

Construction and Validation of a Computerized Open-ended Bi-functional Translation Assessment System

Azam Akrami

MA. In Translation Studies, Department of English Language, Imam Reza International University, Mashhad, IRAN

Behzad Ghonsooly (Corresponding Author)

Professor in Applied Linguistics, Department of English Language, Ferdowsi University of Mashhad, Mashhad, IRAN

Mohammad Yazdani

Assistant Professor in Translation Studies, Department of English Language, Imam Reza International University, Mashhad, IRAN

Pouria Mohammad Alami

M.Sc. In Computer Engineering-Software, Department of Computer Engineering, Imam Reza International University, Mashhad, IRAN

ABSTRACT

Within the educational sphere, instruction and assessment have been acknowledged as two distinct realms which must be brought into closer nexus (Lantolf and Poehner 2004). In the translation educational setting, apart from the distinction between instruction and assessment, the apparent lack of an effective model of translation quality assessment has led us to develop a translation assessment software, which covers both diagnostic and instructional functions. More specifically, following *Vygotsky's sociocultural theory* and *dynamic assessment* approaches, this study pursued the design of a computerized open-ended bi-functional translation assessment system (COBTAS) to, firstly, enhance translation trainees' performance during the procedure by providing pre-set instructional mediations to the students. Then, it was attempted to diagnose the examinees' actual and learning potential abilities in the specified constructs. Furthermore, the study set out to ensure validity and reliability of the newly-developed software. The results revealed that COBTAS is a reliable instrument for demonstrating translation trainees' actual as well as proximal

development and enhancing students' level of translation ability within the test context.

KEYWORDS: assessment software, dynamic assessment, mediation, interventionist dynamic assessment, translation assessment, translation instruction, zone of proximal development

1. Introduction

Translation quality assessment has received considerable attention within the academic sphere (Hatim and Mason 1997; House 1997; Zequan 2003; Secară 2005). Though the topic has been explored from a variety of angles and perspectives– e.g. assessment of published translations, assessment of professional translators' works, and assessment of translation trainees' competence (Melis 1997, quoted in Melis and Albir 2001: 273), to the best of the researchers' knowledge most scholarly work has concentrated on the assessment of published works. In contrast to most research carried out in this area, this study seeks to focus on developing and providing a computerized model of dynamic assessment of translation within a pedagogical context.

The fundamental role of translation assessment in translation training has been acknowledged by a number of translation educators and researchers (e.g. Angelelli 2009; Angelelli and Jacobson 2009; Zehnalová 2013). However, scholarly work in this area remain scarce. This somewhat restricts constructing a satisfactory translation test instrument applicable to a translation pedagogical context. Besides, factors like the subjective nature of translation (Bowker 2000; Williams 2009), the lack of generally accepted measurement criteria (Bassnett-McGuire 1991; Kim 2009), and the unavailability of a universally accepted definition of translation competence may also be influential in developing an objective translation test instrument. The restrictions constitute a serious challenge to translation trainers who need to objectively assess students' translation competence and provide reliable feedback (Kim 2009). So far, a number of researchers have attempted to provide systematic methods of translation evaluation to be applied in the educational setting (e.g. Kim 2009; De Sutter, Cappelle, De Clercq, Looock, and Plevoets 2018). However, "it remains difficult, if not impossible, to define objectively what a good translation is" (De Sutter et al. 2018).

Within the current educational sphere, teaching and assessment seem to be regarded as two distinct activities, and, translation is no exception. While instruction usually takes place in order to develop students' abilities in a specified area, assessment aims only to gather information about students' past performance. In contrast to the common approaches of assessment, *Dynamic Assessment (DA) research* integrates assessment and instruction into a unified activity, aiming at simultaneously assessing and enhancing students' development. Based on Vygotsky's sociocultural theory and the concept of *zone of proximal development (ZPD)*, DA makes use of *instructional mediations* (e.g. probing questions, prompts, hints) during the assessment process. Within this approach, the provision of mediation is essential for a comprehensive understanding of the examinee's actual and potential abilities as well as promoting students' development in the assessment procedure (Lantolf and Poehner 2004). Although DA has been widely researched within the fields of education and psychology, it has been neglected within translation pedagogy and assessment.

