



Akademik Eğitim ve Sosyal Bilimler Dergisi (AJESS Journal)

ÖĞRETİM YÖNTEMLERİNİN ÖTESİNDE: BANGLADEŞ'TEKİ ORTAOKUL İNGİLİZCE SINIFLARINDA ÖĞRENCİ BİREYSELLİĞİ VE ÖĞRENME HIZI KARMA YÖNTEM ÇALIŞMASI

BEYOND TEACHING METHODS: LEARNER INDIVIDUALITY AND LEARNING PACE IN SECONDARY ENGLISH CLASSROOMS IN BANGLADESH MIXED- METHOD STUDY

Md. Abu BAKKAR , Teacher., Quepara J. C. Sen A & I Institution, abubakrcrtg@gmail.com

Makale Tarihçesi/Article History

Geliş/Received: 02/12/2026 Düzeltme/Revision: 14/06/2026 Kabul/Accepted: 14/08/2026

Abstract

Teaching methods have long been treated as the primary determinant of learning outcomes in English language teaching (ELT). This study examines how learner individuality and learning pace relate to English language achievement and task-based performance among secondary students in Bangladesh, a context in which classroom instruction typically blends communicative and grammar-translation elements rather than implementing either in isolation. The study adopted a convergent mixed-methods design involving 120 students from Grades 8-10 (40 per grade; class sizes ranging from 35 to 45) in one secondary school. Quantitative data were collected through a general pre-test and post-test of English achievement and through a speaking task and a translation task completed by all participating students. Qualitative data were gathered through classroom observations, reflective journals, and semi-structured interviews. Paired-samples *t*-tests showed statistically significant improvement in post-test scores across all three grades (all $p < .001$); the associated effect sizes (Cohen's *d*) were unusually large because individual gain scores varied little around a consistent mean improvement of approximately 19 points, a pattern that is explained in full, together with the underlying formulas, software, and robustness checks. Across all grades, translation-task scores were consistently higher than speaking-task scores by an equally consistent margin; because these two tasks assess different language constructs (accuracy-focused written translation versus spontaneous oral communicative production) using different scoring criteria, and because class-level instructional-method membership was not

Öz

İngilizce dil öğretiminde (ELT) öğretim yöntemleri uzun zamandır öğrenme çıktılarının temel belirleyicisi olarak ele alınmaktadır. Bu çalışma, öğretimin genellikle iletişimsel ve dilbilgisi-çeviri unsurlarını birbirinden bağımsız olarak uygulamak yerine harmanladığı Bangladesh bağlamında, öğrenen bireyselliği ve öğrenme hızının ortaöğretim öğrencilerinin İngilizce dil başarıları ve görev temelli performansıyla nasıl ilişkili olduğunu incelemektedir. Araştırmada, tek bir devlet tarafından düzenlenen ortaöğretim okulunda 8-10. sınıflarda öğrenim gören 120 öğrenciyi (her sınıftan 40 öğrenci; sınıf mevcudu 35 ile 45 arasında) kapsayan yakınsayan karma yöntemli bir araştırma deseni benimsenmiştir. Nicel veriler, İngilizce başarılarını ölçmeye yönelik genel bir ön test ve son test ile araştırmaya katılan tüm öğrenciler tarafından gerçekleştirilen bir konuşma görevi ve bir çeviri görevi aracılığıyla toplanmıştır. Nitel veriler ise sınıf gözlemleri, yansıtıcı günlükler ve yarı yapılandırılmış görüşmeler yoluyla elde edilmiştir. Eşleştirilmiş örneklemeler için yapılan *t* testleri, üç sınıf düzeyinin tamamında son test puanlarında istatistiksel olarak anlamlı bir gelişme olduğunu göstermiştir (tüm $p < .001$). İlgili etki büyüklükleri (Cohen'in *d*) olağandışı derecede büyük bulunmuştur; bunun nedeni, bireysel kazanım puanlarının yaklaşık 19 puanlık tutarlı bir ortalama artış etrafında çok az değişkenlik göstermesidir. Bu örüntü, kullanılan formüller, yazılım ve sağlamlık kontrolleriyle birlikte ayrıntılı olarak açıklanmıştır. Tüm sınıf düzeylerinde çeviri görevi puanları, konuşma görevi puanlarından tutarlı biçimde daha yüksek bulunmuş ve bu fark da benzer ölçüde tutarlı bir düzeyde gerçekleşmiştir. Bununla birlikte, bu iki görev farklı dil yapılarını (doğruluğa odaklanan yazılı çeviri ile spontane sözlü iletişimsel üretim) farklı puanlama ölçütleri kullanarak değerlendirdiğinden ve sınıf düzeyindeki öğretim yöntemi

retained as an analytic variable, this pattern is interpreted as a difference between task types rather than as evidence that one instructional method is more effective than another. Qualitative analysis identified four interrelated themes: learner motivation, confidence, and willingness to participate; learning pace and classroom constraints; the influence of examination orientation on learning behaviour; and the differential benefits that communicative and grammar-focused activities offered to different learners. Together, the findings suggest that instructional activities function as supportive tools whose relationship with learning outcomes is shaped by learner individuality, learning pace, and assessment orientation, rather than as an overriding determinant of outcomes in itself. Implications include a need for post-method, principled-eclectic pedagogy, differentiated instruction, more balanced assessment, and teacher training to support diverse learners in examination-oriented contexts such as Bangladesh. Given the single-school context, the absence of a class-level instructional-group comparison, and the short intervention period, the findings should be interpreted as context-sensitive and descriptive rather than as a causal or generalisable comparison of teaching-method effectiveness.

Keywords: Teaching methods, learner individuality, learning pace, mixed-method, post-method pedagogy, secondary education, Bangladesh, CLT, GTM

gruplandırmasına ilişkin değişken analitik değişken olarak korunmadığından, bu örüntü bir öğretim yönteminin diğerinden daha etkili olduğuna ilişkin bir kanıt olarak değil, görev türleri arasındaki bir farklılık olarak yorumlanmıştır. Nitel analiz, birbiriyle ilişkili dört tema ortaya koymuştur: öğrenen motivasyonu, özgüven ve katılım isteği; öğrenme hızı ve sınıf içi kısıtlamalar; sınav odaklılığının öğrenme davranışları üzerindeki etkisi; ve iletişimsel ve dilbilgisi odaklı etkinliklerin farklı öğrenenlere sağladığı farklı yararlar. Bulgular bir bütün olarak değerlendirildiğinde, öğretim etkinliklerinin kendi başına öğrenme çıktılarının baskın belirleyicisi olmaktan ziyade, öğrenen bireyselliği, öğrenme hızı ve değerlendirme yönelimi tarafından şekillenen öğrenme çıktılarıyla ilişkili destekleyici araçlar olarak işlev gördüğünü göstermektedir. Çalışmanın sonuçları, Bangladeş gibi sınav odaklı bağlamlarda farklı öğrenenleri desteklemek amacıyla yöntem-sonrası ve ilkeli-eklektik pedagojik yaklaşıma, farklılaştırılmış öğretime, daha dengeli bir değerlendirme anlayışına ve öğretmen eğitime duyulan ihtiyaca işaret etmektedir. Araştırmanın tek okul bağlamıyla sınırlı olması, sınıf düzeyinde öğretim grupları arasında bir karşılaştırmanın bulunmaması ve müdahale süresinin kısa olması nedeniyle, bulgular öğretim yöntemlerinin etkililiğine ilişkin nedensel veya genellenebilir bir karşılaştırma olarak değil, bağlama duyarlı ve betimleyici sonuçlar olarak yorumlanmalıdır.

Anahtar Kelimeler: Öğretim yöntemleri, öğrenci bireyselliği, öğrenme hızı, karma yöntem, yöntem sonrası pedagoji, ortaöğretim, Bangladeş, CLT, GTM

Atıf için: Bakkar, M. A. (2026) Beyond teaching methods: Learner individuality and learning pace in secondary English classrooms in Bangladesh mixed-method study, *Akademik Eğitim ve Sosyal Bilimler Dergisi (AJESS Journal)*, 3(3), 13-31.

Beyond Teaching Methods: Learner Individuality and Learning Pace in Secondary English Classrooms in Bangladesh Mixed-Method Study

Introduction

The field of English language teaching (ELT) has been dominated for decades by the search for the "best" method. From grammar-translation to the direct method, audiolingualism, and later Communicative Language Teaching (CLT), teachers and researchers have often treated method as the central factor determining success in language learning. However, the complexity of classroom reality suggests that teaching methods are not the sole drivers of learning outcomes. Learners differ in cognitive styles, motivation, prior knowledge, socio-economic background, and learning pace, all of which shape how they respond to classroom activities. This study explores the proposition that teaching methods are supplementary rather than decisive: they support learning but cannot override learner individuality and learning pace.

