QO‘QON DAVLAT PEDAGOGIKA INSTITUTI ILMIY XABARLARI
(2026-yil 10-son)

PEDAGOGIKA
PEDAGOGY

https://doi.org/10.70728/c.series.ped.v08.i10.066

UDK 372.881.111.1

PREPARING UPPER-SECONDARY STUDENTS FOR CEFR-BASED EXAMINATIONS: CHALLENGES AND SOLUTIONS

Ne’matov Abdulloh
Farg‘ona davlat universiteti,
Amaliy ingliz tili kafedrasi o’qituvchisi
ORCID: https://orcid.org/0009-0008-8588-7451

Abstract. This article examines the challenges of preparing 9th and 10th grade students in general secondary schools for examinations based on the Common European Framework of Reference for Languages (CEFR). A mixed-methods design was employed with [186] students and [8] English teachers from [4] schools in Fergana region. Data were collected through a questionnaire, a CEFR-format mock examination and semi-structured interviews. The findings indicate that students’ writing and speaking scores were considerably lower than their listening and reading scores, that the school curriculum is not fully aligned with examination requirements, and that the washback of the examination on classroom practice is largely narrow and strategy-oriented. A four-stage model integrating examination preparation with communicative language teaching is proposed.

Keywords: CEFR, language competence, washback, test preparation, productive skills, upper-secondary students, formative assessment.

Annotatsiya. Ushbu maqolada umumta’lim maktablarining 9–10-sinf o‘quvchilarini CEFR (Chet tillarni egallashning umumevropa kompetensiyalari doirasi) asosidagi imtihonlarga tayyorlash jarayonida yuzaga keladigan muammolar aralash metodli tadqiqot orqali o‘rganiladi. Tadqiqotda Farg‘ona viloyatidagi [4] ta maktabning [186] nafar o‘quvchisi va [8] nafar ingliz tili o‘qituvchisi ishtirok etdi. Ma’lumotlar so‘rovnoma, CEFR formatidagi sinov imtihoni va yarim tuzilmali suhbatlar yordamida to‘plandi. Natijalar o‘quvchilarning yozish va gapirish bo‘yicha ko‘rsatkichlari tinglab va o‘qib tushunishga nisbatan sezilarli darajada pastligini, maktab dasturi bilan imtihon talablari o‘rtasida tafovut borligini hamda imtihonning dars jarayoniga ta’siri asosan test strategiyalariga yo‘naltirilgan tor shaklda namoyon bo‘layotganini ko‘rsatdi. Maqolada imtihonga tayyorlovni kommunikativ o‘qitish bilan uyg‘unlashtiruvchi to‘rt bosqichli model taklif etiladi.

Kalit so‘zlar: CEFR, til kompetensiyasi, washback effekti, imtihonga tayyorlov, produktiv ko‘nikmalar, yuqori sinf o‘quvchilari, shakllantiruvchi baholash.

Аннотация. В статье на основе исследования смешанного типа рассматриваются проблемы подготовки учащихся 9–10 классов общеобразовательных школ к экзаменам, основанным на Общеевропейских компетенциях владения иностранным языком (CEFR). В исследовании приняли участие [186] учащихся и [8] учителей английского языка из [4] школ Ферганской области. Данные собраны с помощью анкетирования, пробного экзамена в формате CEFR и полуструктурированных интервью. Результаты показали, что показатели учащихся по письму и говорению значительно ниже, чем по аудированию и чтению, школьная программа не в полной мере соответствует требованиям экзамена, а обратное влияние экзамена на учебный процесс носит преимущественно узкий, стратегически ориентированный характер. Предлагается четырёхэтапная модель, интегрирующая подготовку к экзамену с коммуникативным обучением.

Ключевые слова: CEFR, языковая компетенция, эффект обратного влияния, подготовка к экзамену, продуктивные навыки, старшеклассники, формирующее оценивание.

INTRODUCTION. Over the past fifteen years, Uzbekistan has reformed foreign-language education. Presidential Resolution No. PQ-1875 (2012) promoted foreign-language learning, while later reforms introduced CEFR-linked standards and certification. Consequently, certificates have become important for university admission, benefits, and career opportunities.

