# Differentiated Analytic Assessment in Multilingual Classrooms: A Case Study of B1-Level Learners Preparing for The B2 Uzbekistan State English Exam

Yulduz Sultanova

MA TESOL Candidate, Webster University, USA

**Abstract:** This research explores the application of differentiated analytic testing in the multilingual environment of English as a Foreign Language (EFL), with particular regard to four learners of B1 level preparing to take the essential B2 Multilevel English test in Uzbekistan. Utilizing the case study research method based on Krashen's Input Hypothesis and Kinginger's sociocultural models, the research presents the vital roles of the adapted rubrics, scaffolded testing, and emotional preparedness in providing equal and reliable evaluations of learners' abilities. The main findings underscore the essential need for research methods with cultural sensitivity and tailored support in tests to facilitate linguistic balance and provide equal learning opportunities.

**Keywords:** Differentiated assessment, analytic rubrics, CEFR, multilingual learners, formative assessment, Uzbekistan, sociocultural theory.

**Introduction:** The field of language testing in educational contexts has undergone considerable change, evolving from rigid, standardized testing practices to more adaptable and context-responsive approaches. These innovations reflect a deep understanding that the process of language learning is not one-size-fits-all across varying contexts. This understanding is especially important in multilingual contexts, like Karakalpakstan in Uzbekistan, where students routinely interact with multiple languages—often using Karakalpak at home, learning the official Uzbek language in schools, attending Russian-language schools, and learning English as a foreign language—thereby requiring flexible testing systems. Traditional tests that are largely monolingual in orientation do not reflect the higher-order abilities that such students exhibit. Such shortcomings are particularly evident with regard to high-stakes language testing, since its results can have a considerable impact on both academic pathways and career prospects.

This study looks into the measures used in designing and administering varied and analytic evaluations in linguistically diverse classroom learning, taking into consideration four male teenagers who share different cognitive, affective, and linguistic traits. These teenagers are preparing to take the Uzbekistan State Multilevel B2 English test, the key entry to accessing further education. Although in total they are placed on the B1 level of proficiency, they have different capabilities with regard to the four skills of English. Such difference in capabilities requires the evaluation to be adapted to meet CEFR specifications and, at the same time, cover the special demands of each individual.

Differentiated assessment is an adaptive, learner-focused approach that adjusts instruments and methods of evaluation to meet the varied readiness, interests, and learning styles of learners. Such an approach is contrasted with standardized testing, which can provide unbalanced representations of learners' skills. The use of analytic rubrics—breaking down language skills into separate criteria such as vocabulary, grammatical correctness, text coherence, and task achievement—enriches teachers' ability to offer more detailed and helpful feedback. These instruments support formative assessment activities by identifying individual strengths and weaknesses and allowing targeted instructional intervention.

This paradigm's underlying theory is Stephen Krashen's Input Hypothesis, which emphasizes the role of affective filters in second-language learning processes. The provision of comprehensible input in affectively secure and engaging environments greatly maximizes the potential of language learning. The work of Claire Kramsch and Claire Kinginger extends these sociocultural theory tenets, hypothesizing second-language learning within social and emotional settings influencing learners' self-image and level of commitment. These perspectives collectively indicate that tests need to be technically proper and also responsive to learners' experiences.

Methodologically, the research utilizes a qualitative case study design. Four learners attended an extra-curricular English program designed explicitly to prepare learners to take national proficiency tests. A range of sources were drawn on to obtain data, including classroom observations, portfolios of work from the learners, audio recordings of spoken tasks and analysis of examination artifacts. Linguistic difference and socio-economic inequality in the classroom, combined with differences in the learners' academic backgrounds, created a rich setting in which to assess the efficiency of differentiated analytic assessment.

Each of the students showed their own distinct learning profile. Student A was described as an independent learner with capable receptive abilities; however, he struggled with speech fluency due to Tourette syndrome. Student B showed notable reliance on memorized structures and limited inherent motivation, often requiring quite significant scaffolding to engage in effective activities. Student C showed scholastic potential but was introverted and subjected to significant familial responsibilities that impacted his ability to meet formal timelines. Student D showed emotional sensitivity and struggled in traditional classroom settings, performing better in asynchronous and low-stakes tasks.

The varied learner profiles called for an overall evaluation method. With regard to vocabulary and reading improvement, material was adapted in accordance with research by Brown (2004), including tasks of multiple-choice and short-answer with inferences. Vocabulary improvement was promoted by the use of visual and contexts-specific illustrations before utilizing them to form meaningful sentences. Task organization followed a logical input-to-output sequence, thus catering to less proficient learners by setting the ground before requesting their external involvement.

For writing, the students worked on the formal letter task under the auspices of Hughes's (2002) research.

They studied letter samples, identified features of cohesion, and redesigned poor models in collaboration before attempting the main task. The prewriting tasks enabled internalization of text structure and lexical choices among the students, particularly students B and D. The final writing task was assessed with an analytic rubric that included task achievement, vocabulary use, grammatical accuracy, cohesion, tone, and mechanics. Time limits and word constraints were imposed to simulate testing conditions.