In this study, *learner individuality* refers to observable and reported differences among learners in motivation, anxiety, prior knowledge, confidence, and engagement, as evidenced through task performance, reflective journals, classroom observations, and interviews. *Learning pace* is understood as the relative speed at which learners comprehend, practise, and produce language tasks, operationalised through classroom observation of task completion time, need for scaffolding, and differential progress between the pre-test, post-test, and task-based assessments.

In Bangladesh, English is taught as a compulsory subject from primary through secondary education and is widely considered a key instrument for social mobility. Despite an official curricular preference for communicative approaches, classroom practice often reflects a mixture of methods, shaped by examination pressure, large class sizes, and limited resources. Research in Bangladesh indicates that even where CLT is officially promoted, the Grammar Translation Method (GTM) remains prevalent in practice, sometimes in disguised forms, and that high-stakes examinations exert strong negative washback on classroom instruction, narrowing it toward grammar, reading, and writing at the expense of speaking and listening (Abedin, 2013; Ali & Hamid, 2020; Matin, 2021; Rahman et al., 2019, 2021). This mixed-method, examination-oriented environment provides a context in which learner factors and task demands may matter as much as, or more than, instructional method alone.

The present study is grounded in a post-method pedagogy perspective, which argues that teaching must move beyond rigid method adherence and instead foreground local context, learner needs, and teacher judgement (Kumaravadivelu, 2001, 2006). This perspective is consistent with complex dynamic systems theory in second language acquisition, which highlights the dynamic interaction of multiple factors, including individual differences (de Bot,

Lowie, & Verspoor, 2007). Recent research on individual differences similarly emphasises the socio-ecological and holistic nature of learner variation, suggesting that motivation, anxiety, identity, and learning pace interact in complex ways rather than operating as isolated, stable traits (Spring, Pawlak, & Li, 2021).

This study examines English teaching in secondary schools in Bangladesh, focusing on Grades 8-10 with class sizes of 35-45 students. The research questions are:

1. To what extent did students' English achievement change over the course of the instructional intervention, and how did performance differ between communicative (speaking) and accuracy-oriented (translation) language tasks?
2. How are learner individuality and learning pace associated with students' engagement and performance across these different task types?
3. What are the implications of these patterns for teaching practice, assessment, and teacher training in Bangladeshi secondary English education?

Using a mixed-methods approach, this study aims to produce empirical evidence on the interaction between instructional context, task type, and learner factors. It contributes to ELT scholarship by providing data from a South Asian context in which large classes, examination-driven instruction, and resource constraints shape teaching and learning, and it contributes to debates on post-method pedagogy by illustrating how learner diversity can be foregrounded in the interpretation of classroom assessment data.

Literature Review

Evolution of Teaching Methods in ELT

The history of ELT methods reflects a continuous search for a universal approach to language teaching. Early methods such as the Grammar Translation Method emphasised explicit grammar instruction and translation exercises, prioritising accuracy and written competence. The direct method shifted attention to oral communication, while the audiolingual method focused on repetition and habit formation. The rise of CLT in the late twentieth century marked a significant shift toward meaning-based communication, learner interaction, and real-life language use, and CLT is often considered the dominant paradigm in contemporary ELT, emphasising fluency, task-based interaction, and learner-centred instruction (Richards & Rodgers, 2014).

The "method wars," however, have been criticised for oversimplifying teaching and ignoring the complexity of classrooms. Kumaravadivelu (2001, 2006) argued that methods are often inadequate because they cannot account for the particularity of local conditions and learner diversity, proposing instead a postmethod pedagogy built on the parameters of particularity, practicality, and possibility. This perspective aligns with broader critiques suggesting that teaching methods should be viewed as tools rather than solutions (Brown, 2014; Kumaravadivelu, 2001, 2006).

In Bangladesh specifically, the official curriculum promotes communicative competence, but classroom realities often involve a hybrid of methods. Abedin (2013) found that ELT classes at the higher secondary level often practise a "disguised" GTM despite CLT rhetoric, and subsequent washback research has reinforced this picture: Rahman et al. (2019) documented that teachers in Bangladesh continue to hold ambivalent attitudes toward CLT more than two decades after its official adoption, while Ali and Hamid (2020) and Rahman, Seraj, Hasan, Namaziandost, and Tilwani (2021) showed that the grammar- and accuracy-weighted format of the Secondary School Certificate (SSC) English examinations exerts strong negative washback, encouraging teachers to narrow instruction toward the skills that are tested (grammar, reading, and writing) and to de-emphasise speaking and listening even where these are nominally part of the curriculum. Matin (2021) similarly argued that a blended approach incorporating both GTM and CLT can be effective in this context, especially where examination performance is a central concern for students and teachers alike. Together, these findings suggest that method alone cannot explain learning outcomes in the Bangladeshi context; rather, examination orientation, institutional constraints, and learner readiness shape how any given instructional activity is taken up in the classroom.

CLT and GTM: Strengths and Limitations

CLT emphasises interaction, communicative competence, and learner engagement, aiming to develop learners' ability to use language in real-life situations and to prioritise meaning over form. Research suggests that CLT can enhance speaking confidence, fluency, and communicative awareness (Richards & Rodgers, 2014), but it also requires sufficient classroom time, teacher competence, and learner willingness to participate; in contexts with large classes, limited resources, and examination-driven teaching, CLT can face practical constraints (Ali & Hamid, 2020).

GTM, by contrast, emphasises accuracy, grammar, and translation skill, and can be effective in developing reading comprehension and grammatical awareness, particularly in examination-oriented environments. Núñez-Vázquez and Crismán-Pérez (2021) found that GTM can support reading comprehension and language awareness in ways that CLT does not fully address, and, in a quasi-experimental comparison with ninth-grade secondary students,

Çiftçi and Özcan (2021) similarly reported that a traditional grammar-translation approach produced fewer grammatical errors and stronger vocabulary production than CLT on written and error-analysis measures, while CLT-taught students showed corresponding strengths in extended oral production. GTM is nonetheless routinely criticised for neglecting speaking skills and communicative competence and for encouraging rote learning over meaningful communication (Richards & Rodgers, 2014).

This body of work illustrates that the comparative literature on CLT and GTM is genuinely mixed, with different studies favouring one approach or the other depending on the outcome measured (grammatical accuracy versus communicative production) and the research design used. The present study does not attempt to adjudicate this broader debate; as explained in Section 3.1, its design does not support a controlled comparison between CLT-taught and GTM-taught groups. Its contribution instead lies in describing, within a single sample, how performance differs across two task types that mirror the constructs these two traditions emphasise, and in exploring, through qualitative data, why learners might respond differently to each.

The debate between CLT and GTM is, in any case, not merely theoretical; it reflects classroom realities in which teachers routinely blend methods. Maqbool (2024) reported that teachers in rural Bangladesh adapt methods according to context, often preferring GTM because of examination pressure and limited resources, suggesting that method choice is shaped by external constraints as much as by pedagogical belief.

Learner Individual Differences and Learning Pace

Learner differences have long been recognised in second language acquisition research and include cognitive style, motivation, aptitude, anxiety, personality, prior knowledge, and learning strategies, all of which shape how learners perceive, process, and retain language input. The complex dynamic systems perspective emphasises that individual differences are not stable traits but dynamic variables interacting with context and experience (de Bot et al., 2007), while Spring, Pawlak, and Li (2021) proposed a socio-ecological and holistic view of individual differences, arguing that learners' backgrounds, identities, and social contexts jointly shape their language-learning trajectories. In Bangladesh, socio-economic factors, exposure to English outside school, parental support, and cultural attitudes toward English can all influence learning outcomes; students from urban backgrounds, for example, may have more exposure to English through media and private tutoring than students from rural backgrounds with more limited resources.

Learning pace is a related but distinct construct: some learners grasp new material quickly, while others require repetition, scaffolding, and additional time. In large classes, teachers often cannot accommodate this diversity of pace, which can produce frustration for slower learners and disengagement for faster ones. Tomlinson's (2001) work on differentiated instruction offers a pedagogical framework for responding to exactly this kind of variation, arguing that teachers can adapt content, process, and product according to learners' readiness, interest, and learning profile rather than pacing instruction to a single hypothetical average learner. In large, mixed-proficiency classes of the kind common in Bangladeshi secondary schools, such differentiation is difficult to implement in practice, but it may be particularly important for reconciling the needs of faster- and slower-progressing learners (Brown, 2014; Nation, 2013; Tomlinson, 2001).

