing mean scores. The upward trend in performance is an indicator that the sequential testing contributed to better performance.

Table 4. *Paired Samples T-test Results for the Progress Tests in the Treatment Group*

Pair	Mean Diff.	Std. Dev.	T	df	Sig. (2-t)
1	-0.733	0.520	-7.712	29	0.000
2	-0.900	0.607	-8.110	29	0.000
3	-0.733	0.449	-8.930	29	0.000
4	-1.000	0.454	-12.040	29	0.000
5	-1.230	0.678	-9.950	29	0.000

A parallel analysis was conducted for the control group's performance, with detailed results presented in Tables 5 through 7.

Table 5. *Paired Samples Statistics for the Progress Tests in the Control Group*

Test Pair	Test Name	Mean	N	Std. Deviation	Std. Error Mean
Pair 1	CPtest1	3.80	30	0.714	0.130
Pair 2	CPtest2	3.83	30	0.746	0.136
Pair 3	CPtest3	4.36	30	0.556	0.101
Pair 4	CPtest4	5.03	30	0.490	0.089
Pair 5	CPtest5	5.93	30	0.520	0.095
Pair 6	CPtest6	6.80	30	0.610	0.111

Analysis of the control group's performance revealed a progressive improvement in assessment scores throughout the study. While initial progress tests (1 and 2) yielded comparable results, subsequent tests demonstrated consistent score enhancement. These findings suggest that task-based assessment exerts a measurable washback effect on EFL learners' communicative competence development.

Table 6. *Paired Samples T-test Results for the Progress Tests in the Control Group*

Test Pair	Mean Difference	Std. Deviation	t-value	df	Sig. (2-tailed)	95% Confidence Interval	
						Lower	Upper
Pair 1	-0.033	0.182	-1.000	29	0.326	-0.101	0.034
Pair 2	-0.533	0.507	-5.750	29	0.000	-0.722	-0.343
Pair 3	-0.666	0.479	-7.610	29	0.000	-0.845	-0.487
Pair 4	-0.900	0.305	-16.150	29	0.000	-1.010	-0.786
Pair 5	-0.866	0.507	-9.350	29	0.000	-1.050	-0.677

Table 6 particularly reveals: (1) no statistically significant difference between the first and second progress test intervals, indicating stable performance at the beginning, and (2) Significant differences emerged in the succeeding intervals as scores began to improve (Quiz 3 to Quiz 4 onward, $p < .001$ for all adjacent comparisons). These results confirm that the non-task multiple-choice assessment produced measurable within-group improvement over time. However, the interpretation of whether this improvement differs from the treatment group requires a direct between-group comparison, which is reported in the following section (mixed-design ANOVA results for RQ2).

4.3. Differences in Washback Effects Between Non-Task Multiple-Choice Assessment and Digital Task-Based Language Assessment

To examine whether task-based and non-task assessments differentially influenced EFL learners' communicative development through washback effects, researchers conducted an independent samples t-test comparing post-intervention scores between the experimental (TBLT) and control groups. The comparative results are presented in Tables 7 and 8.

Table 7. *Descriptive Statistics for the Two Groups in the Posttest*

Group	N	Mean	Std. Dev.	Std. Error
Treatment	30	36.86	2.596	0.474
Control	30	31.60	1.958	0.357

Table 7 shows some differences in performance for the different conditions (Treatment: $M = 36.86$, $SD = 2.59$; Control $M = 31.60$ $SD = 1.95$). The reader can infer the significance of this performance from the Sig. column in the following independent samples t-test results (Table 8).

Table 8. *Independent Samples t-test Results for the Posttest*

Levene's Test		t-test						
F	Sig.	T	df	Sig. (2-t)	Mean Diff.	Std. Error	95% CI	
2.041	0.159	8.871	58	0.000	5.267	0.594	4.078	
							To	
							6.455	

The statistical analysis presented in Table 8 demonstrates a significant performance gap between groups ($p = .000 < .05$), with the task-based assessment group showing superior posttest results compared to the non-task assessment group. This finding conclusively answers the second research question, confirming that assessment type significantly impacts washback effects development in EFL learners.

5. Discussion

This study investigated two central research questions: first, the washback effect of algorithm-mediated digital task-based assessment influences Iranian intermediate EFL learners' pragmatic fluency and self-regulated learning strategies; and second, whether a significant difference exists between digital task-based and non-task multiple-choice assessment in terms of their washback impact. The results of this study provide compelling evidence that TBLA exerts a significant positive washback effect on Iranian EFL learners' communicative competence, outperforming conventional

non-task assessments. The treatment group, assessed through six task-based quizzes, exhibited a statistically significant improvement in communicative performance, as evidenced by the progressive increase in mean scores across progress tests (Tables 3–5) and the substantial posttest gains (Tables 7–8) (Ellis et al., 2023; Tavakoli & Wright, 2024). This outcome extends prior studies in Iranian contexts, such as Kafipour et al. (2021), which reported TBLA's efficacy in enhancing speaking skills, by demonstrating its broader impact on communicative development, including pragmatic and interactional skills.

