Rethinking Translation Evaluation in Academic Contexts: Performance-Based Assessment as an Alternative Practice

[1*]Hossein Heidari Tabrizi, Professor, Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
heidaritabrizi@gmail.com
[2]Juliane House, Professor Emerita, University of Hamburg, Hamburg, Germany
juliane.house@uni-hamburg.de

Abstract

In the context of translator education, evaluating students' academic translations is a common, crucial activity. Evaluations are usually done by assigning numerical scores or letter grading in order to detect the relation established between intended instructional objectives and learning outcomes. Research shows that the dominant translation evaluation methods currently and commonly practiced by university teachers in undergraduate English translation programs are mainly based on the theoretical principles of the conventional testing paradigm. This paper tries to "address the 'problem' of evaluative practices provoked from the challenges and criticisms leveled against principles and procedures of Classical True Score Measurement Theory and conventional testing tradition. In doing so, the paper first reviews the status quo of translation evaluation and then pinpoints the problems and limitations which are typically associated with translation tests. As a result of the shift from conventional testing to language assessment and from language tests to alternative assessment practices, performance-based evaluation is introduced as one of the potential solutions that may help fill the gaps left by traditional testing. The paper concludes that Performance-based assessment is well-suited for translation evaluation, as both share a lot in terms of nature, features, and purposes. The contributions made by the current work hopefully may open new avenues for further exploration in this decisive facet of research on translation studies, namely, academic translation evaluation.

Keywords: Performance-based Assessment; Students' Academic Translation Evaluation; Translation Assessment; Translation Evaluative Practices, Translation Tests

INTRODUCTION

Language assessment and evaluation are essential and integral constituents of the teaching-learning process done for pedagogical purposes in any academic setting. Effective testing and assessment are critical in revealing the extent to which students learn what they should learn. Equally, translation assessment through quality evaluation is an indispensable part and parcel of every translation program. In the context of translator education, a common activity of crucial concern is evaluating students' academic translations through assigning numerical scores or letter grading in order to detect the relation established between intended instructional objectives and learning outcomes. Inevitably, in the academic career of every translation teacher/student, there are always diagnostic tests, mid-term and final examinations, and less possibly other more formative assessments (Huertas-Barros et al., 2019), which are to be developed,

administered, evaluated, and scored. Based on such scores, translation teachers make decisions that might have serious impacts on students' educational careers, particularly and their entire lives, generally. That is why translation assessment and evaluation are of significance in translation programs. The more severe these impacts are, the more critical translation evaluation will be. Therefore, evaluations made must be fair, justifiable, strongly dependable, and highly credible. The consequential validity and accountability of these measuring instruments used for evaluation are of utmost importance, too.

An instructor of translation, much like teachers of any other discipline, is expected to help learners develop their actual performance; as an academic member of staff, s/he is required to assess the quality of their students' work. In fact, the central role with which translation teachers are normally associated is twofold: they serve as facilitators of the learning process, and, at the same time, they must act as evaluators of what students have achieved. In other words, as Honig (1998) argues, in the academic context, judging the translation quality "should not be an end but a means" (p. 32). Thus, as a rule of thumb, the process of evaluating students' academic translations and decision-making is certainly one of the most challenging tasks a translation teacher faces. In the words of Adab (2000, p. 227), "Translation assessment in the university environment is a problematic issue." The problem gets much more decisive as more and more students are attracted to translation programs, "mainly oriented towards training future professional commercial translators and interpreters and serve as highly valued entry-level qualifications for the professions" (Munday et al., 2022, p.11).

There has been a massive spreading out of translator education programs throughout the world since the turn of the century to meet the excessive demand for translation. In consort with such rapid proliferation in quantity, today, a growingly sophisticated body of research and knowledge has also been devoted to improving the quality of translation curricula, including its various aspects such as curriculum design, teaching methods, instructional materials, etc. However, lesser attention has been given to the often neglected teacher evaluative practices of student translations. In fact, in the context of translator education, translation evaluation is a widespread yet challenging activity of vital significance typically carried out by teachers in translation courses. Accentuating its significant role in translator training and education, Campbell and Hale (2003) explain that "better assessment means better translators and interpreters" (p. 221).

Nevertheless, this common practice has received the least attention in Translation Studies during the past decades (Arango-Keeth & Koby, 2003; Bowker, 2000). At the turn of the last century, Hatim and Mason (1997, p. 197) assert that in comparison to other areas in Translation Studies, "little is published on the ubiquitous activity of [translation] testing and evaluation"; thus considering it an "under-researched and under-discussed" area where, in words of McAlester (2003), the field suffers from "its worst failure." More recently, Tsagari and Van Deemter (2013), while admitting that a growing number of studies can be found in the field of translation assessment and evaluation, argue that "there is still much work to be done" (p. 11). Dungan (2013) echoes the same idea that the area of assessment in the applied branch of Holmes' (1988/2000) map of translation studies "requires more critical and systematic research" (p. 132). In the same vein, Conde (2013) also believes that translation evaluation within this field "is still a much-debated topic" (p. 108) on which many questions have remained unanswered yet. In the early 2020s, most recently published works confirm that the situation remains rather unchanged (Abdel Latif, 2020; Heidari Tabrizi, 2021a, 2022a; Yazdani et al., 2020, 2023).