Language Anxiety, Motivation, and Willingness to Communicate

Among learner-internal factors, language anxiety and willingness to communicate (WTC) are particularly relevant to understanding why students in this study responded differently to communicative activities. Horwitz (2001) describes foreign language anxiety as a largely situation-specific form of anxiety that can impede performance independently of a learner's underlying ability, and MacIntyre (1999) outlines how apprehension about evaluation and fear of negative judgement can suppress spontaneous language production even when linguistic competence is otherwise adequate. MacIntyre, Clément, Dörnyei, and Noels (1998) proposed an influential pyramid model conceptualising WTC as the outcome of layered situational and enduring influences, including self-confidence, intergroup attitudes, and classroom climate, suggesting that a learner's readiness to speak in a second language is shaped by considerably more than the instructional activity in front of them at a given moment. These constructs anticipate the qualitative findings reported in Section 4.4, in which willingness to participate in speaking tasks varied markedly among students with broadly similar linguistic competence.

Post-Method Pedagogy and Context-Sensitive Teaching

Post-method pedagogy emphasises teacher autonomy, contextual awareness, and principled eclecticism. Kumaravadivelu (2001) argued that teachers should develop their own teaching strategies based on particularity (sensitivity to a specific group of learners in a specific context), practicality (a workable relationship between theory and classroom practice), and possibility (attention to learners' social and political realities); Kumaravadivelu (2006) subsequently extended this framework into a fuller account of the shift from method-based to postmethod pedagogy. This approach is particularly relevant in Bangladesh, where classroom constraints and examination pressures require flexible teaching rather than fidelity to a single method.

Post-method pedagogy aligns with the idea that teaching methods are supplementary: they provide frameworks and techniques, but effective teaching requires adaptation, teacher judgement, and sensitivity to learners. In this perspective, CLT and GTM are not competing ideologies but tools that can be used selectively according to learners' needs and the skills being targeted.

Research Gap

Although research has examined CLT and GTM in Bangladesh, there is limited empirical work that jointly considers general achievement gains, task-based performance differences, and learner individuality and learning pace within a single mixed-methods study. Much of the existing literature focuses on teacher perceptions, general classroom observation, or between-group method-comparison designs conducted at a single grade or proficiency level (Çiftçi & Özcan, 2021; Núñez-Vázquez & Crismán-Pérez, 2021). This study contributes to that literature by combining general achievement data, task-based (speaking versus translation) performance data, and qualitative accounts of learner individuality and learning pace across three secondary grade levels within one examination-oriented school context.

Theoretical Framework

This study is grounded in post-method pedagogy and complex dynamic systems theory (CDST). Post-method pedagogy argues that teaching must move beyond rigid method adherence and focus on local context, learner needs, and teacher judgement (Kumaravadivelu, 2001, 2006). CDST emphasises the dynamic interaction of multiple factors in language learning, including individual differences, social context, and classroom environment (de Bot et al., 2007). Together, these frameworks motivate the study's central proposition: that classroom activities are supplementary tools whose relationship with learning outcomes is shaped by learner individuality and learning pace, rather than an overriding determinant of outcomes considered in isolation.

Methodology

Research Design

This study employed a convergent mixed-methods design (Creswell & Plano Clark, 2018), in which quantitative and qualitative data were collected concurrently and considered together to develop a fuller understanding of English language learning among secondary students in Bangladesh, including general achievement gains, task-based performance, and the role of learner individuality and learning pace.

The intervention was conducted over 12 weeks within existing, intact Grade 8-10 classes at one secondary school; students were not randomly assigned to instructional conditions. Reflecting normal classroom practice in this context, two of the four participating English teachers placed relatively greater emphasis on Communicative Language Teaching (CLT) in their general classroom instruction, while the other two placed relatively greater emphasis on the Grammar Translation Method (GTM); all four teachers followed the same curricular objectives, instructional schedule, and intended learning outcomes throughout the intervention, and periodic structured classroom observations were used to monitor instructional practice.

This instructional detail describes the general classroom context rather than a controlled between-group comparison. Class-level instructional emphasis was not retained as an analytic grouping variable in the dataset analysed for this study, and all 120 participating students, regardless of which teacher's class they were enrolled in, completed the same assessment battery: a general pre-test and post-test of English achievement, together with a speaking task and a translation task administered near the end of the intervention period (Section 3.3). Consequently, the quantitative analyses reported in Section 4 compare performance between task types (speaking versus translation) and between time points (pre- versus post-intervention) within the same students; they do not compare CLT-taught and GTM-taught groups against one another, and no claim of relative instructional-method effectiveness is made on the basis of the quantitative data. This reflects a genuine constraint of the dataset available for analysis: a valid between-method comparison would require class- or teacher-level identifiers to be retained for every student record and analysed using an appropriate design (for example, a time $\times$ method interaction analysis or ANCOVA controlling for pre-test score), which was not done here. A formal between-method comparison of this kind remains an important direction for future research and is discussed further in Section 8.

Because students were drawn from three intact classes taught by four teachers, scores from students in the same class cannot be assumed to be fully independent of one another; students within a class may resemble each other more than they resemble students in other classes simply by virtue of sharing a teacher, a classroom environment, and a peer group (Raudenbush & Bryk, 2002). The analyses reported in Section 4 (paired-samples t-tests conducted within each grade) treat individual students as the unit of analysis and do not model this classroom-level clustering explicitly. This is acknowledged as a limitation in Section 8, where multilevel modelling is discussed as a preferred approach for future datasets that retain class-level identifiers.

Implementation fidelity, understood as the extent to which the intended instructional approach was actually delivered in the classroom (Century, Rudnick, & Freeman, 2010), was monitored descriptively through periodic structured classroom observations rather than through a formal fidelity checklist scored for every lesson. Fidelity is therefore reported qualitatively in Section 4.4 rather than as a quantitative index, which is also acknowledged as a limitation in Section 8.

Participants and Sampling

The study involved 120 students from Grades 8-10 (40 students per grade) enrolled in one secondary school in Bangladesh.

The school was selected purposively because administrative permission, sustained classroom access, and teacher cooperation were available for conducting the intervention. Within the selected school, the study drew on three intact classes, identified in cooperation with the school administration (Section 3.1). Inclusion criteria were: (a) regular enrolment in one of the four participating classes for the full duration of the 12-week intervention; (b) provision of informed consent by a parent or legal guardian; and (c) provision of informed assent by the student. Students enrolled in a participating class who did not meet all three criteria (for example, because consent was not provided, or because a student was absent for a substantial portion of the intervention or of the pre-test/post-test sessions) were not included in the analysed sample.

Of the students enrolled across the three participating classes, none were excluded under the criteria above: all students met the inclusion criteria described above and are retained in full in the analysed sample. The 120 students retained in the analysed sample constitute a convenience sample within a purposively selected school and set of classes.

The analysed sample included both male and female students across the three grade levels. Gender was not systematically recorded as part of the study's participant data and is therefore not reported as a demographic variable; this is noted as an unmeasured background characteristic, alongside the socio-economic factors discussed in Section 8. Because class-level instructional emphasis was not retained as an analytic variable (Section 3.1), the distribution of students across the more CLT-oriented and more GTM-oriented classes is not reported separately.

Proficiency-level representation was addressed through the inclusion of three grade levels spanning early to late secondary English instruction, and through pre-test performance itself, which varied meaningfully across students within every grade. Pre-test scores ranged from 34 to 54 in Grade 8 ($M = 44.68$, $SD = 4.74$), from 37 to 57 in Grade 9 ($M = 47.40$, $SD = 5.01$), and from 36 to 55 in Grade 10 ($M = 45.28$, $SD = 4.66$), indicating that each grade-level sample included students entering the intervention with a substantial range of prior achievement rather than a narrow proficiency band. Because the study was conducted in one school, the findings should nonetheless be interpreted as context-specific rather than statistically generalisable to other schools or regions.

Data Collection Instruments

Pre-test and post-test: Students completed equivalent achievement tests before and after the intervention. The tests assessed reading comprehension, grammar, vocabulary, and writing, based on the National Curriculum and previous Secondary School Certificate examination formats. Content validity was established through curriculum alignment and review by two experienced secondary English teachers.

Speaking task: Students completed communicative activities, including role plays, pair discussions, oral presentations, and information-gap tasks, designed to elicit spontaneous oral production. Performance was evaluated using an analytic rubric assessing fluency, grammatical accuracy, pronunciation, interaction, and communicative effectiveness.