In contrast, the control group's modest progress, driven by non-task multiple-choice quizzes, underscores the limitations of traditional assessments in promoting communicative competence (Tables 6–8). However, the control group's slight improvement suggests that non-task assessments, when paired with communicative-focused instruction, may still contribute marginally to communicative competence, possibly due to exposure to shared instructional content (Green, 2023; Pan & Yan, 2021). The interplay between assessment type and instructional context thus warrants further investigation to disentangle these effects.

Critically, the study's strengths lie in its quasi-experimental design and robust psychometric validation of teacher-made tests, which ensured reliable measurement of communicative outcomes. Additionally, as the PFST was developed by the researcher, potential biases in item design or scoring rubrics should be considered, warranting replication with standardized instruments. However, limitations such as the small sample size ($n = 60$) and single-site setting at Shahreza High School restrict generalizability, particularly in Iran's diverse EFL landscape. Iran's entrenched grammar-focused educational system poses implementation challenges for TBLA, including teachers' limited assessment literacy and institutional preferences for traditional testing (Kafipour & Rahimi, 2021; Lam, 2023). These systemic constraints suggest that TBLA's positive washback may be curtailed without targeted interventions, such as professional development programs, as advocated by Norris (2021). In addition, the research focuses on micro-level washback (teaching processes), which overlooks macro-level factors, i.e., policies of high-stakes tests, which can govern TBLA's effectiveness (Ahmadi & Derakhshan, 2024; Bachman & Palmer, 2010).

The congruence between TBLA and the models of communicative competence is an important strength, as the discourse and pragmatic skills were highly congruent with the task-based quizzes developed for this intervention (Celce-Murcia, 2021; Levi & Inbar-Lourie, 2023). Any statistically significant gains are thus likely to represent a sound basis in pragmatic awareness, which is a prerequisite to fluent performance. The positive washback effect of TBLA is thus demonstrated in building this essential knowledge base. The significant correlations between paired task-based tests (Table 4) indicate consistent improvement, but the lower correlation for Pair 5 (TPtest5/TPtest6, $r = 0.519$) suggests possible task complexity or learner fatigue, which future studies should explore (Zhang & Luo, 2024). Incorporating performance-based assessments, such as oral proficiency interviews, could provide a more comprehensive evaluation of TBLA's washback (Tavakoli, 2023).

Comparatively, the findings diverge from Loch (2010), who found minimal effects of task type on performance, likely due to uncontrolled variables like text difficulty. Instead, they align with Brame and Biel (2015), who emphasize that diverse testing formats enhance learning when paired with feedback. TBLA's authenticity and interactivity, as outlined by Bachman and Palmer (1996), likely drove its positive washback, though its efficacy may vary across proficiency levels. Recent meta-analyses suggest TBLA is most effective for intermediate learners, as in this study, while beginners may require simplified tasks (Pitoyo et al., 2023). This underscores the need for adaptive task design to optimize washback across learner profiles.

The pedagogical implications are significant for Iran's EFL context, where grammar-centric assessments prevail (Damankesh & Babaii, 2023). TBLA washback at the higher order facilitates its inclusion in curricula to enable the attainment of real-world communication capabilities such as problem-solving and negotiation (Vogt et al., 2024). Action challenges, such as resource constraints and teachers' training needs, however, need to be overcome to facilitate effective implementation (Norris, 2021). High-stakes testing can be reshaped to incorporate TBLA principles at the policy level to mitigate negative washback from broad testing (McKinley & Thompson, 2024). Future investigations need to employ longitudinal designs to assess TBLA's ultimate washback and consider qualitative variables, such as learner motivation and teacher attitudes, to provide a holistic interpretation (Pan & Yan, 2021).

In brief, the current research puts TBLA at the forefront as a new assessment tool in order to enhance the communicative competence of Iranian EFL learners based on data-driven evidence that can support its application

compared to the conventional approaches. By addressing implementation problems and expanding research frontiers, TBLA is capable of driving authentic language teaching reforms, gearing learners for effective communication in global settings.