Language Assessment: Status quo

Generally speaking, in the historical development of the discipline, a distinction has been made in the related literature between language testing and language assessment in terms of their definitions, foci, purposes, and scopes (Bachman & Damböck, 2018; Cheng & Fox, 2017). Initially, the primary focus of assessing language proficiency tended to be on tests as the sole measuring instruments. In the new millennium, however, the focus has shifted away from language testing to language assessment, which encompasses a broader range of multiple evaluation purposes, methods, and measures beyond standardized tests. The two systems and their defining characteristics are briefly introduced here.

Language testing refers to the tradition of measuring the language ability, competence, or proficiency of an individual, typically dominated by using paper-and-pencil tests. In other words, it is used specifically for the process of measuring individuals' language proficiency through standardized tests (Bachman, 1990; Bachman & Damböck, 2018). It typically involves the administration of formal tests or examinations designed to measure specific language skills or overall proficiency according to predefined criteria. Conventional language testing often focuses on producing numerical or categorical scores or ratings that can be used for purposes such as certification, admission, placement, or employment. They tend to emphasize the measurement aspect and may not always encompass the broader goals and practices associated with language assessment. By providing one single sample of a student's performance in a specific domain on the measure, these tests are designed to measure their learning achievement status at a particular point in time. They are usually administered after the instruction process ends under strict formal procedures and time limitations.

On the other hand, language assessment, in its broadest sense, is used for the entire discipline and its tools. It refers to any formal and informal processes of systematic gathering of information for the purposes of judgments about individuals using tests as well as non-tests, whether done quantitatively or qualitatively and used to make decisions or not (Brown & Abeywickrama, 2019). As defined by Farhady (2021), it is "an ongoing process of collecting information about the students learning" (p.123). In brief, language assessment includes various assessment practices aimed at promoting learning, providing feedback, and supporting language development; language testing can be considered a subset or one aspect of language assessment.

Traditionally, language/translation testing has been seen as an outside force imposed upon the curriculum generally and the learner specifically. In the last century, the prevalent use of standardized paper-and-pencil tests in academic contexts was universally taken for granted with no serious objection. The approval of these tests by many shareholders was largely due to the relative efficiency with which a large number of students on multiple learning outcomes could be measured. Moreover, they are assumed to be inexpensive, easy to score, and simple to record and interpret the results. In sum, due to their optimal practicality and high reliability, conventional tests have typically been welcomed by practitioners in the field (Salmani Nodoushan, 2008). Nevertheless, due to the dramatic changes in assessment purposes, methods, and measures, these tests have been criticized for the way they measure student achievement.

Criticisms against Conventional Tests

Conventional tests are still extremely popular in educational contexts as they enjoy so-called high reliability and optimal practicality. They are cost-efficient, relatively easy to administer, and score in a little time; in addition, the results are also rather easy to report and interpret. The results are said to be 'objective' in the sense that personal biases and values of the scorers do not affect the scores, or at least their impact is minimized. As the main –and usually the only— measuring instruments, conventional tests

enjoy uniformity in all students taking the same test, which adds to their objectivity. Used as typical gatekeepers, they are accepted as a standard norm in educational programs in general and translation programs in particular.

However, severe criticisms have been leveled against the widespread use of paper-and-pencil tests. According to Brown and Abeywickrama (2019), "In the public eye, tests have acquired an aura of infallibility in our culture of mass-producing everything" (p. 16). In fact, the test results tend to be accepted by many 'uncritically' (Bailey, 1998). Surprisingly enough, most telling criticisms commonly arise from the same characteristics contributing to their benefits. As discussed earlier, these features can be summarized as one single sample taken at the end of instruction under time limit pressure and formal procedure. It is rightly proclaimed that one single sample of performance is not enough and may not be representative of the real performance of an individual in different contexts and occasions. If the student has a bad day (like an illness at the time of the test), it sticks with them! On the other hand, there are also deep concerns about potential cheating, coaching to the test, test-taking strategies, and test-wiseness regarding students taking a test successfully without having adequate knowledge of the subject matter in question. In both cases, the sample taken is faulty, and the decisions made on the basis of test results are under serious question in terms of reliability and validity. In sum, conventional language tests cannot serve as true indicators of how students might perform in real, authentic situations.

Moreover, tests tend to provoke adverse feelings such as fear of failure, stress, anxiety, and fatigue more as they are administered under time limit pressure and formal procedure, thus impeding cognitive abilities. The main point, as explained by Brown and Abeywickrama (2019), is that traditional tests "have a way of scaring students" (p.1), whereas language assessment practices "need not be degrading and threatening" (p. 2). Traditional testing tends to act as an "add-on" to instruction and a crushing burden to students in terms of workload and adverse consequences. With the same token, another criticism comes from the negative washback effects traditional testing has on teaching. The emphasis on preparing students for standardized tests can lead to a "teaching to the test" approach, where educators prioritize test-taking strategies and content coverage over deeper learning experiences. The authoritative power of standardized tests is said to force teachers to narrow their instruction to certain discrete points so that students are prepared for the mandated test, helping them improve their scores. Traditional tests have frequently been criticized for having both potentially negative academic and social consequences.

Conventional tests are also criticized for looking for perfection rather than student progress. Instead of stimulating collaboration and weighting progress or the act of learning itself, tests promote competition among students as individuals in isolation, overemphasizing achievements and outcomes alone. They have a contribution to the 'academic race' where only the most knowledgeable student matters, not others. A standardized test "penalizes students who are achieving well but are not in the top percentage of students, and thus can cause these students to lose their motivation and interest in learning" (Bachman & Dambӧck, 2018, p. 222).