Translation task: Students completed English-Bengali and Bengali-English translation tasks requiring lexical and grammatical accuracy in written language transfer. Performance was evaluated using an analytic rubric assessing grammatical accuracy, lexical choice, preservation of meaning, and overall language accuracy.

Classroom observations: Structured classroom observations were conducted periodically throughout the 12-week intervention, focusing on instructional practices, learner participation, classroom interaction, learner engagement, and classroom management.

Reflective journals: Students maintained weekly reflective journals documenting their learning experiences, challenges, motivation, confidence, and preferred learning activities.

Semi-structured interviews: Semi-structured interviews with students and with all four participating teachers explored perceptions of communicative and grammar-translation activities, learner differences, learning pace,

classroom challenges, and instructional effectiveness. A total of 18 student interviews were conducted, in addition to interviews with all four participating teachers. Student interviewees were selected through purposive sample of students capturing grades 8-10 and proficiency levels. Student interviews typically lasted approximately 15-20 minutes, and teacher interviews typically lasted approximately 25-35 minutes.

Scoring Procedure

To enhance scoring consistency, two independent, experienced English teachers first scored a randomly selected 20% subset of speaking- and translation-task performances independently, using the analytic scoring rubrics described in Section 3.3. Before scoring began, the two raters were jointly briefed on the analytic rubrics described in Section 3.3 and discussed the scoring criteria for each rubric dimension; however, no formal calibration exercise was conducted. Where scoring differences occurred on this double-scored 20% subset, the two raters discussed the specific performance and reached a consensus score. The remaining 80% of performances were then divided between the two raters, with each rater independently scoring a separate, non-overlapping portion of the remaining subset.

A formal inter-rater reliability coefficient (for example, the intraclass correlation coefficient, ICC; Koo & Li, 2016) was not calculated for the double-scored subset. Consequently, no quantitative claim regarding the magnitude of inter-rater agreement is made, and this is acknowledged as a methodological limitation in Section 8.

Data Analysis

Quantitative analysis

Quantitative data were analysed using paired-samples t-tests, conducted separately for each grade, to compare (a) pre-test and post-test achievement scores and (b) speaking-task and translation-task scores. All analyses were conducted in Python (version 3.12) using the pandas, NumPy, and SciPy libraries; every value reported in Section 4 was calculated directly from the dataset described in Section 3.2.

Prior to inferential analysis, distributions of the difference scores were examined through visual inspection of histograms and Q-Q plots, which suggested approximately normal distributions. Formal Shapiro-Wilk tests, however, indicated statistically significant deviations from normality for the pre-test/post-test difference scores in all three grades (Grade 8: $W = .93$, $p = .012$; Grade 9: $W = .93$, $p = .020$; Grade 10: $W = .93$, $p = .011$). Given the sensitivity of this test at $n = 40$, such a result does not necessarily indicate a practically important departure from normality, and the paired t-test is known to be robust to modest non-normality at this sample size. As a robustness check, non-parametric Wilcoxon signed-rank tests were additionally conducted for every comparison; results were consistent with the paired t-tests in direction and significance in every case (all $p < .001$; Table 3). Outlier assessment using the $1.5 \times$ interquartile range (IQR) rule identified no outlying difference scores for the pre-test/post-test comparison in any grade. For the speaking-task/translation-task comparison, the difference scores were tightly clustered within an 11-to-14-point range in every grade, which limits the diagnostic value of a mechanical IQR-based rule for this particular comparison (the interquartile range itself approaches zero); visual inspection did not suggest that any individual score was anomalous relative to the rest of the distribution.

Effect sizes for these paired, within-subjects comparisons are reported as Cohen's d_z , the effect-size index appropriate for paired designs, calculated as the mean of the difference scores divided by the standard deviation of the difference scores: $d_z = M_{diff} / SD_{diff}$ (Lakens, 2013). This differs from the between-groups Cohen's d , which is calculated from the pooled standard deviation of two independent samples, and the two indices are not directly comparable. Where useful for interpretation against conventional effect-size benchmarks, the corresponding d_{av} (Lakens, 2013) is also reported; d_{av} standardises the mean difference by the average of the two conditions' individual standard deviations and does not adjust for the correlation between the paired scores. For each comparison, the following statistics are reported in accordance with APA (7th ed.) recommendations: the mean and standard deviation of each measure; the mean and standard deviation of the difference scores; the paired-samples t statistic and degrees of freedom; the exact p value, alongside the conventional " $p < .001$ " notation used in-text for very small values; Cohen's d_z ; and the 95% confidence interval for the mean difference.

Because students were nested within three intact classes rather than independently sampled (Section 3.1), the paired t-tests reported here treat individual students as the unit of analysis and do not model potential classroom-level clustering explicitly. Multilevel or hierarchical linear modelling (Raudenbush & Bryk, 2002) would be a more appropriate approach for data that retain class-level identifiers, allowing classroom-level variance to be estimated directly; this is noted as a direction for future analysis in Section 8.

Qualitative analysis

Qualitative data were analysed using inductive thematic analysis, following the approach described by Braun and Clarke (2006). One researcher conducted all stages of coding: familiarisation with the data through repeated close reading of interview transcripts, reflective journal entries, and classroom observation notes; generation of initial codes grounded in participants' own language and observed behaviour; and iterative comparison and consolidation of related codes into candidate subthemes and, ultimately, the four broader themes reported in Section 4.4.

The researcher's role during data collection was that of a non-participant observer during classroom observations and as interviewer for the semi-structured interviews, and the same researcher was the sole analyst responsible for coding and theme development. Because coding was conducted by one researcher, investigator triangulation in the strict sense was not possible; this is acknowledged as a limitation in Section 8. To strengthen the trustworthiness of the analysis, several credibility strategies recommended by Lincoln and Guba (1985) were employed, including data triangulation across interviews, journals, and observations; maintenance of an audit trail documenting coding decisions; reflexive memo-writing; repeated comparison between themes and the original data; and peer debriefing with experienced ELT colleagues regarding theme interpretation.

Peer debriefing involved discussing emerging codes, candidate themes, and alternative interpretations with experienced ELT colleagues who were not otherwise involved in data collection or analysis. Peer debriefing was conducted with two experienced ELT colleagues. Debriefing sessions took place after the initial coding stage and again during theme refinement, allowing colleagues to review both the emerging coding framework and the more developed thematic structure. Debriefing colleagues were asked to consider whether alternative readings of the data were plausible and whether the researcher's interpretations were adequately grounded in the data. Where a debriefing colleague raised an alternative interpretation, the researcher returned to the original data to evaluate whether the existing theme adequately accounted for it or whether the theme required revision; disconfirming or negative cases identified through this process are reported explicitly in Section 4.4 (Theme 4).

Ethical Considerations

Formal ethics-committee approval was not obtained for this study. In Bangladesh, mandatory national ethical review of research involving human participants is principally administered through the Bangladesh Medical Research Council's National Research Ethics Committee (NREC), operating under the National Health Research Strategy; this mechanism was established primarily to oversee medical, clinical, and health-related research, and independently conducted, non-medical educational research of the kind reported here does not fall under a comparable, unified mandatory review requirement. The study was conducted independently within a secondary school and was not undertaken under the auspices of a university or an institutional research programme that would have required committee review under its own policies.

Institutional permission to conduct the study was obtained from the school administration. The school's Head Teacher granted permission following a verbal request after a discussion of the study's purpose, importance, procedures, and duration. As the school did not have a formal research approval system, no written approval or reference number was issued.

Informed consent was obtained from the parents or legal guardians of the participating students, as well as from the participating teachers, in cooperation with the school administration. Participating students additionally provided informed assent after the study procedures had been explained in age-appropriate language. Participation was voluntary, and students were informed that they could withdraw from the study at any time without academic consequence. All responses were anonymised and used exclusively for research purposes.

No formal exemption, waiver, or acknowledgement letter regarding ethics review was sought from or issued by any external body. This is consistent with the absence of a mandatory ethics-review mechanism applicable to independently conducted, non-medical educational research of this kind, as explained above, and is acknowledged as a limitation of the study's ethical documentation.

Findings

Quantitative Findings

Pre-test and Post-test Achievement

Table 1: *Descriptive and Inferential Statistics for Pre-test and Post-test Scores, by Grade*

Grade	n	Pre-test (SD)	M	Post-test (SD)	M	Mdiff (SDdiff)	t(39), p	dz	95% CI
8	40	44.68 (4.74)		64.25 (5.67)		19.58 (1.82)	67.87, <.001	10.73	[18.99, 20.16]
9	40	47.40 (5.01)		66.88 (5.45)		19.48 (1.60)	76.93, <.001	12.16	[18.96, 19.99]

10	40	45.28 (4.66)	64.55 (5.12)	19.28 (1.68)	72.60, <.001	11.48	[18.74, 19.81]
----	----	--------------	--------------	--------------	--------------	-------	----------------

Note. *Mdiff* and *SDdiff* are the mean and standard deviation of the within-student post-test-minus-pre-test difference scores. $dz = Mdiff / SDdiff$ (Lakens, 2013); see Section 3.5 and Section 4.2 for the corresponding *dav* values and full robustness checks (Table 3). 95% CI is the confidence interval for the mean difference. All degrees of freedom = 39.