For many students, certification has become a major goal, sometimes taking priority over practical communication. This is especially visible in Grades 9 and 10, when students prepare for high-stakes examinations.

National standards expect learners to reach A2 by the end of basic secondary education and B1 by the end of general secondary education. Yet classroom practice does not always provide sufficient opportunities to develop all four skills.

This issue relates to washback: the influence of tests on teaching and learning. Although international research has examined washback extensively, evidence from Uzbekistan—particularly concerning upper-secondary students’ productive skills—remains limited.

This study identifies the main challenges in preparing Grade 9 and 10 students for CEFR-based examinations and proposes a practical approach to integrating preparation with communicative teaching. It addresses skill-level differences, factors associated with performance, and teachers’ and students’ experiences.

LITERATURE REVIEW

CEFR and its role in educational practice

Published by the Council of Europe in 2001, the CEFR describes language proficiency across six levels, from A1 to C2, and defines each level through descriptors that specify what learners “can do” using the language [4]. Its central principle is an action-oriented approach, in which the language learner is viewed as a social agent carrying out particular tasks. The 2020 Companion Volume updated the descriptors and introduced areas such as mediation, online interaction, and plurilingual competence [5]. Thus, the CEFR is not itself an examination; rather, it is a reference framework intended to align teaching, learning, and assessment through a shared conceptual system.

However, CEFR implementation varies across contexts. Little argues that it has become influential in language policy, although classroom practice may not fully reflect its action-oriented principles [6]. Figueras likewise highlights the challenges of translating CEFR descriptors into teaching and assessment [7].

The concept of washback and its contributing factors

The influence of examinations on teaching and learning is referred to in language testing as washback or backwash. In their influential article, J. C. Alderson and D. Wall argued that washback should not simply be assumed to exist but must be demonstrated empirically. They formulated hypotheses about how examinations may affect what teachers teach, how they teach, what learners learn, and how they learn [1]. S. Messick linked washback to test validity: when an examination adequately represents the competence being measured, preparation for it can help develop that competence; when the test covers only part of the competence, preparation may become restricted to that narrow area [8].

K. Bailey proposed a model of participants, processes, and products to explain washback [9]. According to this model, examinations first influence the beliefs of students, teachers, materials writers, and administrators; these beliefs then affect their actions and the processes of teaching and learning, which ultimately influence educational outcomes. A. Hughes identifies several ways to promote positive washback, including directly testing the skills that are intended to be developed, using criterion-referenced assessment, and explaining examination formats and requirements to students and teachers in advance [10].

Studies show that washback is complex and changes over time. High-stakes examinations can shape curricula and classroom priorities, but their effects depend on teachers’ beliefs, institutional conditions, and available resources [11; 13].

Examination-oriented teaching and communicative competence

Teaching to the test can narrow the curriculum to familiar item types and reduce opportunities for meaningful communication. This is especially problematic when learners practise grammar and vocabulary more often than speaking and writing.

Exam preparation can also have positive effects. When tests assess all four skills, they may encourage productive-language development—provided classroom activities reflect the skills being assessed [10].

Private education outside school is also directly relevant to this issue. M. Bray describes private tutoring as a “shadow education system” that follows the demands of formal education but may increase social inequality and change how students perceive school lessons [15]. Where high-stakes examinations are common, shadow education typically expands, and students may begin to regard school lessons as secondary.

Formative assessment and teachers’ assessment literacy

Formative assessment helps learners understand their current performance, goals, and next steps. Regular feedback can support progress more effectively than a final score alone.

Its impact depends on teachers’ assessment literacy: the ability to design tasks and rubrics, interpret results, and provide useful feedback. Limited training and time can make this difficult, especially in large classes.

The Uzbek context and the research gap

In Uzbekistan, CEFR standards and national certification have strengthened the focus on English proficiency. However, the relationship between school curricula, examination tasks, and everyday classroom practice remains under-researched, especially for speaking and writing in Grades 9 and 10.