Instead of evaluating a student's work according to some standard rating scale, traditional tests assess their ability in comparison with others. As Cheng and Fox (2017) explain, "For assessment to be effective and to enhance, not harm, students' learning, students must compete with themselves to continue to improve, and teachers should use assessment events to help students to develop effective learning strategies" (p. 188). This alternative is beneficial to them for more enduring and memorable learning in the classroom and real-life situations outside the academic center, making them prepared to be integrated into the 21st-century world. In short, effective assessment should engage students in showing rather than

telling what they learn, in explaining how they reach the answer, in showing their depth of understanding, and in making something new from what they know. Conventional tests, to a great extent, fail to do so.

Another major dissatisfaction hinges on the argument against the traditional view that conventional tests can universally and uniformly measure all individuals and all abilities; that is to say, a "one-size-fits-all" testing mindset. The traditional testing system ignores individual differences, including students' diverse backgrounds, learning styles, abilities, and needs. Thus, it may not effectively capture these differences. Critics argue for the use of multiple measures of assessment that encompass a variety of methods and perspectives, providing a more comprehensive and nuanced understanding of student learning and achievement. They also point out that traditional tests may perpetuate inequality and bias, as they may be culturally biased or disadvantage students from underprivileged backgrounds who lack access to resources for test preparation. The emphasis on grades in traditional testing systems can lead to a narrow focus on performance outcomes rather than fostering a love of learning or intrinsic motivation to explore and understand new concepts.

Addressing these criticisms has led to calls for alternative assessment methods, including PBA, that may offer more comprehensive and meaningful evaluations of students' existing knowledge and actual performance. Today, the significance of language assessment is not just limited to its role as an instrument for measuring learning. Instead of just auditing instruction (sorting and selecting students or justifying a grade), the assessment system must inform language teaching and promote language learning. A single testing form is not sufficient to measure complex subject matters in units, degrading language assessment to a numerical "snapshot" of student achievement. Using several types of assessments gives a more comprehensive appreciation of student learning. What is assessed, how it is assessed and evaluated, and how the results are reported all send out a strong signal to students about what they should learn, how they should learn it, what elements of quality are most important, and how well they are expected to perform. Thus, since the turn of the century or even earlier in the late 1990s, there has been a momentous movement from traditional testing toward language assessment in general and alternatives in assessment in particular.

Translation Evaluation in Academic Context: Current Situation

At present, the dominant trend for evaluating translation quality in most translation programs in academic settings follows the principles of Classical True Score Measurement Theory employing a traditional testing system, which relies heavily on inauthentic, summative, competence-measuring, decontextualized tests. The heavy weight given to these conventional tests and the excessive emphasis on their results, especially in translation programs, have terrible consequences. Translation teachers carelessly replace any awareness of the context complexity of the evaluative practices with the authority of their personal stances and opinions (Heidari Tabrizi, 2021a, 2022a, Heidari Tabrizi & Chalak, 2021). Students feel that the evaluation of their translations is done on the basis of arbitrary, subjective practices. In most cases, translation students do not know by what criteria their work will be evaluated. They spend most of their energy adapting themselves to the personal criteria of their teachers and feel that it is a waste of time to gain insights into the nature of translation processes as provided by translation theories. Consequently, they lack the self-awareness as well as the self-confidence they need to carry out translation tasks when they are on their own in the real — and sometimes confusing — world of translations. One piece of evidence for this assertion can be the frequent negative feedback teachers are likely to receive from the students about the final translation tests every semester. Still, another piece of supporting evidence is the countless anecdotes one hears in professional conferences about the deficiencies of translation tests.

Following the critical evaluation of Meylaerts and Marais (2023), the landscape of translation has evolved significantly in response to the rapid proliferation of hyper and multimodal objects dispersed widely across vast spatial and temporal dimensions, coupled with the diminishing dominance of literary written texts once considered the primary focus of translation. Translation is no longer regarded as the mere art of conveying meaning from one language to another; translators are not treated as language mediators either. In fact, the evolution has led to a re-evaluation of traditional definitions and approaches to translation, and there is a need for more expanded conceptualizations of translation to capture a broader spectrum of activities and explore the intricate dynamics of translation in its full richness and diversity. Translation is increasingly understood as a complex and unpredictable process rather than merely the creation of a finished product. It encompasses a wide array of modalities beyond written texts, including visual, auditory, gestural, and interactive forms of communication. This recognition challenges the bias towards a certain communication mode (i. e., focusing solely on the 'written' form of language) that has historically constrained the study of translation.

In this expanded understanding, translation becomes a dynamic and multifaceted process that engages with the complexities of culture, communication, and meaning-making across diverse contexts. It involves negotiation, adaptation, and creative interpretation rather than mere replication or transfer of content from one language to another. Embracing this complexity allows for a more nuanced and inclusive approach to translation that acknowledges its role in shaping and mediating our interconnected global reality. According to Meylaerts and Marais (2023), "Such expanded definitions consider translation not merely as a research object but also as a (research) practice, a process constructing, (re)assembling, and (re)connecting the social' and as 'an inter-semiotic all-encompassing epistemological tool and ontological concept' which produces knowledge" (p. 1).