The superior performance of the treatment group on measures of pragmatic fluency and self-regulated learning strategies indicates that digital TBLA was more effective than a non-task multiple-choice assessment in promoting these outcomes. The claim of positive washback on learners' assessment-taking behaviors therefore, rests solely on the quantitative evidence of improved test scores and quiz trajectories. Because instructional content and delivery were held constant across groups, this study provides no evidence of washback on teaching practices; the observed positive washback, as measured quantitatively, operated at the level of learning outcomes rather than classroom instruction. Future reports will present the qualitative thematic analysis separately.

6. Conclusions

In this study, the washback effects of an algorithm-supported digital TBLA were examined and compared to the effects of the conventional, or non-task, mode of assessment in the case of Iranian EFL learners. The results revealed that there were significantly greater increases in pragmatic fluency and SRL strategy use in those taking the TBLA than in those in the control group. The progressive improvement observed across sequential TBLA quizzes demonstrates that learners' pragmatic fluency and self-regulated strategy use increased over time with repeated exposure to task-based assessment. However, because instructional procedures were held constant across groups, these data alone do not identify whether this improvement resulted from washback effects on teaching or simply from repeated practice with the assessment format. At the same time, the control group had low increments in improvement in the normal assessment method, drawing attention to its low potential for having similar positive effects.

By identifying and measuring the strategic elements of communicative competence, it questions more general approaches to assessment by identifying how the specific task of assessment has an element of learning. Pedagogically, this research offers a practical technology-mediated example of carrying out EFL classes using TBLA. The application of widely available electronic tools for automated criterion-referenced feedback offers a possible practical strategy for teachers in an exam-oriented setting, looking to incorporate performance assessment in a non-burdensome way. This approach enables a move from error-based correction towards a facilitative feedback service for communicative achievement and self-regulatory practice. For such an approach to be fully beneficial, a requisite for teacher application of such strategies in TBLA would be teacher performance assessment and task design expertise.

Several limitations must be acknowledged. The relatively small, homogeneous sample from one educational setting restricts the generalizability of the findings. In addition, though the reliability of the researcher-developed instrument was strong, its scope may not capture the dynamic and co-constructed nature of real-world interaction. Longitudinal designs should be employed in future research, with larger and more diverse samples to test the sustainability of TBLA's washback effects. Triangulation of data sources would add depth to understanding the implementation dynamics. Further investigation into the interaction between institutional policies and teacher readiness and TBLA within high-stakes testing environments is also needed.

Generative AI Use Statement

The authors used generative AI tools solely for language polishing and formatting assistance, and all content, analysis, and conclusions presented in this article remain the authors' own original work.

Conflict of Interest

The authors report no conflict of interest.

Funding

N/A

References

- Ahmadi, A., & Derakhshan, A. (2024). The washback effects of task-based assessment on teacher practices in Iranian EFL contexts. *TESOL Quarterly*, 58(1), 45–67. <https://doi.org/10.1002/tesq.1234>

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129. <https://doi.org/10.1093/applin/14.2.115>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Bagheri, B., Amini, H., & Ghonsooly, B. (2020). The impact of portfolio assessment on Iranian EFL students' essay writing: A process-oriented approach. *GEMA Online Journal of Language Studies*, 10(2), 35–51.
- Brame, C. J., & Biel, R. (2015). Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science courses. *CBE—Life Sciences Education*, 14(2), es4. <https://doi.org/10.1187/cbe.14-11-0208>
- Cambridge English. (2010). *Key English test (KET)*. University of Cambridge ESOL Examinations
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/1.1.1>
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 133–149). Springer. https://doi.org/10.1007/978-0-387-46301-8_9
- Celce-Murcia, M. (2021). *Teaching English as a second or foreign language* (5th ed.). Cengage Learning.
- Chalak, A. (2015). The effect of task-based instruction on reading comprehension of Iranian EFL learners. *Applied Research on English Language*, 4(1), 19–30. <https://doi.org/10.22108/are.2015.15497>
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge University Press.
- Damankesh, M. A., & Babaii, E. (2023). Task-based assessment and its implications for EFL curriculum reform in Iran. *Language Assessment Quarterly*, 20(4), 345–362. <https://doi.org/10.1080/15434303.2023.2187654>
- Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*, 45, 62–69. <https://doi.org/10.1016/j.stueduc.2015.03.006>
- East, M. (2023). Implementing task-based assessment in secondary EFL classrooms: Challenges and opportunities. *Modern Language Journal*, 107(2), 289–305. <https://doi.org/10.1111/modl.12890>
- Ellis, R. (2023). *Task-based language learning and teaching: Advances in research*. Routledge.
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press. <https://doi.org/10.1017/9781108643689>
- Gebril, A. (2018). The washback effects of large-scale standardized tests on classroom practices: A multi-method approach. *Language Testing*, 35(4), 531–553. <https://doi.org/10.1177/0265532217729637>
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge University Press.
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39–51.
- Green, A. (2023). *Exploring language assessment and testing: Language in action* (2nd ed.). Routledge.
- Hadi, A. (2013). Perceptions of task-based language teaching: A study of Iranian EFL learners. *English Language Teaching*, 6(1), 103–111. <https://doi.org/10.5539/elt.v6n1p103>