RESEARCH METHODOLOGY

Research design

The study used an explanatory sequential mixed-methods design [16]. First, students completed a questionnaire and a CEFR-format mock examination. Follow-up interviews with students and teachers were then used to interpret the quantitative findings.

Participants

Participants were Grade 9 and 10 students from [4] general secondary schools in Fergana region ([3] public and [1] private). The sample included N = [186] students recruited through convenience sampling.

Table 1. Characteristics of the study participants

Measure	Grade 9	Grade 10	Total
Number of students, n	[94]	[92]	[186]
Girls, n (%)	[51 (54.3)]	[49 (53.3)]	[100 (53.8)]
Boys, n (%)	[43 (45.7)]	[43 (46.7)]	[86 (46.2)]
Age, M (SD)	[15.1 (0.4)]	[16.2 (0.5)]	[15.6 (0.7)]
Attends tutoring or a learning centre, n (%)	[38 (40.4)]	[47 (51.1)]	[85 (45.7)]
Previously took a certification	[3 (3.2)]	[11 (12.0)]	[14 (7.5)]

Measure	Grade 9	Grade 10	Total
examination, n (%)			

The study followed the principles of voluntary participation and confidentiality. Written parental consent was obtained for underage participants; students were informed about the study’s purpose and their right to withdraw at any time. All data were anonymised: students and teachers are identified in the article using codes (e.g., S1, T2).

Research instruments

The mock examination assessed listening ([30] items), reading ([30] items), writing (two tasks: an email and an essay), and speaking (three tasks). Tasks were designed to reflect CEFR-style formats.

To make the skill scores comparable, each section score was converted to a 0–100 scale by dividing the raw score by the maximum possible score and multiplying by 100. The overall score was calculated as the arithmetic mean of the four section scores. For the purposes of this study, overall scores were mapped onto approximate CEFR levels using the following ranges: 0–29, A1 or below; 30–49, A2; 50–69, B1; 70–84, B2; and 85–100, C1. These ranges are not official cut scores for a certification examination; they were established only to indicate an approximate distribution within this study.

A 28-item student questionnaire used five-point Likert scales and covered perceived difficulty by skill, classroom activities, independent preparation, motivation, and examination anxiety.

Semi-structured interviews were conducted with [8] teachers and [12] students. Students were selected to represent different mock-examination performance levels; interviews explored preparation practices, classroom experiences, and perceived challenges.

Data analysis

Quantitative data were analysed in SPSS 26 using descriptive statistics and Cronbach’s alpha. Skill scores were compared using repeated-measures ANOVA, and Grade 9 and 10 results using independent-samples t-tests. Pearson correlations examined associations between preparation, anxiety, and performance.

To answer the first research question, scores across the four skills were compared using repeated-measures ANOVA. Sphericity was assessed using Mauchly’s test, and the Greenhouse–Geisser correction was applied where the assumption was violated. Pairwise comparisons were conducted with the Bonferroni correction. Effect size was expressed as partial eta squared (ηp^2). Differences between Grades 9 and 10 were examined using an independent-samples t-test, and effect size was calculated using Cohen’s d: $d = (M_1 - M_2) / SDpooled$, where $SDpooled = \sqrt{[((n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2) / (n_1 + n_2 - 2)]}$. Following Cohen’s recommendations, d = 0.2, 0.5, and 0.8 were interpreted as small, medium, and large effects, respectively [18]. The association between independent preparation and examination

performance was assessed using Pearson’s correlation coefficient (r). Statistical significance was set at $p < 0.05$ for all tests.

Interview data were analysed thematically following Braun and Clarke’s six-phase approach [19]. The quantitative and qualitative findings were then compared to identify convergent patterns and explain differences.

RESULTS

Reliability of the questionnaire subscales

Table 2 presents Cronbach’s alpha coefficients for the questionnaire subscales. Values ranged from [0.74] to [0.81], exceeding the accepted threshold of 0.70. The questionnaire therefore demonstrated satisfactory internal consistency and was considered suitable for subsequent analyses.