The assignments of speeches were both adaptable and accessible. The opportunity to take home and record their answers lessened anxiety on the part of A and D. The inquiry questions were designed to meet the standards of the task and facilitate variability in question form. The students were encouraged to listen to their recordings and, optionally, re-record their answers, thus engendering metacognitive awareness and self-assessment. The rubrics treated fluency with lesser importance in situations of speech difficulty on the part of the students, instead accentuating idea development, vocabulary richness, and coherence.

Rubrics were also found to be useful not just in assessment but also in instructional strategies. CEFR-based analytic rubrics, which were made available to students prior to and following tasks, explained the evaluation standards and helped to inform improvement. Standards set were not only evaluation in scope but also instructional, enabling both teachers and students to identify exactly areas of improvement needed. Students also demonstrated increased confidence and reduced anxiety as they were very familiar with the evaluation criteria and improvement steps needed.

The cumulative impact of different measures of assessment was realized in regard to learners' performances and engagement. Learners who previously showed hesitance to participate showed enhanced willingness to take on challenges. The degree of freedom and empowerment also increased, particularly with the learners who previously struggled with standardized testing formats. Such findings support the assertion by Taylor and Chen (2016) that equitable measures of assessment need to fit the varied needs of learners and maintain construct integrity.

To foster a more nuanced understanding of the implications of this practice, it is crucial to explore how these evaluation strategies impacted classroom relationships. When assessment was viewed as supportive and equitable and not punitive in nature, there was a significant shift towards increased learner motivation. Student reflections and casual interviews

implied that the learners found assessment to be the way forward and not the punitive evaluation fraught with risk of failure. The learners came to see feedback as an educational asset, which increasingly shifted the nature of the classroom to be more open, collaborative, and learner-centered. This was evident with regard to students B and D, who previously showed reluctance to participate due to low self-efficacy and anxiety. These developments highlight the emotional implications of inclusive assessment models and the potential of personalized, formative assessment to transform.

In addition, the dual role of the classroom teacher as both facilitator and assessor were essential in creating these beneficial outcomes. Classroom spaces guided by sociocultural considerations enhanced learners' capacities to construct meaning, jointly construct knowledge, and engage in reflective practices about their learning in collaboration. Assessment was linked to learning and not seen as a distinct activity. This embedding of instruction and assessment is of particular importance in contexts where learners are faced with various challenges such as limited access, pressure situations, and socio-emotional challenges.

The other important implication of this case is related to the extension of differentiated assessment strategies to more general or institutional contexts. While the bespoke accommodations that were illustrated in this case study were practicable in the small group setting, the same principles can be transferred to large classroom settings by utilizing flexible assessment modules, multiple formats of assessments, or more systematic use of peer and self-assessment in instructional design. When supplemented with institutional support and professional development, these approaches can be scaled into entire educational systems, thus advancing systemic inclusivity.

In addition, it is important to integrate policy discourse to harmonize national and institutional evaluation models with various methodological approaches. Assessment bodies, ministries of education, and curriculum writers would need to infuse flexibility into the design of the examination to take into consideration differences in modalities, timelines, and critical benchmark points. This would avoid exclusion of learners with heterogeneous backgrounds and competencies. Pursuit of equitable access policies not only ensures fairness in individual classrooms but also fosters educational justice on a wider systemic level.

On a practical level, this study underlines the need to balance flexible and empathetic evaluation techniques in linguistically diverse learning settings. Educators

need to balance the conflicting imperatives of academic integrity with the unique learning requirements of each student. Strategies like scaffolded instruction, analysis rubrics, and assessments based on differentiation provide the tools to balance the equation. Educators' professional development programs need to emphasize such methods, and accountability policies need to integrate accommodations as standard strategies and not as special measures.

The results of this case study confirm that analytic differentiated assessment goes beyond simple conceptual goals and becomes a fundamental pedagogy of instruction in multilingual and cognitively disparate settings. Far from undermining standards of education, this model increases the validity of tests by matching tasks more closely to the true capabilities and histories of learners. By utilizing systematically constructed tasks, flexible delivery, and CEFR-aligned analytic rubrics, teachers can not only produce better examination results but also create more stimulating and more equitable learning experiences. As such, this transforms the nature of language testing from a limiting barrier into a facilitative tool—ultimately furthering the rights of all learners to access, engage, and succeed.

## REFERENCES

Ayhan, Ü., & Türkyılmaz, M. U. (2015). Rubrics and rubric design. International Journal of Language and Linguistics, 2(2), 82–92.

Brown, H. D. (2004). Language assessment: Principles and classroom practices. Longman.

Buck, G. (2009). Challenges and constraints in language testing. In D. Tsagari & J. Banerjee (Eds.), Handbook of second language assessment (pp. 203–216). De Gruyter Mouton.

Chapelle, C. A. (2021). Argument-based validation in testing and assessment. SAGE Publications.

Gottlieb, M., & Ernst-Slavit, G. (2019). Promoting educational equity through teacher education: Examples from global contexts. Multilingual Matters.

Hughes, A. (2002). Testing for language teachers (2nd ed.). Cambridge University Press.

Kinginger, C. (2009). Language learning and study abroad: A critical reading of research. Palgrave Macmillan.

Krashen, S. D. (1985). The Input Hypothesis: Issues and implications. Longman.

Taylor, L., & Chen, N. N. (2016). Assessing students with learning and other disabilities/special needs. In D. Tsagari & J. Banerjee (Eds.), Handbook of second

language assessment (pp. 377–395). De Gruyter.