Nowadays, both scholars and practitioners acknowledge the complex, dynamic, and multifaceted nature of the translation as "a socio-psychological phenomenon" (Almanna & House, 2024, p. 1). By framing translation as "intercultural communication and socio-cognitive action," House (2014, 2015, 2024) argues that translation goes beyond the mere transfer of words from one language to another. Rather, it involves various social factors in navigating intercultural differences and engaging in cognitive processes to achieve accurate and meaningful cross-cultural communication. By treating translation "as an act of 're-contextualization'" (Almanna & House, 2024, p. 198), this perspective underscores the importance of considering both cultural and cognitive dimensions in translation theory and practice.

With the expansion of the definitions and conceptualizations, translation is a complex, multifaceted process; the main problem in evaluating students' academic translations is the fact that inauthentic, summative, competence-measuring, decontextualized standardized paper-and-pencil tests fail to work effectively at least when using alone as the only measuring instrument. Translation, by nature, is more an act of performance than mere competence; it is an ongoing cognitive complex activity that is sensitive to contexts of situation and culture. As a task-focused, project-centered activity, translation, is considered to be a rather an objective-driven, result-oriented process that generates a product meeting a series of covert or overt conditions. This suggests that the success of a translation depends more on the translator's ability to perform the act of translation, to interpret skillfully, and to express the text in the target language rather than solely relying on linguistic competence. It underscores the dynamic and interpretive nature of translation, highlighting the role of the translator as an active participant who must engage with the text and bring it to life in the target language. While linguistic competence is undoubtedly essential for accurate translation, the performative aspect of the craft is emphasized, where the translator's creativity, intuition, and understanding of the cultural context play crucial roles in producing a successful

translation. In essence, translating is not merely about transferring words from one language to another, but rather about capturing the essence of the original text and delivering it to the audience in a compelling and impactful manner, much like a performer on stage. In academic contexts, conventional translation tests alone cannot capture this knowledge-informed performance task.

Translation can be more conveniently approached within the characteristics of a non-linear interactive task cycle of three stages: pre-translation, in-translation (translating proper), and post-translation activities. During the pre-translation activities, students may do some preliminary research to become familiar with the text and its topic, the writer and their style, as well as gathering and ordering ideas about which translation strategy should be selected throughout the translation process in general and which translation technique can be employed for particular cases. They can also contemplate over the workflow that the task goes through from initiation to completion, making resources and mechanisms required for successful task accomplishment available. Focused reading and several rereading of the text and connecting it to its context are key steps in this stage.

The in-translation stage begins with translators producing an initial draft translation of the source text into the target language. This step includes all the activities done by a translator while doing the actual translation. It involves transferring the meaning and nuances of the original text into the target language as accurately as possible and considering the cultural and linguistic differences between the two languages. In this stage, translators, as well as translation students, should lay a solid foundation for further refinement and improvement in subsequent drafts.

After drafting the translation, translators are advised to take breaks periodically to rest their minds and prevent fatigue, which can impact the quality of their translations. In this stage, translators should empty their minds of all the deliberations and thoughts they have made hitherto and forget their translations, preparing themselves for the post-translation activities, which usually include reviewing, revising, editing, and proofreading. These can be done with a fresh look after a longer break, revisiting the translation, and approaching it with the same fresh perspective as if reading it for the first time. By checking for possible linguistic and translation mistakes and errors, (would be) translators can add to the quality of their translations (including accuracy, consistency, fluency, and clarity) before finalizing them for submission. More often than not, it is possible for them to seek feedback from peers, instructors, or experts to gain deeper insights and have suggestions in hand for refinements.

As the stages are non-linear and cyclic, students can jump forward or move backwards across the stages and sub-stages, too; the translation process involves a recurring sequence of actions or steps in completing translation tasks. The translation cycle does not strictly follow a linear progression from one stage to the next; translators can revisit earlier stages or skip ahead to later stages as they refine their workflow. In all these stages, like any professional translators in real-life situations, translation students have access to printed, and online sources and databases, can consult dictionaries or more knowledgeable people in the field, can ask for 'correct answers,' and may do the translation with the help or collaboration of others. Newmark (1988) is right to say that "there is no such thing as a perfect, ideal or 'correct' translation" (p. 6). To him, "Translation is for discussion… Nothing is purely objective or subjective- There are no cast-iron rules. Everything is more or less… there are no absolutes" (p. 21). As he explains further, "Translation is enjoyable as a process, not as a state. Only a state is perfect" (p. 225). If translation is open to discussion and negotiations, how is it possible for the students, who are left alone on their own and deprived of any discussion under the constraints of a test session, to produce a translation that is going to be treated as a 'perfect' product?

If formal translation programs and classes are to prepare students for their future profession as translators, they should stimulate workplaces and the conditions under which would-be translators should work effectively. In real-life situations, the translation projects can be completed within a reasonable amount of time in a couple of days or even weeks, not within a one or two-hour session allocated by the test developers. Hence, translators have enough time to go through different stages of the translation process several times by doing and re-doing their translation again and again. This allows them to apply lessons learned from previous attempts and potentially achieve better results. They can improve the quality of their final translation product and refine it through discussion, negotiation, and collaboration too. Again, it is not clear how standardized paper-and-pencil tests can take care of these issues in evaluating students' academic translations. These tests, by nature, require testees to act individually, autonomously, and independently and regard teamwork, collaboration, and consultation as fraud and cheating.