Table 2. Reliability of the questionnaire subscales

Subscale	Number of items	Cronbach’s α
Perceived difficulty by skill	8	[0.81]
Frequency of classroom activity types	8	[0.76]
Independent preparation	6	[0.79]
Examination anxiety and motivation	6	[0.74]

Mock-examination results

Table 3 presents students’ results for each section of the mock examination, broken down by grade.

Table 3. Mock-examination results (0–100 scale)

Skill	Grade 9 M (SD)	Grade 10 M (SD)	Total M (SD)	Min–Max
Listening	[54.2 (15.8)]	[59.6 (16.3)]	[56.9 (16.2)]	[17–93]
Reading	[56.8 (16.9)]	[62.1 (17.2)]	[59.4 (17.2)]	[13–97]
Writing	[38.5 (17.6)]	[41.3 (18.9)]	[39.9 (18.3)]	[5–88]
Speaking	[41.7 (19.2)]	[44.0 (20.1)]	[42.8 (19.6)]	[8–91]
Overall score	[47.8 (13.2)]	[51.8 (13.9)]	[49.8 (13.6)]	[15–88]

Students scored highest in reading (M = [59.4], SD = [17.2]) and lowest in writing (M = [39.9], SD = [18.3]). Productive-skill performance (M = [41.4]) was notably below receptive-skill performance.

Bonferroni-adjusted pairwise comparisons showed no significant difference between listening and reading ($p = [0.071]$), while scores for both receptive skills were significantly higher than those for writing and speaking ($p < [0.001]$ in all cases). The difference between writing and speaking was small but statistically significant ($p = [0.032]$). Thus, Hypothesis 1

(H1) was supported. It should also be noted that the standard deviations for productive skills were relatively large, indicating substantial variation among students in these skills.

Distribution by CEFR level

The distribution of students across approximate CEFR levels, based on their overall scores, is presented in Table 4.

Table 4. Distribution of students by approximate CEFR level, n (%)

Level	Grade 9	Grade 10	Total
A1 or below	[21 (22.3)]	[14 (15.2)]	[35 (18.8)]
A2	[37 (39.4)]	[33 (35.9)]	[70 (37.6)]
B1	[28 (29.8)]	[32 (34.8)]	[60 (32.3)]
B2	[7 (7.4)]	[11 (12.0)]	[18 (9.7)]
C1	[1 (1.1)]	[2 (2.2)]	[3 (1.6)]

The largest group ([37.6]%) was at A2, while [32.3]% were at B1 and [11.3]% reached B2 or above. Some Grade 9 students ([22.3]%) had not reached the expected A2 level.

Differences between grades

Table 5. Comparison of Grade 9 and Grade 10 students’ results

Skill	t	df	p	Cohen’s d
Listening	[2.29]	184	[0.023]	[0.34]
Reading	[2.12]	184	[0.035]	[0.31]
Writing	[1.05]	184	[0.297]	[0.15]
Speaking	[0.80]	184	[0.426]	[0.12]
Overall score	[1.99]	184	[0.048]	[0.29]

The independent-samples t-test showed that Grade 10 students’ overall score was significantly higher than that of Grade 9 students: t(184) = [1.99], p = [0.048], d = [0.29]. However, the effect size was small, and the difference was mainly attributable to receptive skills; differences between grades in writing and speaking were small or non-significant. Thus, Hypothesis 3 (H3) was partially supported: overall performance increased over one school year, but growth in productive skills was not substantial.

Questionnaire results

Table 6. Descriptive statistics for questionnaire subscales (1–5 scale)

Subscale	M	SD
Perceived difficulty: listening	[2.94]	[0.92]
Perceived difficulty: reading	[2.71]	[0.88]
Perceived difficulty: writing	[3.68]	[0.95]
Perceived difficulty: speaking	[3.87]	[1.01]

Subscale	M	SD
Grammar and vocabulary exercises in class	[4.12]	[0.71]
Speaking activities in class	[2.46]	[0.98]
Writing tasks and feedback in class	[2.08]	[0.91]
Independent preparation	[2.89]	[0.84]
Examination anxiety	[3.71]	[0.87]
Motivation	[3.94]	[0.76]

Students rated speaking (M = [3.87], SD = [1.01]) and writing (M = [3.68], SD = [0.95]) as the most difficult skills. Grammar and vocabulary exercises were frequent (M = [4.12]), while writing tasks and detailed feedback were less common (M = [2.08]); only [17.2]% often received detailed essay comments.