Across all three grade levels, post-test scores were substantially higher than pre-test scores. Grade 8 students improved from a pre-test mean of 44.68 ($SD = 4.74$) to a post-test mean of 64.25 ($SD = 5.67$), an average gain of 19.58 points. Grade 9 students' mean scores increased from 47.40 ($SD = 5.01$) to 66.88 ($SD = 5.45$), a gain of 19.48 points. Grade 10 students improved from 45.28 ($SD = 4.66$) to 64.55 ($SD = 5.12$), a gain of 19.28 points. These descriptive results indicate consistent and substantial improvement in English language achievement over the course of the instructional intervention, with very similar average gains across all three grade levels.

Paired-Samples *t*-Tests and Effect-Size Interpretation

Paired-samples *t*-tests demonstrated statistically significant improvement in achievement for all three grade levels. For Grade 8, post-test scores ($M = 64.25$, $SD = 5.67$) were significantly higher than pre-test scores ($M = 44.68$, $SD = 4.74$), $t(39) = 67.87$, $p < .001$ (exact $p = 4.21 \times 10^{-42}$), $Mdiff = 19.58$, $SDdiff = 1.82$, $dz = 10.73$, $dav = 3.75$, 95% CI for the mean difference [18.99, 20.16]. For Grade 9, students similarly improved from pre-test ($M = 47.40$, $SD = 5.01$) to post-test ($M = 66.88$, $SD = 5.45$), $t(39) = 76.93$, $p < .001$ (exact $p = 3.29 \times 10^{-44}$), $Mdiff = 19.48$, $SDdiff = 1.60$, $dz = 12.16$, $dav = 3.72$, 95% CI [18.96, 19.99]. For Grade 10, students improved from pre-test ($M = 45.28$, $SD = 4.66$) to post-test ($M = 64.55$, $SD = 5.12$), $t(39) = 72.60$, $p < .001$ (exact $p = 3.11 \times 10^{-43}$), $Mdiff = 19.28$, $SDdiff = 1.68$, $dz = 11.48$, $dav = 3.93$, 95% CI [18.74, 19.81].

These dz values are exceptionally large by conventional benchmarks (Cohen, 1988, treats $dz = 0.8$ as a "large" effect). This is not a computational error, and all values have been independently reproduced from the raw dataset for this revision (Section 3.5). Because pre-test and post-test scores were highly correlated within students in every grade ($r = .95$ to $.96$), the standard deviation of the difference scores was small relative to the mean gain ($SDdiff = 1.6$ to 1.8 points, against mean gains of 19.3 to 19.6 points), and dz is mathematically sensitive to this ratio. Substantively, this reflects a consistent improvement across nearly all students rather than a small number of students driving the average: individual gain scores ranged narrowly within each grade (for example, 16 to 23 points in Grade 8), and no outliers were identified using the $1.5 \times IQR$ rule (Section 3.5). Readers are cautioned against comparing these dz values directly to between-groups Cohen's d benchmarks; the more conventional dav values (3.72 to 3.93 across grades) remain very large but are more directly comparable to typical effect-size benchmarks and are reported for that purpose. Table 3 reports the corresponding Shapiro-Wilk and Wilcoxon signed-rank results in full; the non-parametric results corroborate the parametric findings in every comparison.

Overall, the paired-samples *t*-test results indicate that the instructional intervention was followed by statistically significant, and unusually consistent, improvement in students' English language performance across all three grade levels.

Speaking-Task and Translation-Task Performance

Table 2: Descriptive and Inferential Statistics for Speaking-Task and Translation-Task Scores, by Grade

Grade	n	Speaking M (SD)	Translation M (SD)	Mdiff (SDdiff)	t(39), p	dz	95% CI
8	40	56.68 (3.05)	69.63 (2.71)	12.95 (0.71)	114.66, <.001	18.13	[12.72, 13.18]
9	40	57.88 (2.71)	70.90 (2.62)	13.03 (0.42)	194.79, <.001	30.80	[12.89, 13.16]
10	40	59.68 (2.85)	72.70 (2.75)	13.03 (0.28)	298.26, <.001	47.16	[12.94, 13.11]

Note. Speaking and translation tasks are described in Section 3.3. *Mdiff* is Translation-minus-Speaking. As explained in Section 4.3, this comparison describes within-student differences between two task types and should not be read as a comparison of instructional-method effectiveness.

Across all three grade levels, mean scores on the translation task were consistently higher than mean scores on the speaking task, by a highly consistent margin of approximately 13 points. Grade 8 students obtained a mean speaking-task score of 56.68 ($SD = 3.05$) compared with a mean translation-task score of 69.63 ($SD = 2.71$), a difference of 12.95 points, $t(39) = 114.66$, $p < .001$, $dz = 18.13$. Grade 9 students showed a similar pattern (speaking $M = 57.88$, $SD = 2.71$; translation $M = 70.90$, $SD = 2.62$; difference = 13.03 points, $t(39) = 194.79$, $p < .001$, $dz = 30.80$), as did Grade 10 students (speaking $M = 59.68$, $SD = 2.85$; translation $M = 72.70$, $SD = 2.75$; difference = 13.03 points, $t(39) = 298.26$, $p < .001$, $dz = 47.16$).

As with the pre-test/post-test comparisons in Section 4.2, these dz values are extremely large because speaking-task and translation-task scores were even more highly correlated within students ($r = .98$ to $.996$ across grades) than pre-test and post-test scores were, producing a very small standard deviation of the difference scores (0.28 to 0.71

points) relative to the mean difference. Because the difference scores were tightly clustered within an 11-to-14-point range in every grade, the standard $1.5 \times \text{IQR}$ outlier rule is not informative for this comparison (Section 3.5); visual inspection did not suggest that any individual score was anomalous relative to the rest of the distribution.

These statistics describe a highly consistent descriptive pattern, in which nearly every student scored higher on the translation task than on the speaking task by a broadly similar margin, and they should not be interpreted as a test of instructional-method effectiveness. The speaking task and the translation task assess different language constructs (spontaneous oral communicative production versus written grammatical accuracy and meaning transfer) using different analytic rubrics (Section 3.3), and, as explained in Section 3.1, class-level instructional-method membership was not available as an analytic variable in this dataset. Accordingly, the difference reported here is best understood as a difference between task types within the same students, not as evidence that the Grammar Translation Method was more effective than Communicative Language Teaching, or vice versa. Sections 4.4 and 5 draw on the qualitative data to explore why this task-based pattern may have occurred.

Table 3: *Robustness Checks: Shapiro-Wilk Normality Tests, Wilcoxon Signed-Rank Tests, and d_{av}*

Comparison	Grade	Shapiro-Wilk W	Shapiro-Wilk p	Wilcoxon W	Wilcoxon exact p	d_{av}
Pre-test → Post-test	8	.93	.012	0.00	<.001	3.75
Pre-test → Post-test	9	.93	.020	0.00	<.001	3.72
Pre-test → Post-test	10	.93	.011	0.00	<.001	3.93
Speaking → Translation	8	.82	<.001	0.00	<.001	4.49
Speaking → Translation	9	.58	<.001	0.00	<.001	4.89
Speaking → Translation	10	.36	<.001	0.00	<.001	4.65

Note. Wilcoxon $W = 0.00$ in every comparison indicates that every student showed a difference in the same direction (post-test > pre-test; translation > speaking), consistent with the very small SD_{diff} values reported in Tables 1 and 2. d_{av} standardises the mean difference by the average of the two conditions' individual standard deviations (Lakens, 2013) and is reported for interpretive comparison against conventional effect-size benchmarks; see Section 3.5 and Section 4.2 for the distinction between d_z and d_{av} .

Qualitative Findings

Rather than merely describing participants' comments, the qualitative data were analysed to explain how learner characteristics interacted with classroom activities and to help interpret the quantitative results. Four interrelated themes emerged from classroom observations, reflective journals, and semi-structured interviews (Table 4).