Consistent with other fields of education, the distinction between translation instruction and its assessment underlines the need for coalescing them into a unified activity. In light of the profound impact of assessment on translation instruction, using DA approaches, the present study sought to construct and validate a computer-based *open-ended*¹ translation test instrument to:

- firstly, provide a fine-grained diagnosis of examinees' development while generating information pertinent to subsequent translation instruction;
- and, then, advance translation students' translation performance during the assessment process.

2. Review of Related Literature

The process of designing a test instrument commonly starts with posing some fundamental questions such as: *what construct is to be measured?*; *how is the assessment instrument to be constructed?*; and *how are the results to be used?* (Cohen 1994). In our view, upon embarking to design any translation assessment instrument, the first fundamental question to be answered is *what is to be measured* – i.e. pinpointing a definition of translation

¹ A test in which the respondents are allowed "the formulation of any answer, rather than a selection from a set of possible answers" (Oxford Living Dictionaries 2017).

competence. The thorough description of translation competence and its sub-competences provides the guidance test developers need to “define in measurement terms what knowledge or ability the translator needs to have and what skills are required from a candidate to function as a qualified translator” (Angelelli 2009: 23).

Broadly speaking, the models that have been constructed to gauge translation competence fall under two general categories: straightforward models and complicated models. The former (e.g. Pym 2003; Faber 1998; American Translators Association [ATA] 2015; etc.) are too general to be employed in a pedagogical setting. As an example, in his problem solving model, Pym (2003) describes translation competence as “the ability to generate a series of more than one viable target (TT1, TT2,...,TTN) for a pertinent source text (ST), and the ability to select only one viable from this series, quickly and with justified confidence” (489). The model, however, does not offer a comprehensive definition of translation ability or competence. In more explicit models of translation ability, ability is perceived to be the sum of several translation sub-competences. Such a description of translation ability is adopted by the PACTE² model (see PACTE 2003).

The models belonging to the second category (e.g. PACTE 2003; Angelelli 2009; EMT 2009; 2017) are influenced by more recent communication-oriented approaches. These are deemed too complex covering a multiplicity of competences and sub-competences to be employed in practical situations, specifically in assessing students’ translation competence. As a case in point, PACTE’s translation-competence model is, recognized as one of the most widely used empirical translation-competence models. The model presents five translation inter-related sub-competences including bilingual sub-competence (knowledge of L1 and L2), extra-linguistic sub-competence (knowledge of the world), knowledge about translation (knowledge of process and method), instrumental sub-competence (knowledge of tools and databases), and strategic sub-competence (knowledge of how to solve translation problems), and psycho-physiological components (such as critical thinking ability) (PACTE 2003). Nevertheless, in practice, the diversity of the competences and sub-competences it measures and the inter-relations among the components make the model too complicated to be used as a framework for designing a translation curriculum or assessing a translation product (Ghonsooly 2017).

² Process of Acquisition of Translation Competence and Evaluation

The other well-known framework for translation competence was designed by European Master`s in Translation network and it is the so-called EMT “Wheel of Competence” (EMT 2009). The framework, which was recently redesigned and published by the network in 2017, allows various skills, competences, and sub-competences to be acquired by students studying master`s level translation programs (EMT 2017). Five main areas of competence included in this model are: 1) Language and culture, 2) translation, 3) technology, 4) personal and interpersonal, and 5) service provision, each encompassing various sub-competences. Despite the considerable merits of this multi-componential model, and the attempts made at revising it, the model is not comprehensive enough to cover the spectrum of all competences, skills and knowledge that translation graduates should acquire (EMT 2017). This highlights the complexity and relativity of the concept of TC, and that no single description can be offered for the concept of translation competence which would cover all relevant sub-competences and aspects. In other words, establishing the dimensions of translation competence is in itself subjective (Eyckmans, Anckaert and Segers 2009). Space constraints do not allow the exploration of further translation-competence models here. But, based on the literature reviewed, one could argue that the selection of an appropriate translation-competence model applicable to translation assessment within the academic sphere has chiefly been the prerogative of translation instructors and assessors (Ghonsooly 2017).