Examination anxiety was relatively high (M = [3.71]); [68.3]% feared that not obtaining a certificate would put their future at risk. Motivation was also high (M = [3.94]) but appeared largely linked to certification and examinations.

Association between independent preparation and performance

Table 7. Pearson correlations between independent preparation, anxiety, and mock-examination results

Variable	1	2	3	4
1. Independent preparation	—			
2. Examination anxiety	[−0.14]	—		
3. Productive-skills score	[0.41**]	[−0.22**]	—	
4. Overall score	[0.36**]	[−0.18*]	[0.91**]	—

Note: * p < 0.05; ** p < 0.01.

Correlation analysis indicated a moderate positive association between independent preparation and overall score (r = [0.36], p < [0.001]). The association was somewhat stronger for productive-skill scores (r = [0.41], p < [0.001]). Examination anxiety had a weak negative association with overall score (r = [−0.18], p = [0.014]). Thus, Hypothesis 2 (H2) was supported. Because correlation does not establish causation, this result should be interpreted cautiously.

Qualitative findings

Interview analysis identified five recurring challenges: curriculum–exam mismatch, limited time and large classes, insufficient feedback on productive skills, reliance on private tutoring, and examination anxiety with extrinsic motivation.

Teachers reported that textbook topics and task types did not fully match the examination. Textbooks often emphasised vocabulary and grammar, while examinations

required reading, supported arguments, and oral discussion. One teacher noted that textbooks offered few complete essay models.

Limited lesson time and large classes restricted speaking practice. One teacher estimated that in a 45-minute lesson each student spoke for only one or two minutes; in classes of over 30, participation often came from a few active students.

Teachers assigned writing infrequently because marking was time-consuming, and feedback often focused on grammar rather than criteria. Students said they did not understand their scores or how to improve.

Students who attended private learning centres relied on them for exam preparation and viewed school lessons as less central. Teachers also found mixed proficiency levels challenging to address in one class.

Many students were motivated mainly by certification. Those with little experience of extended classroom speaking reported particular anxiety about the speaking test.

DISCUSSION

Writing and speaking were substantially weaker than listening and reading. This reflects a gap between CEFR-based assessment and classroom practice: lessons often prioritised grammar and vocabulary, while large classes and limited feedback reduced opportunities to practise productive skills [6; 7].

The results support research showing that washback affects learners’ and parents’ priorities more quickly than teaching processes [2; 9]. In this study, examination pressure was strong, but its influence on classroom methods appeared limited. Private learning centres had partly taken over preparation, consistent with Bray’s account of shadow education [15].

Students described private preparation as model-task practice and test-taking strategy training. Such preparation may improve familiarity with exam formats but cannot replace broader language development [3]. The positive association between independent preparation and performance suggests that meaningful writing and speaking practice is particularly important.

The small difference between grades, especially in productive skills, suggests that one additional school year does not necessarily lead to systematic progress in speaking and writing. Teachers recognised their importance but cited time, class size, and assessment literacy as barriers [12].

High motivation was largely extrinsic, and limited speaking practice appeared to intensify anxiety. Regular pair and small-group speaking can build competence and familiarity with the test while creating a safer environment for making mistakes [11].

Reliance on fee-paying learning centres also raises equity concerns. Students receiving tutoring scored higher (M = [53.2]) than those without tutoring (M = [46.9]), t(184) = [3.21], p = [0.002]. Schools should therefore provide more consistent access to effective preparation, regardless of family income.

Because CEFR-based examinations assess all four skills, they offer an opportunity for positive washback. Preparation should be integrated into communicative teaching rather than delivered as separate test-focused lessons [10].

A four-stage approach is proposed: (1) diagnose students’ skill levels; (2) build communicative practice through regular pair/group speaking and short writing tasks; (3) integrate CEFR-style tasks into subject lessons; and (4) use rubrics, peer assessment, and targeted feedback throughout the school year.