Table 4: *Qualitative Themes, Illustrative Subthemes, Data Sources, and Representative Quotations*

Theme	Illustrative subtheme(s)	Data source(s)	Representative quotation
1. Learner motivation, confidence, and willingness to participate	Confidence-driven participation; performance anxiety	Student journal; teacher interview; classroom observation	"Role play is enjoyable because I can practise speaking, although I still become nervous when everyone is listening." (Grade 9 Student Journal)
2. Learning pace and classroom constraints	Differential task-completion speed; need for scaffolding in large classes	Classroom observation; teacher interview	"Large classes make it difficult to adjust activities for every learner. Some students require much more practice before they feel ready to speak." (Teacher Interview)
3. Assessment orientation and learning behaviour	Exam-driven effort allocation toward translation and grammar	Student journal; teacher interview	"Translation helps me get marks in exams. Speaking is useful, but it is not examined." (Grade 10 Student Journal)
4. Differential benefits of communicative and grammar-focused activities	Preference variation by learner confidence; disconfirming/negative cases	Student interview; student journal	"I enjoy discussions because I learn to express my own ideas." (Grade 8 Interview); negative case: some high-achieving students found translation repetitive

Theme 1: Learner Motivation, Confidence, and Willingness to Participate

Across all three data sources, learner motivation emerged as a key factor influencing engagement with classroom activities. Students generally recognised the importance of English for academic achievement and future employment; however, their willingness to participate differed considerably. Several students described communicative activities as enjoyable because they encouraged authentic language use and increased confidence over time.

"Role play is enjoyable because I can practise speaking, although I still become nervous when everyone is listening." (Grade 9 Student Journal)

Teachers similarly observed that learners who were already confident participated more actively in pair work and group discussions.

"The students who already have some confidence usually volunteer first. Once they begin speaking, others gradually become involved." (Teacher Interview)

Conversely, classroom observations showed that some students consistently avoided oral participation despite demonstrating adequate written ability. Their reflective journals frequently mentioned anxiety, fear of making mistakes, and concern about peer evaluation, consistent with the broader literature on foreign language anxiety and willingness to communicate (Horwitz, 2001; MacIntyre, 1999; MacIntyre, Clément, Dörnyei, & Noels, 1998). This convergence across observations, journals, and interviews suggests that communicative participation was strongly influenced by learners' confidence and emotional readiness rather than by classroom activity type alone; these findings help explain why speaking-task scores remained comparatively lower despite substantial overall post-test improvement.

Theme 2: Learning Pace and Classroom Constraints

Learners progressed at noticeably different rates throughout the intervention. Observation records consistently showed that higher-performing students completed communicative tasks rapidly and often requested additional activities, whereas slower learners required repeated explanations, vocabulary support, and extended practice before participating confidently.

Observation (Week 5): Several groups completed discussion activities within the allocated time, while two groups continued requesting teacher assistance in constructing simple sentences.

Teachers reported that accommodating these differences was particularly difficult in classes containing approximately forty students.

"Large classes make it difficult to adjust activities for every learner. Some students require much more practice before they feel ready to speak." (Teacher Interview)

These observations echo Tomlinson's (2001) account of the practical difficulty of differentiating instruction by readiness and pace in large, mixed-proficiency classes, and they help explain part of the quantitative variation observed within each grade. Although mean achievement increased substantially, classroom evidence demonstrated that this improvement occurred at different rates among individual learners.

Theme 3: Assessment Orientation and Learning Behaviour

A consistent finding across interviews and journals was the strong influence of examination requirements on students' learning priorities. Many learners regarded grammar exercises and translation tasks as directly relevant to public examinations and therefore invested more effort in these activities.

"Translation helps me get marks in exams. Speaking is useful, but it is not examined." (Grade 10 Student Journal)

Teachers similarly acknowledged that examination pressure shaped both instructional decisions and student expectations. This pattern is consistent with a substantial body of washback research in Bangladesh, which has documented that the grammar- and accuracy-weighted format of the SSC English examinations encourages teachers and students to prioritise tested skills over communicative competence (Ali & Hamid, 2020; Rahman et al., 2019, 2021). Consequently, higher translation-task scores in the present study should be understood within this assessment context rather than interpreted as demonstrating the superiority of grammar-translation-oriented instruction; the qualitative evidence indicates that assessment practices, more than instructional method as such, strongly influenced learner engagement with different language skills.

Theme 4: Differential Benefits of Communicative and Grammar-Focused Activities

Evidence from multiple qualitative sources demonstrated that neither communicative nor grammar-focused activities benefited all learners equally. Students with higher confidence and greater prior exposure to English generally preferred communicative activities because these opportunities improved fluency and self-confidence.

"I enjoy discussions because I learn to express my own ideas." (Grade 8 Interview)

Conversely, students reporting lower confidence frequently preferred translation activities because they perceived grammar-based instruction as more structured and predictable.

"Translation feels easier because I can think carefully before writing." (Grade 9 Interview)

Importantly, several negative cases emerged during analysis. For example, a small number of high-achieving students expressed dissatisfaction with translation exercises because they considered them repetitive and insufficiently challenging, while some initially anxious learners reported increasing confidence following repeated communicative practice. These contrasting experiences demonstrate that learner responses cannot be explained solely by activity type; instead, learner characteristics interacted with classroom context and assessment expectations to shape individual outcomes.

Integration of Quantitative and Qualitative Findings

The mixed-methods design enabled the quantitative and qualitative findings to complement one another. Quantitatively, all three grade levels demonstrated statistically significant, and unusually consistent, improvement between pre-test and post-test. Qualitatively, this improvement was accompanied by accounts of increased learner engagement, greater familiarity with instructional activities, and improved confidence over time.

Similarly, although translation-task scores exceeded speaking-task scores across all grades, the qualitative evidence suggests that this difference reflects examination-oriented learning priorities, greater familiarity with translation-style tasks, and the comparatively demanding nature of spontaneous oral communication, rather than the superiority of one instructional approach. Overall, the convergence of multiple qualitative data sources strengthens the interpretation that learner characteristics and contextual factors, including assessment orientation, shaped how students engaged with different classroom activities and, in turn, their performance on the two task types.

Discussion and Conclusion

This study examined how learner individuality and learning pace relate to English language achievement and task-based performance among secondary students in Bangladesh, using a convergent mixed-methods design. Consistent with the general aim of the instructional intervention, quantitative findings demonstrated statistically significant, and notably consistent, improvement across all three grade levels between pre-test and post-test. The qualitative findings suggest that this improvement cannot be attributed to any single instructional element in isolation; rather, learner motivation, confidence, prior knowledge, learning pace, and assessment expectations collectively shaped students' engagement with classroom activities and their subsequent performance.

One contribution of this revised analysis is a more disciplined separation between task-based performance differences and instructional-method comparison. Although translation-task scores were consistently higher than speaking-task scores, these two tasks assessed different language constructs using different scoring criteria: translation tasks primarily evaluated grammatical accuracy and meaning transfer, whereas the speaking task required spontaneous oral production, interaction, pronunciation, and communicative fluency. Because class-level instructional-method membership was not available as an analytic variable in this dataset (Section 3.1), the present findings cannot speak directly to whether CLT or GTM is the more effective instructional method; they speak instead to how the same students performed differently across two task types, and to why, as illuminated by the qualitative data.

This more cautious framing sits alongside a genuinely mixed comparative literature. Some quasi-experimental studies comparing CLT and GTM directly have found that grammar-translation-oriented instruction produces fewer grammatical errors and stronger vocabulary gains than communicative instruction, particularly on written and accuracy-focused measures (Çiftci & Özcan, 2021; Núñez-Vázquez & Crismán-Pérez, 2021), while others emphasise CLT's advantages for fluency and communicative confidence (Richards & Rodgers, 2014). The present study's task-based pattern, translation scores consistently exceeding speaking scores, is broadly compatible with this literature, but because the design does not isolate instructional method as an independent variable, it cannot adjudicate between these positions; instead, it highlights, through the qualitative data, the assessment-related and affective mechanisms that plausibly contribute to such patterns in an examination-oriented context like Bangladesh.

Chief among these mechanisms is examination washback. The qualitative findings that students prioritised translation and grammar activities because they were directly examined align closely with a substantial body of washback research conducted in Bangladeshi secondary schools, which has documented that the grammar- and accuracy-weighted format of the SSC English examinations narrows classroom practice toward tested skills and away from speaking and listening, regardless of the curriculum's official communicative orientation (Ali & Hamid, 2020; Rahman et al., 2019, 2021). The present findings extend this literature by showing, at the level of individual task performance rather than teacher or student self-report alone, a pattern consistent with that narrowing: students performed comparatively better on the task type that mirrors what is examined.