All these are absent in evaluating the process of students' translations in academic settings where teachers develop tests measuring students' competence rather than their actual performance. To increase the reliability of the measurement, they go through 'standardizing' these tests. The more standardized the teachers try to make a translation test, the less authentic and contextualized it will be. On the one hand, translation is characterized in nature by the aforementioned defining features and more; on the other hand, standardized tests enjoy certain characteristics that make them partially unfitting, if not totally inappropriate, for translation evaluation in translation classes. They fail to provide the whole actual translating performance of translation students as they are expected to act in real-life situations.

Research (Heidari Tabrizi, 2008, 2021a, 2021b, 2022a, 2022b; Yazdani et al., 2020, 2023) has showed how much discrepancy is found among the translation teachers as translation test designers. It revealed that in developing these tests, they do not fuse theory-driven principles and standards. However, rather they follow their intuition and personal experience, and the criteria they employ in scoring students' academic translations are highly subjective. Consequently, the quality and accountability of such tests are under serious question. Thus, the question remains as to the possible solutions that can be offered to escape from this complex situation. In other words, how can the present situation be improved and promoted? The next section deals with a cohort attempt to occupy the existing niche in academic translation evaluation. Accordingly, in the next section, performance-based assessment is introduced as one of the potential solutions that may help fill the gaps left by traditional testing.

Performance-based Assessment: Major Principles and Premises
According to Richards and Schmidt (2010), Performance-based Assessment (PBA) aims "to measure student learning based on how well the learner can perform on a practical, real-world task," and it is believed that PBA is "a better measure of learning than performance on traditional tests such as multiple-choice tests" (p. 428). PBA is characteristically intended to assess more complex abilities with tasks rather than traditional fixed-response tests. PBA employs authentic, contextualized tasks that engage students in the application of learned knowledge and skills in performing real-life tasks successfully within a meaningful, culturally relevant situation rather than involving them in test items that demote mental functioning to discrete, isolated, decontextualized abstract knowledge. Most recently, Farhady (2022) delineates that "any assessment that attempts to measure the produced language" (p. 59) may fall within the realm of PBA. As a socially situated activity, language assessment practices unavoidably happen in authentic, real-world contexts, which "is particularly the case with task- and performance-based assessment" (Wigglesworth & Frost, 2017, p. 130). In PBA, tasks are developed to assess productive,
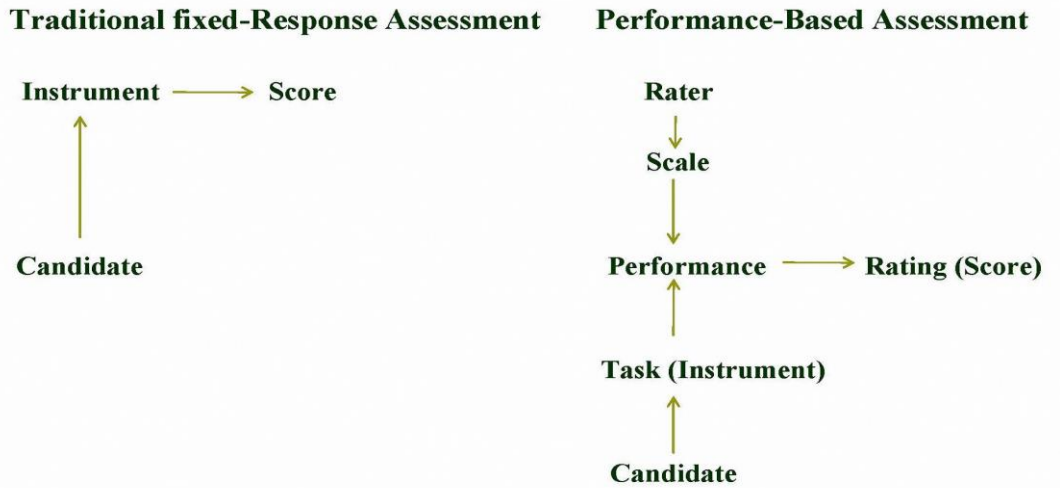
performance-oriented integrated skills essential for real-life situations or simulations. In brief, PBA is the process of involving students in challenging authentic tasks that necessitate various steps, which stimulates a wide range of active responses. By employing stimulus materials, the tasks are authentic (real-life approximations), motivating, challenging, direct, highly contextualized, humanistic, and learner-centered.

Traditional tests normally entail the process of objective measurement of a testee's more abstract demon of knowledge by means of a reliable instrument, typically a formal, discrete-point (standardized) paper-and-pencil test where testees are required to answer some questions 'correctly.' In contrast, PBA normally involves some human raters or scorers judging the quality of candidates' performance on some tasks by using an agreed-upon rating scheme, rubric, or scale. In other words, the degree of achievement is rated and scored by performance observation and professional judgment. As such, PBA entails three major components: tasks, rating scale, and human raters. Traditional tests, in contrast, are claimed to be scored more objectively due to their rather fixed-response nature. The majority of traditional tests are normally accompanied by relatively static scoring keys where the 'correct' answers are predetermined; thus, the end-all, be-all intention is eventually to obtain the right answer. As opposed to such a tradition, PBA practices do not necessarily involve clear-cut right or wrong answers. More often than not, there is more than one acceptable solution to performance-based tasks and projects. PBA tends to be more process-oriented procedures; what matters is not just the correct answer itself (the product) but rather the way students arrive at it (the process or solution). As an ongoing, more humanistic process, PBA displays what students learn not only at the end of the instruction but also, more importantly, throughout the process of learning or working on a problem.