The next central issue which ought to be taken into special consideration in developing an assessment instrument is the *method of measurement or how the construct is to be evaluated*. Responding to this question has been a challenging and demanding task on both a theoretical and practical level. The absence of universally accepted measurement criteria might be the most important factor accounting for this challenge. Based on the relevant literature, two opposing theoretical approaches to TA emerge: subjective and objective methods. On the one hand, advocates of the subjective approach consider the assessment of translation ability as a holistic measurement which is performed on the basis of some general measurement criteria such as fluency, accuracy and naturalness (Bassnett-McGuire 1991; Honig 1998; Sager 1989). In a more objective assessment of translation ability, on the other hand, a set of assessment criteria is proposed to evaluate translations based on detailed analyses of linguistic errors which enable the translation assessor to rate a translation more clearly. This also gives the assessor the opportunity to provide detailed feedback to translation learners about the test outcomes. In this regard, a number of studies have been carried out (e.g. Angelelli 2009; Kim

2009; De Sutter et al. 2018) which have generally focused on providing more systematic and objective methods of translation assessment, applicable within a translation pedagogical setting. Kim (2009), for example, addresses the lack of objective criteria for assessing the translation competence of translation students in the academic settings and the challenge translation teachers face in evaluating students' translations or offering reliable, detailed and constructive feedback. She then discusses how meaning-oriented translation assessment criteria can be applied to provide more objective translation evaluation and detailed and constructive feedback on students' translations. The other model of translation assessment recently introduced is the one produced by De Sutter et al. (2018). De Sutter et al. propose a statistical translation assessment method and contend that the method helps translation teachers objectively measure the acceptability of students' translations and provide clear and specific feedback on translation quality (2018). In the study, the homogenous linguistic behavior of professional translators and professional writers is considered as the norm, against which student performance is measured based on patterns of deviance. This raises a number of fundamental questions concerning the feasibility and reliability of the approach within the academic sphere. More precisely, it is still unclear how specific linguistic features should be regarded as indicators of translation quality, and hence, how the method "should be incorporated in the feedback for students" (De Sutter, et al. 2018). Based on the aforesaid, the call for objective translation assessment, specifically in an academic setting, is becoming increasingly persistent.

Besides the abovementioned principal questions, the other fundamental issue that must be considered upon developing a test, is *how the results are used*. Admittedly, any attempt at designing a test and using its results remains null and void without the explanation of the test functions and purposes. Given that this study sets out to construct a computerized translation assessment instrument, covering *DA functions*, the relevant theoretical background is briefly discussed.

Vygotsky's Sociocultural Theory and Zone of Proximal Development (ZPD)

Generally speaking, sociocultural theory associated with Vygotsky (1978) explores how cognitive processes are derived from sociocultural interactions. Vygotsky introduced *Zone of Proximal Development (ZPD)* as the core concept of the sociocultural theory, which is described as the zone in which a learner is not capable of completing the given tasks

Azam Akrami, Behzad Ghonsooly, Mohammad Yazdani and Pouria Mohammad Alami, Construction and Validation of a Computerized Open-ended Bi-functional Translation Assessment System, 1 – 37.

independently and needs proper social interaction (in Vygotsky's words: mediation). In this regard, social interaction with a more experienced person (e.g. parent, teacher) is essential for the learner's cognitive skills' development (Vygotsky 1978). Although Vygotsky did not directly suggest the application of the concept of ZPD to assessment methods (Duvall 2008), he suggests that learners' use of mediation in ZPD helps assessors/teachers to "study the process of accomplishing a task by the aid of specific auxiliary means; thus" assessors/teachers "are also able to discover the inner structure and development of the higher psychological processes" (Vygotsky 1978: 74). Vygotsky's socio-cultural theory and its concept of ZPD are considered as the fundamental framework of Dynamic Assessment approaches, which are explored in the next section.