The qualitative findings also demonstrated considerable variation among learners in their response to communicative activities specifically. Students with higher confidence generally participated more actively in speaking activities, whereas learners experiencing greater anxiety often preferred more structured, grammar-based activities. These patterns are consistent with foundational work on foreign language anxiety and willingness to communicate (Horwitz, 2001; MacIntyre, 1999; MacIntyre, Clément, Dörnyei, & Noels, 1998), which frames a learner's readiness to speak as the outcome of layered situational and affective influences rather than of linguistic competence alone.

Similarly, the marked differences in learning pace observed through classroom observation echo Tomlinson's (2001) account of the practical difficulty of differentiating instruction for readiness and pace in large, mixed-proficiency classrooms of the kind common in Bangladesh.

Taken together, these findings support post-method pedagogy, which argues that instructional decisions should be adapted to local context and learner diversity rather than resting on the presumed superiority of a single teaching method (Kumaravadivelu, 2001, 2006), and they support the socio-ecological perspective on individual differences proposed by Spring, Pawlak, and Li (2021), which frames learner characteristics as operating dynamically within particular classroom environments rather than as isolated, stable variables. The mixed-methods evidence as a whole indicates that classroom activities are best viewed as complementary pedagogical resources whose relationship with learning outcomes is shaped by learner characteristics, learning pace, and assessment orientation, rather than as an overriding determinant of outcomes considered on its own.

This study investigated English language achievement and task-based performance among secondary students in Bangladesh, with particular attention to learner individuality and learning pace. Both quantitative and qualitative findings demonstrated that students achieved significant, and unusually consistent, improvement over the course of the 12-week instructional intervention. However, the findings also indicate that this improvement, and the pattern of task-based performance observed alongside it, cannot be explained by instructional method alone. Instead, learner motivation, confidence, prior knowledge, learning pace, and assessment expectations shaped how students engaged with different classroom activities.

Although translation-task scores exceeded speaking-task scores in every grade, this difference should be interpreted cautiously: the two assessments measured different language constructs using different scoring criteria, and the present design does not support a controlled comparison of CLT and GTM as instructional methods. Rather than demonstrating the superiority of one instructional approach, the findings suggest that different classroom activities foreground different aspects of language learning, and that examination orientation, learner confidence, and learning pace jointly shape how students perform on each.

Accordingly, the study supports context-sensitive teaching in which instructional activities are selected flexibly according to learner needs, instructional goals, and classroom realities, rather than through adherence to a single methodological tradition. Although the findings are limited to one Bangladeshi secondary school and do not permit a formal comparison of instructional-method effectiveness, they contribute empirical evidence, grounded in both achievement data and task-based performance data, in support of post-method pedagogy within examination-oriented educational contexts.

Implications and Recommendations

Practical Implications

Principled Eclecticism: Teachers should integrate communicative and grammar-based activities according to lesson objectives and learner needs rather than relying exclusively on one instructional approach (Kumaravadivelu, 2001, 2006).

Differentiated Instruction: Instruction should accommodate differences in learner confidence, proficiency, and learning pace through flexible grouping, scaffolded activities, and additional support where necessary (Tomlinson, 2001).

Assessment Reform: Giving greater weight to speaking and communicative assessment within national examinations may reduce negative washback and encourage greater classroom emphasis on communicative competence (Ali & Hamid, 2020; Rahman et al., 2021).

Teacher Training: Professional development should focus on managing learner diversity, adapting instruction to large classes, and balancing communicative and examination-oriented teaching.

Learner Support: Schools should provide additional support for slower-progressing learners through remedial classes or peer tutoring.

Limitations and Future Research

Several limitations should be acknowledged.

- The study was conducted in a single secondary school, limiting the generalisability of the findings.
- The intervention lasted only twelve weeks and therefore cannot capture longer-term instructional effects.
- Students remained within intact classrooms, and class-level instructional-method assignment was not retained as an analytic variable in the dataset analysed for this study. As a result, the study cannot support a formal

between-method (CLT versus GTM) statistical comparison, such as a time $\times$ method interaction analysis or ANCOVA; the quantitative findings reported here describe within-student change over time and within-student differences between task types, not between-method effectiveness (Section 3.1).

- Because students were nested within three intact classes, observations within a class cannot be assumed fully independent; the analyses reported here do not model this clustering explicitly. Multilevel or hierarchical linear modelling (Raudenbush & Bryk, 2002) would be a more appropriate approach for future datasets that retain class-level identifiers.
- Although two raters independently scored a subset of performances before consensus scoring, a formal inter-rater reliability coefficient (e.g., ICC; Koo & Li, 2016) was not calculated; rating consistency should therefore be interpreted cautiously.
- Implementation fidelity was monitored descriptively through periodic classroom observation rather than through a formal, quantitatively scored fidelity checklist (Century, Rudnick, & Freeman, 2010); the degree to which each teacher's classroom practice adhered to its intended instructional emphasis over the full 12 weeks is therefore not quantified.
- Qualitative coding was conducted by one researcher. Although credibility was strengthened through triangulation, an audit trail, reflexive memo-writing, and peer debriefing (Lincoln & Guba, 1985), future studies employing multiple independent coders would further strengthen analytical trustworthiness.
- Unmeasured socio-economic differences among learners, such as access to private tutoring or exposure to English outside school, may have affected outcomes and should be examined more systematically in future research.
- The very low variability in individual gain scores (pre-test to post-test) and in speaking-translation differences is a notable feature of this dataset; while it does not appear to reflect data-entry error or the presence of outliers (Section 3.5), the sources of this consistency (for example, a highly structured intervention, a ceiling effect in some rubric components, or characteristics of the specific assessment instruments) were not directly investigated here and may warrant closer examination in future work.

Despite these limitations, the study provides empirical insight into how learner individuality, learning pace, and assessment orientation relate to English language performance in a large-class, examination-oriented ELT context. Future research could include multiple schools and classes with class-level instructional-method identifiers retained for formal between-method comparison, a longer intervention duration, and more diverse regional contexts. Longitudinal designs could track learners over multiple years to examine how learning pace and instructional approach interact over time, and studies with larger numbers of independently sampled classes could apply multilevel modelling to separate student-, class-, and school-level sources of variation (Raudenbush & Bryk, 2002).

References

- Abedin, M. M. (2013). The present mode of teaching in the ELT classes at the Higher Secondary level in Bangladesh: Is it the practice of CLT or disguised GTM? *Stamford Journal of English*, 7, 1-15. <https://doi.org/10.3329/sje.v7i0.14459>
- Ali, M. M., & Hamid, M. O. (2020). Teaching English to the test: Why does negative washback exist within secondary education in Bangladesh? *Language Assessment Quarterly*, 17(2), 129-146. <https://doi.org/10.1080/15434303.2020.1717495>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Brown, H. D. (2014). *Principles of language learning and teaching* (6th ed.). Pearson.
- Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, 31(2), 199-218. <https://doi.org/10.1177/1098214010366173>
- Çiftçi, H., & Özcan, M. (2021). A contrastive analysis of traditional grammar teaching and communicative language teaching in teaching English grammar and vocabulary. *International Online Journal of Education and Teaching*, 8(2), 709-729.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

- de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7-21. <https://doi.org/10.1017/S1366728906002732>
- Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, 21, 112-126. <https://doi.org/10.1017/S0267190501000071>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly*, 35(4), 537-560. <https://doi.org/10.2307/3588427>
- Kumaravadivelu, B. (2006). *Understanding language teaching: From method to postmethod*. Routledge.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.
- MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom environment* (pp. 24-45). McGraw-Hill.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545-562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Maqbool, M. (2024). Adapting language teaching methods: Comparative acceptability of GTM and CLT in Bangladesh's rural secondary education. *International Journal of Language, Literature and Culture*, 4(12), 1-4.
- Matin, A. A. (2021). Incorporation of GTM and CLT: An effective way to teach. *Global Journal of Human-Social Science*, 21(G3), 47-51. <https://doi.org/10.34257/GJHSSGVOL21IS3PG47>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Núñez-Vázquez, I., & Crismán-Pérez, R. (2021). A study of the differences among EFL/ESL methods for reading comprehension and language awareness. *Porta Linguarum: International Journal of Foreign Language Teaching and Learning*, (35), 9-24. <https://doi.org/10.30827/portalin.vi35.15412>
- Rahman, K. A., Seraj, P. M. I., Hasan, M. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, 11(1), Article 12. <https://doi.org/10.1186/s40468-021-00129-2>
- Rahman, M. M., Islam, M. S., Karim, A., Chowdhury, T. A., Rahman, M. M., Seraj, P. M. I., & Singh, M. K. M. (2019). English language teaching in Bangladesh today: Issues, outcomes and implications. *Language Testing in Asia*, 9(1), Article 9. <https://doi.org/10.1186/s40468-019-0085-8>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press.
- Spring, U., Pawlak, M., & Li, S. (2021). Individual differences in language learning and teaching: A complex/dynamic/socio-ecological/holistic view. *English Teaching & Learning*, 45, 339-353. <https://doi.org/10.1007/s42321-021-00085-3>
- Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Association for Supervision and Curriculum Development.