The approach uses existing lesson time. Pair and group work expands speaking opportunities, while peer assessment and simple rubrics make feedback more manageable and help students understand assessment criteria.

The findings should be interpreted cautiously. The sample was a convenience sample from one region; the mock test was not an official certification examination; questionnaire responses may reflect self-report bias; and the cross-sectional design cannot establish causality. The proposed model requires experimental evaluation.

CONCLUSION AND RECOMMENDATIONS

This study found that Grade 9 and 10 students performed less well in writing and speaking than in receptive skills. Curriculum–exam mismatch, limited classroom practice, and reliance on private tutoring appear to constrain preparation, while independent practice was positively associated with performance.

Teachers should include productive-skill practice in every lesson, use pair/group work, and assess writing with clear CEFR-informed rubrics. Schools should use diagnostic and end-of-year assessments to guide support. Textbooks and teacher-development programmes should better address examination tasks and assessment literacy.

Future experimental and longitudinal studies should test the proposed four-stage approach and examine how school-based preparation can improve language outcomes over time.

REFERENCES

1. Alderson J. C., Wall D. Does washback exist? // Applied Linguistics. – 1993. – Vol. 14, No. 2. – P. 115–129.

2. Cheng L. Changing language teaching through language testing: A washback study. – Cambridge: Cambridge University Press, 2005.

3. Green A. IELTS washback in context: Preparation for academic writing in higher education. – Cambridge: Cambridge University Press, 2007.

4. Council of Europe. Common European Framework of Reference for Languages: Learning, teaching, assessment. – Cambridge: Cambridge University Press, 2001.

5. Council of Europe. Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume. – Strasbourg: Council of Europe Publishing, 2020.

6. Little D. The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy // The Modern Language Journal. – 2007. – Vol. 91, No. 4. – P. 645–655.

7. Figueras N. The impact of the CEFR // ELT Journal. – 2012. – Vol. 66, No. 4. – P. 477–485.

8. Messick S. Validity and washback in language testing // Language Testing. – 1996. – Vol. 13, No. 3. – P. 241–256.

9. Bailey K. M. Working for washback: A review of the washback concept in language testing // Language Testing. – 1996. – Vol. 13, No. 3. – P. 257–279.

10. Hughes A. Testing for language teachers. 2nd ed. – Cambridge: Cambridge University Press, 2003.

11. Shohamy E., Donitsa-Schmidt S., Ferman I. Test impact revisited: Washback effect over time // Language Testing. – 1996. – Vol. 13, No. 3. – P. 298–317.

12. Spratt M. Washback and the classroom: The implications for teaching and learning of studies of washback from exams // Language Teaching Research. – 2005. – Vol. 9, No. 1. – P. 5–29.

13. Tsagari D. The complexity of test washback: An empirical study. – Frankfurt am Main: Peter Lang, 2009.

14. Brown H. D., Abeywickrama P. Language assessment: Principles and classroom practices. 2nd ed. – White Plains, NY: Pearson Longman, 2010.

15. Bray M. The shadow education system: Private tutoring and its implications for planners. – Paris: UNESCO International Institute for Educational Planning, 1999.

16. Creswell J. W., Plano Clark V. L. Designing and conducting mixed methods research. 3rd ed. – Thousand Oaks, CA: SAGE, 2018.

17. Dörnyei Z. Research methods in applied linguistics. – Oxford: Oxford University Press, 2007.

18. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. – Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

19. Braun V., Clarke V. Using thematic analysis in psychology // Qualitative Research in Psychology. – 2006. – Vol. 3, No. 2. – P. 77–101.

20. Resolution No. PQ-1875 of the President of the Republic of Uzbekistan, 10 December 2012, “On Measures to Further Improve the System of Learning Foreign Languages.”

21. Resolution No. 124 of the Cabinet of Ministers of the Republic of Uzbekistan, 8 May 2013, “On Approval of State Educational Standards for Foreign Languages in the Continuing Education System.”

22. Resolution No. PQ-4312 of the President of the Republic of Uzbekistan, 19 May 2019, “On Measures to Bring the Promotion of Foreign-Language Learning to a Qualitatively New Stage.”