Dynamic Assessment

Vygotsky's sociocultural theory and its focal concept of ZPD have led to a plethora of research in education collectively referred to as Dynamic Assessment (DA). From a DA perspective, instruction and assessment have been distinguished as two separate realms in education which must be brought into closer nexus (Lantolf and Poehner 2004). Poehner and Lantolf contend this notion, stating that on the one hand, efficient instruction requires assessment, "because it must be sensitive to what the individual is capable of achieving when acting independently". On the other hand, "a complete assessment requires instruction, which follows from Vygotsky's argument that higher mental development finds its source in sociocultural activity rather than in the recesses of the brain" (2010, P. 316).

Integrating instruction into assessment, DA primarily seeks to enhance students' performance during the assessment process. That is, if the test objective is to modify learners' performance during the assessment procedure, the test is dynamic, otherwise, it is static (Sternberg and Grigorenko 2002). Crucially, Vygotsky contends that providing instructional mediation noticeably helps learners internalize the tasks they have learned and gain a certain level of independence (Vygotsky 1978).

Apart from improving learners' abilities during the assessment process, DA intends to provide an accurate diagnosis of learners' abilities within the specified knowledge area. Specifically, DA aims to comprehensively investigate learners' actual abilities as well their potential developments. It appears that the typical tests presently in use do not test learning potential,

but they commonly test what has been taught. Kozulin and Garb (2002) argue that this contradiction is because of the differentiation between the goal of students' assessment and the assessment means. That is, as the goal of assessment is chiefly to evaluate learners' learning abilities, the means employed to do so are commonly restricted to only identifying the learners' current performance level (Kozulin and Garb 2002). Vygotsky's theory firmly bridges the gap between *what has been learned*, known as the learners' zone of actual development (ZAD), and *the capacity to learn* (ZPD). From this standpoint, a thorough diagnosis can reveal learners' fully developed competencies as well as those still in the process of forming (Vygotsky 1998). It could be argued that assessing learning potential is considered the mounting objective of DA (Kozulin 1990; Kozulin and Garb 2002; 2004). Chiefly focusing on learners' perception, attention and meta-cognitive strategies, the quality of assessor-learner interaction is assumed to be a major concern in DA approaches.

Mediation in DA

In line with Vygotsky's notion of the particular significance of assessor/teacher-learner interaction for full diagnosis of learner's abilities and advancing his level of performance, to meet instructional as well as diagnostic objectives, DA includes *instructional mediation* in the assessment process. Mediation, from this perspective, is considered as external assistance in the form of leading questions, hints, prompts, or other directives offered to the learners in DA procedure. Thus, mediation does not concern direct assistance which gives information in a very explicit manner but rather guidance which provides indirect assistance through which students endeavour to expand their competencies and cognitive abilities.

DA Approaches

In accordance with the type of mediation offered, a heterogeneous range of approaches fall under the umbrella term DA. However, this paper is limited to exploring two more general and holistic approaches, associated with Lantolf and Poehner (2004): *interactionist* and *interventionist approaches*.

Within an interactionist DA approach "mediation is negotiated with the individual, which means it is continually adjusted in accordance with the learner's responsivity" (Poehner and Lantolf 2010 318). This qualitative DA approach focuses on instructional problems each

individual encounters during the assessment process. In other words, within this approach “the instruction takes centre stage” (ibid.). The fundamental principle of this approach is that it does not consider numeric scores but aims at diagnosing learners’ learning abilities during the process of development and to identify learners’ learning requirements (Sternberg and Grigorenko 2002).