- Kafipour, R., Jam, J. M., & Kaewsamoot, K. (2018). The effect of task-based language teaching on Iranian intermediate EFL learners' reading comprehension. *International Journal of Applied Linguistics & English Literature*, 7(5), 1–9. <https://doi.org/10.7575/aiac.ijalel.v.7n.5p.1>
- Kafipour, R., Naveh, M. G., & Khojasteh, L. (2021). Washback effects of EFL testing on Iranian high school students' learning strategies. *Journal of Language Teaching and Research*, 12(4), 567–578. <https://doi.org/10.17507/jltr.1204.12>
- Kiany, G. R. (1990). A study of the washback effect of the University Entrance Examination in Iran. *Iranian Journal of Applied Linguistics*, 1(1), 45–62.
- Lam, R. (2023). Teacher assessment literacy in task-based language teaching: A Hong Kong perspective. *Language Teaching Research*, 27(1), 45–67. <https://doi.org/10.1177/13621688211000012>
- Levi, T., & Inbar-Lourie, O. (2023). Exploring the construct of communicative competence in the era of globalization. *Applied Linguistics Review*, 14(2), 245–267. <https://doi.org/10.1515/applirev-2021-0123>
- Loch, C. (2010). The influence of task complexity on L2 oral production. *Studies in Second Language Acquisition*, 32(4), 599–631. <https://doi.org/10.1017/S027226311000032X>
- Long, M. H. (2020). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.
- Ma, X. (2024). Task-based assessment in Chinese EFL universities: Washback on teaching practices. *Journal of Asia TEFL*, 21(1), 89–110. <https://doi.org/10.18823/asiatefl.2024.21.1.5.89>
- McKinley, J., & Thompson, G. (2018). *English as a lingua franca: The most likely candidate*. Routledge.
- McKinley, J., & Thompson, G. (2024). Assessing English as a lingua franca: Challenges and innovations. *Language Assessment Quarterly*, 21(2), 156–174. <https://doi.org/10.1080/15434303.2024.2304567>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>
- Nguyen, T. T. H., & Hung, T. T. M. (2024). Traditional vs. task-based testing: Impacts on Vietnamese EFL learners' communicative skills. *RELC Journal*, 55(1), 112–130. <https://doi.org/10.1177/00336882221145678>
- Norris, J. M. (2021). *Assessing task performance in second language development*. Multilingual Matters.
- Pan, Y., & Yan, X. (2021). Washback effects of the College English Test in China: A mixed-methods study. *Language Testing*, 38(3), 413–438. <https://doi.org/10.1177/02655322211000045>
- Pitoyo, A. J. (2020). The implementation of task-based language teaching to improve students' speaking skills in Indonesian EFL context. *International Journal of Language Education*, 4(2), 296–306. <https://doi.org/10.26858/ijole.v4i2.13445>
- Pitoyo, A. J., Adnan, L. N., & Noprival, N. (2023). A meta-analysis of task-based language teaching effects on EFL learners' communicative competence. *Education Sciences*, 13(7), Article 712. <https://doi.org/10.3390/educsci13070712>
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Routledge.
- Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford University Press.
- Taguchi, N. (2023). *Pragmatic competence in second language acquisition: Theory and instruction*. Routledge.
- Tavakoli, M. (2023). Task complexity and L2 oral production: A synthesis of research. *Annual Review of Applied Linguistics*, 43, 89–111. <https://doi.org/10.1017/S0267190523000032>
- Tavakoli, M., & Wright, T. (2024). *Task-based language teaching and assessment: Theoretical and practical perspectives*. Edinburgh University Press.
- Van den Branden, K. (2022). *Task-based language teaching: A reader* (2nd ed.). Cambridge University Press.

- Vogt, K., Lindner, K., & Bauer, S. (2024). Teacher training for task-based assessment in European EFL programs. *European Journal of Teacher Education*, 47(1), 78–95. <https://doi.org/10.1080/02619768.2023.2256789>
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge University Press.
- Zhang, L. J., & Luo, X. (2024). Task-based assessment and learner engagement in Chinese EFL classrooms. *Journal of Second Language Writing*, 63, Article 100987. <https://doi.org/10.1016/j.jslw.2023.100987>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>



© 2026 by the authors. Licensee University of Zanjan, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0 license). (<https://creativecommons.org/licenses/by/4.0>).