In fact, PBA measures the extent to which an individual is successful or unsuccessful in performing a task or project. They may ask candidates to create their personally constructed responses to solve or overcome a problem, usually along with the process by which they approach the problem and solve it by explaining how they arrive at their solutions. Moreover, they are required to defend their choice or solution by explaining and justifying it. It can be concluded that while traditional testing is 'done' to a student, PBA is done by the student as an active participant. Figure 1 illustrates the similarities and dissimilarities between the two methods while highlighting the differences:

Figure 1
Traditional Testing vs. PBA (adopted from McNamara 1996, p.9)

Traditional fixed-Response Assessment

Instrument ⟶ Score

Candidate

Performance-Based Assessment

Rater

Scale

Performance ⟶ Rating (Score)

Task (Instrument)

Candidate

PBA aims to enhance student learning by assessing students on what they do in their daily class activities, promoting collaborative working and teamwork, and expecting students to present their work publicly. While traditional testing tradition merely fosters rote memorization of 'facts,' mechanical filling of the gaps, reproduction of models taught, and the like, PBA encourages the value of deep and reflective learning. Table 1 summarizes the defining characteristics of PBA as compared to those of traditional testing. It is mainly built on what Brown and Abeywickrama (2019) have concluded from the literature at hand. As they correctly notify, the content of the table must be treated with caution as it depicts "overgeneralized" notions. Moreover, the features are represented as if they are absolute, binary opposing sets of all-or-nothing nature, whereas, in reality, the border between the sets is fuzzy with no clear-cut line of distinction. Thus, at the top of Table 2, an arrow between two extremes of the spectrum (i.e., knowing and showing) has been added to help avoid such a misinterpretation and to indicate that these characteristics are of, rather more or less, continuum-based nature.

Table 1

Traditional Testing vs. PBA (based on Brown & Abeywickrama 2019, p.17)

Knowing ⟵⟶	Showing
Assessment driven by curriculum	Assessment driving curriculum
One-shot, standardized tests	Continuous, longitudinal
Speeded, timed, multiple-choice format	Untimed, free-response format
Decontextualized, inauthentic tests	Contextualized, authentic tasks
Scores sufficient for feedback	Individualized feedback and washback
Discrete-point, "right"/"best" answers	Open-ended, creative answers
Summative	Formative
Product-oriented	Process-oriented
Non-interactive performance	Interactive performance
Fostering extrinsic motivation	Fostering intrinsic motivation
Individual projects	Group projects

10

Norm-referenced scores	Criterion-referenced scores
Indirect Testing	Direct assessment

In a nutshell, PBA is claimed to have the following merits. It compensates for the negative aspects of traditional standardized tests, such as negative washback, test biasedness, and consequential validity. Moreover, it is more valid than conventional tests in predicting students' abilities in future real-world settings. It also documents and promotes creativity, critical thinking, and self-reflection. PBA decreases the so-called 'Lake Wobegon effect' where traditional testing tradition reinforces testees' illusory superiority and overstimulation. Overall, PBA provides a robust framework for integrating diversity, equity, and inclusion into translation teaching, learning, and assessment in translation programs across the world. By leveraging authentic contexts, culturally relevant content, differentiated assessment, critical reflection, community engagement, and personalized support, translation instructors can create inclusive learning environments where all students feel valued, respected, and empowered to succeed.

Just like other assessment approaches, PBA is not without its share of shortcomings and limitations. Like any other alternative assessment practice, it is fraught with certain challenges that persist, especially when it comes to implementing it in academic contexts. PBA typically faces challenges in terms of design and administration. According to Wigglesworth and Frost (2017), PBA is "one of the most expensive approaches to assessment, and in terms of development and delivery, one of the most complex" (p. 129). As for its practicality, it is quite exhausting in terms of time and energy that both teachers and students should spend on the rehearsal and for the actual performances. For teachers, it is also intensively laborious to judge and evaluate the quality of the learner's performance and grade. Performance assessments can be resource-intensive, requiring significant time and effort from both teachers and students. This can be challenging to scale up, especially when assessing large groups of individuals. Scaling such assessments can be challenging without compromising the quality of the evaluation.

Another noteworthy drawback of PBA that must be taken into account is related to the issue of generalizability. Despite the daunting workload imposed on its stakeholders, the outcomes tend to be less generalizable to other contexts as performance tasks, by nature, are precisely contextualized and complex. Another concern related to the problem of reduced transferability of the findings is that the performance judgments may not be replicable in other settings as tasks tend to be hard to replicate, keeping the measurement reliable and consistent.

DISCUSSION AND CONCLUSION

As a move forward, PBA can be introduced as one of the potential solutions that may help fill the gaps left by conventional testing tradition in translation evaluation in academic contexts and for pedagogical purposes. During the past two decades, PBA has slowly and steadily expanded and gained momentum, akin to a snowball rolling downhill (Salmani Nodoushan, 2008). Due to the growing prominence and applicability of assessing the quality of the performance of students in educational contexts (Wigglesworth & Frost, 2017), there is little doubt that PBA is quite suitable for evaluating student's academic translations. It has proved not to be a fad or fashion but a way to improve instructional practices in translation classrooms, too. Its expansive nature enables both teachers and students to envision a more defined path toward success and accomplishment. As a promising alternative to traditional testing tradition, PBA helps translation teachers to get deeper insights into what and how students learn and help them grow and become better learners.