In contrast to the interactionist approach, the interventionist approach is a method of assessment in which “a prefabricated and fixed set of clues and hints is determined in advance and offered to learners as they move through a test item by item” (Poehner and Lantolf 2010, 318). Largely influenced by the *graduated prompt approach* associated with Brown and his colleagues (Brown and Ferrara 1985; Campione and Brown 1990) and, based on the assumption that hints given to examinees must meet learners’ requirements, the pre-set mediations, within the interventionist approach, range from implicit to explicit. In cases where an examinee is able to grasp the implicit hints adequately, there is no need to provide more explicit prompts, and the testee will obtain a higher score than those requiring more explicit mediations.

The possibility of computer-based assessment in this approach provides the opportunity for simultaneous administration of a test to a large number of candidates and offers more thorough and accurate description of the learners’ abilities within the test context. However, despite its merits, the interventionist approach, when computer-mediated, suffers from “lack of fine-grained mediation attuned to the specific needs of individuals as these emerge during the course of the procedure” (Poehner and Lantolf 2010: 318). Owing to the constraints involved in developing computerized interventionist dynamic tests, specifically within Applied Linguistics, only a few attempts have been made in this area. Those interventionist tests which were developed within the area of Applied Linguistics were adopted in evaluating receptive skills – i.e. reading and listening (e.g. Pishghadam and Barabadi 2012; Poehner, Zhang, and Lu 2015); not productive ones – i.e. writing and speaking. These systematized tests were mainly conducted in multiple-choice formats where examinees were asked to select the right response and not to generate it. However, the productive nature of translation necessitates the evaluation of the examinees’ translation products, and, hence, learner-generated responses. In practice, the difficulties involved in constructing open-ended computerised translation tests (e.g. the need for constructing a corpus via gathering, sorting,

and storing the plausible translation versions of the same source item) have deterred translation researchers from developing such tests to a great extent.

Building upon the relative literature and theoretical frameworks, with special attention to DA research, this study intends to construct and validate an *interventionist oriented computerized open-ended bi-functional translation assessment system* (COBTAS) to achieve the following functions:

- enhancing the students' level of translation ability through offering instructional prompts during the procedure;
- providing an accurate, descriptive and statistical diagnosis of the students' translation competence.

3. Method

3.1. Participants

The 202 translation students participated in this project were chosen from different universities of Iran including Khayyam, Ferdowsi, and Imam Reza International Universities of Mashhad, and Islamic Azad University (Quchan Branch). Based on convenience sampling, they were selected from three different levels of:

- elementary (second-year BA translation students),
- intermediate (last-year BA translation students) , and
- upper-intermediate (MA translation students)³.

Out of 202 participants, 134 individuals participated in the primary pilot paper-based non-dynamic TA. The main purpose of the test administration was collecting and categorizing the possible translation versions for each translation question item to be employed as a corpus in the COBTAS. In order to provide a more comprehensive list of translation versions, the participants were asked to present more than one translation version for each source item in this phase. Then, 6 students were selected for the one-by-one interactionist dynamic administration of the test to determine the form of mediations most helpful to COBTAS. After the test construction, 5 volunteer students participated in the COBTAS pilot test so the

³ In the higher education system of Iran, the undergraduate course in the field of translation lasts about four years, including eight semesters, and the master's course is a two-year program.

researchers received helpful feedback on the constructed test instrument. Finally, 57 translation students took part in the final test of the software to ensure its validity and reliability. The participants were native speakers of Persian and their age varied from 20 to 46 years. 7 participants did not disclose their age. Considering their assumed unfamiliarity with the project and its main objectives, some instructions and clarifications were made to obtain satisfactory results.