Extended English Summary

Introduction

The study asks whether teaching method is truly the decisive factor in second-language learning, or whether learner-level differences matter just as much. It positions itself against the long-running "method wars" in English Language Teaching (ELT), which have cycled through grammar-translation, direct, audiolingual, and communicative approaches in search of one best practice. Drawing on Kumaravadivelu's post-method pedagogy and complex dynamic systems theory, the author argues that instructional methods are supplementary tools rather than decisive levers, and that learner individuality - motivation, anxiety, confidence, and prior knowledge - together with learning pace, the relative speed at which students grasp and produce language, shape outcomes as much as, or more than, the teaching approach used.

The Bangladeshi secondary-school setting sharpens this argument. English is compulsory and widely regarded as a route to social mobility, and the national curriculum nominally favors Communicative Language Teaching (CLT). In practice, however, large classes, limited resources, and high-stakes Secondary School Certificate (SSC) examinations that reward grammar, reading, and writing over speaking and listening push many teachers toward a grammar-translation orientation, producing the "washback" effects documented in earlier Bangladeshi research.

Set in Grades 8–10 classrooms of 35–45 students, the study poses three research questions: how much did achievement change over a 12-week instructional intervention, and how did performance differ between communicative (speaking) and accuracy-oriented (translation) tasks; how were learner individuality and pace associated with engagement and performance across these task types; and what are the implications for teaching practice, assessment design, and teacher training. A convergent mixed-methods design was chosen to address all three questions together, combining achievement data, task performance, and qualitative accounts of learner experience.

Method

The researcher used a convergent mixed-methods design (Creswell & Plano Clark, 2018), collecting quantitative and qualitative data concurrently across a 12-week intervention in one secondary school in Bangladesh. The sample was 120 students spanning Grades 8–10 (40 per grade), drawn from three intact classes - a purposive, single-school convenience sample. Every enrolled student who completed the intervention and provided parental consent and personal assent was retained, so no participants were excluded, and pre-test scores varied meaningfully within each grade, indicating a genuine spread of starting proficiency.

Four teachers took part; day-to-day, two leaned more toward CLT and two toward grammar-translation, though all followed the same curriculum, schedule, and learning objectives. Which class or teacher each student belonged to was not retained as an analytic variable in this dataset. As a result, the quantitative comparisons reported below track each student's own performance across time and across task type - they do not compare CLT-taught students against GTM-taught students, and the author makes no claim about relative method effectiveness.

Six data sources were used: an achievement pre-test/post-test covering reading, grammar, vocabulary, and writing; a speaking task scored for fluency, accuracy, pronunciation, and interaction; a translation task scored for grammatical accuracy, lexical choice, and meaning preservation; periodic structured classroom observations; weekly student reflective journals; and semi-structured interviews with 18 students and all four teachers. Two raters double-scored 20% of speaking and translation performances and resolved disagreements by consensus, though no formal inter-rater reliability statistic was calculated.

Quantitative data were analyzed using paired-samples t-tests conducted separately for each grade, with non-parametric Wilcoxon signed-rank tests run as a robustness check, all in Python. Qualitative data underwent inductive thematic analysis (Braun & Clarke, 2006), conducted by a single researcher, with credibility supported through triangulation across data sources, an audit trail, reflexive memo-writing, and peer debriefing. No formal ethics-committee review was obtained, as none is mandated for independent, non-medical educational research of this kind in Bangladesh; the school's Head Teacher granted institutional permission, and informed consent and assent were secured from guardians and students respectively.

Findings

Achievement rose sharply and consistently across all three grades. Grade 8 students improved from a pre-test mean of 44.68 to a post-test mean of 64.25 (a gain of 19.6 points), Grade 9 from 47.40 to 66.88 (19.5 points), and Grade 10 from 45.28 to 64.55 (19.3 points), all significant at $p < .001$. The effect sizes were unusually large (Cohen's d above 1.0 in every grade) because pre- and post-test scores were highly correlated within students and gains varied little from one student to the next; the author treats this as evidence of a consistent improvement across nearly the entire sample rather than a statistical artifact, and also reports the more conventional d statistic (3.7–3.9) for comparison against typical benchmarks. Wilcoxon tests confirmed the pattern in every grade, with every student improving.

A second comparison examined task type rather than time. Within the same students, translation-

task scores exceeded speaking-task scores by a strikingly consistent margin of roughly 13 points in every grade (for example, Grade 8: 56.68 versus 69.63), again highly significant. The author are careful to note that this is not evidence that grammar-translation instruction outperforms communicative instruction: the two tasks measure different constructs (spontaneous oral production versus written grammatical accuracy) using different rubrics, and instructional-method group was never tracked as a variable, so no between-method claim follows.

The qualitative strand, drawn from observations, journals, and interviews, produced four themes that help explain these patterns. Motivation, confidence, and willingness to participate varied considerably: some students found speaking activities like role play enjoyable but nerve-wracking, and confident students volunteered first, while others avoided speaking despite adequate writing skills. Learning pace and classroom constraints were also prominent: in classes of around 40 students, faster learners finished tasks quickly and wanted more, while slower learners needed repeated explanation and scaffolding - teachers found this hard to accommodate.

Assessment orientation strongly shaped effort, with students repeatedly explaining that they invested more in translation and grammar because these skills are examined while speaking is not - as one Grade 10 student put it, "Translation helps me get marks in exams" - a pattern consistent with prior Bangladeshi washback research. Finally, the benefits of communicative versus grammar-focused activities were not uniform: more confident students generally preferred discussion-based activities, while less confident students found translation more predictable and manageable, though the researchers also identified genuine exceptions, including high-achieving students who found translation repetitive and initially anxious students who grew more confident through repeated speaking practice.

Discussion and Conclusion

Together, the two strands of evidence support the paper's central claim: instructional method alone does not explain the results. The consistent achievement gains likely reflect broadly effective teaching rather than any single technique, while the translation-over-speaking performance gap is best explained by examination washback, greater familiarity with translation-style tasks, and the greater demands of spontaneous speech - not by the inherent superiority of grammar-translation teaching. The author distinguish their work from studies that directly compare CLT and GTM groups (some favoring grammar-translation for accuracy and vocabulary, others favoring CLT for fluency and confidence); because instructional-method group was not retained as a variable here, the study cannot adjudicate that broader debate.

Instead, its contribution lies in showing, within one context, how affective factors such as confidence and anxiety, pacing differences, and exam-driven priorities jointly shape both engagement and measured performance. This is consistent with complex dynamic systems theory's view of individual differences as dynamic and interacting rather than fixed traits, and with Spring, Pawlak, and Li's (2021) socio-ecological framing of learner variation. It also offers empirical support for Kumaravadivelu's post-method pedagogy: rather than searching for one correct method, the findings suggest classroom activities function as complementary resources whose value depends on who is using them, at what pace, and under what assessment pressure.

The author are candid about the limits of this framing. A single school, a 12-week window, the choice not to retain class-level method identifiers (ruling out a formal CLT-versus-GTM comparison), unmodeled clustering within four classes, no formal inter-rater reliability statistic, and single-coder qualitative analysis all limit how far the findings generalize. Even so, the study concludes that its evidence meaningfully supports treating learner individuality and learning pace as central, rather than peripheral, considerations in English language teaching.

Recommendations

The study closes with five practical recommendations for classroom practice and policy:

- Principled eclecticism - blend communicative and grammar-based activities according to lesson goals and learner needs, rather than committing exclusively to one method.
- Differentiated instruction - accommodate differences in confidence, proficiency, and pace through flexible grouping and scaffolded tasks.
- Assessment reform - give greater weight to speaking and communicative competence in national examinations to ease negative washback and rebalance classroom priorities.
- Teacher training - build professional development around managing learner diversity and large class sizes while balancing communicative and exam-oriented demands.
- Learner support - offer remedial sessions or peer tutoring for slower-progressing learners.

The author also highlights directions for future research: multi-school studies that retain class-level method identifiers to enable a genuine CLT-versus-GTM comparison; longer intervention periods; multilevel modeling to account for classroom clustering; and closer attention to unmeasured factors such as socio-economic background, out-of-school English exposure, and the use of multiple independent qualitative coders.