The shift of attention to multiple assessment purposes in translation programs certainly needs a change in teachers' and students' learning mindsets and perspectives, too. Shifting away from the traditional testing system commonly used in the translation programs towards PBA practices for evaluating academic student translations is certainly a major theoretical as well as practical change, and change takes time. In so doing, translation teachers need to move slowly, keep their students informed, and negotiate with them the aims and justification behind the alternative assessment practices they begin to use. In moving toward alternative assessment practices, including PBA, translation teachers should give an ear to their students, discuss with them, and invite them to offer any feedback. Through such a process of negotiation and renegotiation, translation teachers may guarantee a safe transition from traditional testing towards alternative assessment practices such as PBA. While PBA has not fulfilled its potential as the ultimate solution, the so-called Promised Land, for assessing the actual performance of language learners and translation students, PBA has certainly laid the groundwork for advancements in this field. Whether such a proverbial 'land' truly exists remains uncertain, but these concerted efforts have at least opened up pathways toward its realization.

Utilizing PBA in translation classrooms and programs carries several significant implications. These implications may help improve the current situation of evaluating students' academic translations. As the importance and relevance of evaluating students' performance quality in educational settings continue to increase, a shift towards more direct, performance-based methods seems to be inevitable. Translation teachers should more often integrate task, rating scale, and rater characteristics into their evaluation model. They should also be more selective by leaving room for flexibility by giving priority to the criteria by value assigning; being flexible would assist them in navigating changes more efficiently. As translation is considered to be a socio-cognitive activity demanding high-order thinking skills, it is recommended to incorporate more cognitive, socio-pragmatic macro-structural elements into the evaluation model, too. Moreover, the testee's translation should never be treated as a ready-to-be-published piece of work.

By way of conclusion, translation teachers continue to grapple with the complex decision of whether to prioritize PBA or traditional tests. There is no 'best' or superior way to evaluate students' academic translation. Neither approach can be unequivocally labeled as the optimal choice, as each offers distinct advantages and is hindered by specific limitations and drawbacks. Therefore, it is essential to move between traditional and alternative assessment in a rather balanced way, utilizing both methods wisely and judiciously. It must be re-echoed that translation teachers should inevitably measure their students' mastery of knowledge and skills taught through employing conventional tests along with, or in most cases, even before, implementing PBA tasks. Nevertheless, it can safely be concluded that PBA suits better for evaluating translations of students in academic contexts because there is a relatively comprehensive congruence between the nature and defining characteristics of translation as a project-centered, task-focused, performance-oriented activity and that of PBA as an alternative method of assessment.

Because PBA contributes to fostering the feeling of satisfaction for both the teacher assessors conducting the evaluations and the translation students involved, it is hoped that more translation teachers will be encouraged to shift to a more performance-based evaluation of student translations. By implementing PBA practices, translation teachers can communicate to the translation students much better what constitutes excellence and how they can evaluate their own work. They are also encouraged to gain better insights into the nature of translation processes as informed by translation theories and research studies. PBA empowers students to gain the self-awareness as well as the self-confidence they

need to carry out translation tasks, especially when they are on their own in the real world since they have access to the scoring criteria. It also promotes critical, creative, and self-reflective thought in the students.

As for future directions, it can be concluded that the more teachers recognize the importance of personalized learning, the more PBA will be tailored to individual student needs. Moreover, as students engage with diverse forms of media and communication, PBA will incorporate multimodal approaches that allow students to demonstrate their understanding and skills through various formats, such as video presentations, digital storytelling, or multimedia projects. According to Fox (2017), "the array of alternative assessment approaches [such as PBA] will continue to be enhanced by technology" (p. 11) too. The technology-enhanced PBA using digital tools and platforms such as 3D simulations, virtual reality environments, and online e-portfolios will generate both new affordances and fresh challenges yet to be explored. The contributions made by the current work hopefully may open new avenues for further exploration in this decisive facet of research on translation studies, namely, academic translation evaluation.

References

Abdel Latif, M. M. M. (2020). *Translator and interpreter education research: Areas, methods, and trends*. Springer Nature.

Adab, B. (2000). Evaluating translation competence. In C. Schaffner & B. Adab (Eds.). *Developing translation competence* (pp. 215–228). John Benjamins.

Almanna, A., & House, J. (2024). *Linguistics for translators*. Routledge.

Arango-Keeth, F., & Koby, G. S. (2003). Assessing assessment: Translator training evaluation and the needs of industry quality assessment. In B. J. Baer (Ed.), *Beyond the ivory tower: Rethinking translation pedagogy* (pp. 117–134). John Benjamins.

Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. Heinle & Heinle.

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Bachman, L., & Damböck, B. (2018). *Language assessment for classroom teachers*. Oxford University Press.

Bowker, L. (2000). A corpus-based approach to evaluating student translations. *The Translator, 6*(2), 183–210.

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Longman.

Campbell, S., & Hale, S. (2003). Translation and interpreting assessment in the context of educational measurement. In G. M. Anderman & M. Roger (Eds.), *Translation today: Trends and perspectives* (pp. 205–224). Multilingual Matters.

Cheng, L., & Fox, J. (2017). *Assessment in the language classroom*. Palgrave.

Conde, T. (2013). Translation versus language errors in translation evaluation. *Assessment issues in language translation and interpreting*, pp. 97–112.