3.2. Instruments

In total, three instruments were designed and applied in this study as follows:

- *a 34-item Paper-formed English-Persian Translation Assessment Tool (Non-Dynamic Approach)* which was chiefly designed and administered to provide constructive information for designing test content to be used for building a corpus in COBTAS.
- *a 20-item Paper-Formed English-Persian Translation Assessment Tool (DA interactionist approach)* which was designed and administered to provide a set of standardized mediations applicable in COBTAS.
- *Computerized Open-ended Bi-Functional Translation Assessment System (COBTAS)* designed firstly to enhance translation trainees` performance during the assessment procedure, providing pre-set instructional mediations to the students. It was also constructed to provide accurate diagnostic information of the examinees` actual and potential learning abilities in the specified constructs (the software designing process is elaborated in the procedure section).

3.3. Procedure

According to Dunne (2009: 187-189), software development comprises four overall fundamental phases: *analysis, design, implementation, and verification and validation*. In the first phase, users` requirements and software development restrictions are determined. Within the present study, an effort has been made to elicit the translation students` requirement specifications and the constraints involved in the software construction as fully as possible. The analysis phase is followed by the design process in which software developers attempt to define how the specified users` requirements may be met. At this stage, with considerable assistance from a computer programmer, a comprehensive blueprint of the program was

delineated in conformity with the users` requirements and the special objectives of the study. The following subsections posit the conditions to be met and steps to be taken for the development of the translation assessment software.

3.3.1. Test content Preparation

The absence of a standardized translation test and the restrictions involved in designing an open-ended computerized translation assessment instrument led the researchers to design the test content in different stages based on the following parameters:

- designing the translation question items building on the recurrent translation errors among translation trainees;
- limiting the source items from phrase to sentence level in order to control the range of possible translations for each item;
- constricting the source items to general texts in order to assess the students` general translation ability and not their translation competence in a particular field;
- assessing one or more aspects of translation competence through each item;
- designing the items to be presented progressively based on difficulty (from easy to difficult), in view of the instructional objectives of the test and different levels of participants;
- providing the reader with relevant immediate textual context for the source items in order that the examinee can decode the ST semantic and pragmatic components accurately.

During the first phase, 54 items were selected from a variety of available translation text books –*Translation and Translator Volume 3* (Rashidi 2015), and *Translation Advanced* (English Text 1) (Farahzad 2006). Subsequently, based on the above parameters and taking into consideration the accessibility and storage of the range of possible responses, the items were revised by the authors of this paper, who selected 34 items to be administered non-dynamically as a paper-based open-ended translation test in order to:

- elicit a range of possible responses to the individual items;
- rate the correct responses, in terms of their level of acceptability;
- gauge the level of difficulty of each item and the test overall, based on the frequency of the correct responses each item rendered;

- employ the collected data as a corpus in COBTAS.

Following some corrections and modifications, 20 items along with their relevant instructional prompts were prepared to be re-piloted, but this time in DA format. That is, the test was re-administered, following an interactionist approach of DA, in which an open-ended one-to-one assessor-learner interaction is organized without any constraints upon mediation (see Lantolf and Poehner 2004). The test was administered in paper form, as the researchers engaged dialectically with students during the test procedure to determine students' understanding of the test items and the type of prompts and feedback they need to be offered. Due to the considerable amount of time and energy required for both the test administration and the analysis, the test was piloted with a small number of participants. The interactive test procedure was recorded and transcribed to be employed as a source for drawing out and arranging a set of systematically reliable instructional prompts.

Drawing upon the information yielded from piloting the test following the interactionist approach, two prompts were prepared for each plausible translation error, moving from implicit to more explicit. The implicit prompt was drawn to allow learners to attempt to answer the item again, after having been provided with a few clues (e.g. *Can you provide a more adequate equivalent for the source item?*) to tackle the problem (See Figure 3). If the examinee's response is correct, offering a more explicit mediating prompt is not required.

Given that more than one conceivable translation error was anticipated for each item, the items' problematic aspects and their accompanying mediating prompts were placed in order of priority, based on error typology – i.e. major or minor. Namely, the items' problematic aspects were classified based on the gravity of the error occurring. The pre-set prompts corresponding to potential errors were offered progressively, in cases where more than one error occurred in the student's response.

With special emphasis on the fact that the new test was designed to assess the examinees' translation competence at production level, two extra special prompts were supplemented to the individual items, on the assumption that students would produce some specific types of responses. Students would receive the first prompt when they generated an overall valid answer, but their score was lower than the correctness threshold (90%). It is noteworthy that since each source item usually corresponds to more than one acceptable translational outcome

of different levels of adequacy, those translational outcomes with a value of correctness equal or more than 85% were considered as valid answers. It should be also noted that, based on the assessment of the translation experts who participated in this study, the threshold of 90% was selected to determine the highest scores needed to receive the prompts. More precisely, given that by receiving each prompt, the total score of the relevant item was reduced by 5%, it did not seem reasonable to offer a prompt for the responses that received a level of correctness over 90%. Indeed, in this case, the possibility that the examinee can upgrade her/his score seems to be very low.

As evidenced in Table 1, equivalents such as "رئیس اهل کشور دیگری است"⁴ for the sentence *My boss comes from another country* are adequate in terms of fluency and accuracy. However, there are some more preferable, natural equivalents - such as "رئیس خارجی است."⁵ in Persian.

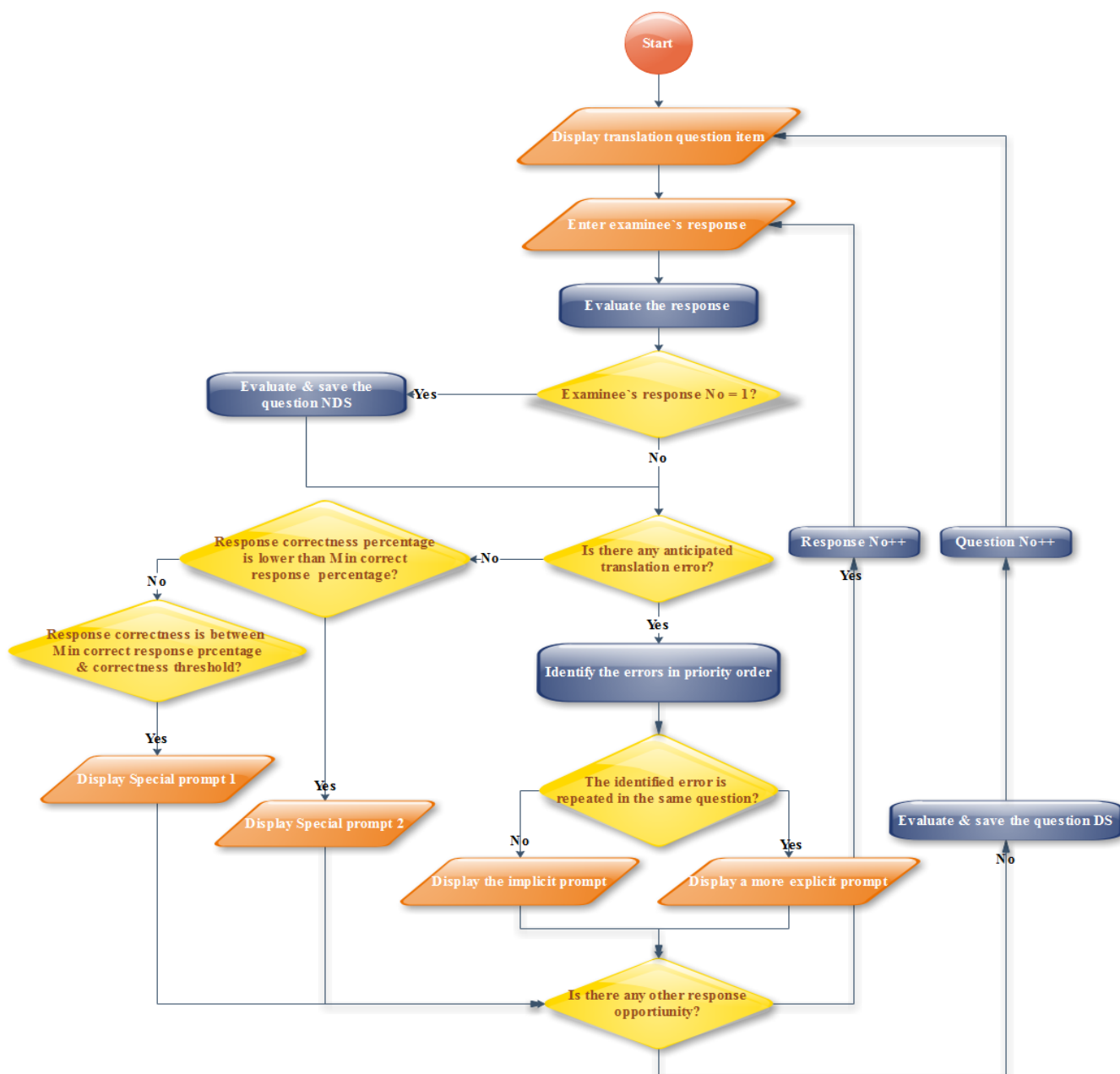
The point of assigning the relevant prompt (Figure 1: special prompt 1) in such cases is to give a chance to the test taker to enhance the level of correctness of his/her response (translation) before moving to the next item. The other prompt (Figure 1: special prompt 2) was prepared for incorrect responses that were not anticipated and are, therefore, not recognized by the software (e.g. part of the answer is left, or an irrelevant translation is generated). For these unpredictable translation errors, some general hints were designed (Table 1).