Dungan, N. (2013). Translation competence and the practices of translation quality assessment in Turkey. In D. Tsagari & R. Van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 131–144). Peter Lang AG.

Farhady, H. (2021). Learning-oriented assessment in virtual classroom contexts. *Journal of Language and Communication*, *8*(2), 121–132.

Farhady, H. (2022). Language testing and assessment in covid-19 pandemic crisis. In K. Sadeghi (Ed.), *Technology-assisted language assessment in diverse contexts: Lessons from the transition to online testing during COVID-19* (pp. 55–68). Routledge.

Fox, J. (2017). Using portfolios for assessment/alternative assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed.; pp. 135–147). Springer.

Hatim, B., & Mason, I. (1997). *The translator as communicator.* Routledge.

Heidari Tabrizi, H. (2008). *Towards developing a framework for the evaluation of Iranian undergraduate students' academic translation* (Unpublished Doctoral Thesis). Shiraz University, Shiraz, Iran.

Heidari Tabrizi, H. (2021a). Evaluative practices for assessing translation quality: A content analysis of Iranian undergraduate students' academic translations. *International Journal of Language Studies*, *15*(3), 65-88.

Heidari Tabrizi, H. (2021b). Pedagogical quality of English achievement tests: An untold story of Iranian high school students' oral scores. *International Journal of Language and Translation Research, 1*(1), pp.17-28.

Heidari Tabrizi, H. (2022a). Assessing quality of pedagogical translations: Dominant evaluative methods in the final tests of undergraduate translation courses. *Journal of Language and Translation*, *12*(3), 21-34.

Heidari Tabrizi, H. (2022b). Mapping out the terminology for judging quality in various translation practices: A key disciplinary desideratum. *International Journal of Language and Translation Research*, *2* (1), pp.1-21.

Heidari Tabrizi, H., & Chalak, A. (2021). Developing a comprehensive framework for evaluation of translated books as MA theses in translation studies in Iranian universities. *Journal of University Textbooks Research and Writing*, *25*(48), 73-87.

Heidari Tabrizi, H., Riazi, A. M., & Parhizgar, R. (2008). On the translation evaluation methods as practiced in Iranian universities' BA translation program: The attitude of students. *Teaching English Language and Literature* (*TELL), 2*(7), 71-87.

Holmes, J. S. (1988a/2000). The name and nature of translation studies. In L. Venuti (Ed.), *The Translation studies reader* (pp. 172–185). Routledge.

Honig, H. (1998). Positions, power, and practice: Functionalist approaches and translation quality assessment. In C. Schaffner (Ed.), *Translation and quality* (pp. 6-34). Multilingual Matters.

House, J. (2014). *Translation quality assessment: Past and present*. Routledge.

House, J. (2015). *Translation as communication across languages and cultures*. Routledge.

House, J. (2024). *Translation: The basics* (2nd ed.). Routledge.

Huertas-Barros, E., Vandepitte, S., & Iglesias-Fernández, E. (Eds.). (2019). *Quality assurance and assessment practices in translation and interpreting*. IGI Global.

McAlester, G. (2003). Comments in the 'Round-table discussion on translation in the New Millennium.' In G. M. Anderman, & M. Rogers, (Eds.). *Translation today: Trends and perspectives* (pp. 13–51). Multilingual Matters.

McNamara, T. F. (1996). *Measuring second language performance.* Longman.

Meylaerts, R. & Marais, K. (Eds.). (2023). *Routledge handbook of translation theory and concepts*. Routledge.

Munday, J., Pinto, S. R., & Blakesley, J. (2022). *Introducing translation studies: Theories & applications* (5th ed.). Routledge.

Newmark, P. (1988). *A textbook of translation*. Prentice hall.

Richards, J. C., & Schmidt, R. W. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Routledge.

Salmani Nodoushan, M. A. (2008). Performance assessment in language testing. *Online Submission*, *3*(4), 1-7.

Tsagari, D., & Van Deemter, R. (2013). *Assessment issues in language translation and interpreting*. Peter Lang AG.

Wigglesworth, G., & Frost, K. (2017). Task and performance-based assessment. In E. Shohamy, I. G. Or & S. May (Eds.), *Encyclopedia of language and education*, (Vol. *7*, pp. 121–133). Springer.

Yazdani, S., Heidari Tabrizi, H., & Chalak, A. (2020). Exploratory-cumulative vs. disputational talk on cognitive dependency of translation studies: Intermediate level students in focus. *International Journal of Foreign Language Teaching and Research, 8*(33), 39-57.

Yazdani, S., Heidari Tabrizi, H., & Chalak, A. (2023). Analyzing exploratory-cumulative talk discourse markers in translation classes: Covertly-needed vs. overtly-needed translation texts. *Journal of Language and Translation*, *13*(1), 15-26.

Biodata

Hossein Heidari Tabrizi is Professor of Applied Linguistics at the English Department at Islamic Azad University, Isfahan Branch, Iran since 1999 where he teaches undergraduate and graduate courses in TEFL and translation. He was the head of the Graduate School of English Department there from 2018 to 2022 and a visiting scholar at the English Department, Albert Ludwig University of Freiburg, Germany from 2023- to 2024. He is the founder and director-in-charge of *Research in English Language Pedagogy* and was selected as the top researcher of the English Department in 2016 and 2020. His research interests include Language Assessment, Translation Studies, and Critical Discourse Analysis.

Email: *heidaritabrizi @gmail.com*