To provide a clear picture of the mediation process, a flowchart of the formulation of offering mediating prompts is represented in Figure 1. The abbreviations NDS (non-dynamic score) and DS (dynamic score) are explained later in the article.

⁴ My boss is from another country.

⁵ My boss is a foreigner.

Figure 1: A Flowchart of the Process of Providing Mediating Prompts



All of the information gathered during the aforementioned stages was employed as a database for designing COBTAS. In sum, the collected data consist of:

- 20 translation items arranged based on level of difficulty from easy to difficult;

Azam Akrami, Behzad Ghonsooly, Mohammad Yazdani and Pouria Mohammad Alami, Construction and Validation of a Computerized Open-ended Bi-functional Translation Assessment System, 1 – 37.

- The range of valid answers (correct translations) and their correctness percentage;
- Identification of the problematic aspects present in the test items, and their categorization under the umbrella terms *adequacy* and *fluency*;
- The relevant mediating prompts supplementing each problematic aspect (moving from implicit to more explicit);
- Arrangement of the problematic aspects and their accompanying prompts in order of priority for each item;
- The negative scores defined for each pre-determined plausible translation error in terms of the type of error and its effects on the adequacy or fluency of the response (translations);
- The number of response opportunities for each item (3-5) which was determined depending on the level of difficulty and the number of problematic aspects associated with each individual item.

Table 1: Sample Test Item, Responses, and Accompanying Prompts

Source item	Item's response (translation)	Correctness percentage	Prompt
<i>My boss comes from another country.</i>	رئیس خارجی است.	۱۰۰%	Null
<i>My boss comes from another country.</i>	رئیس اهل کشور دیگری است.	۸۵%	Special prompt 1: <i>The translation is acceptable. However, to enhance it, attempt to provide an equivalent more common in Persian.</i>
<i>My boss comes from another country.</i>	او رئیس	0%	Special prompt 2: <i>the translation is not accepted. Please translate with more precision and attention.</i>

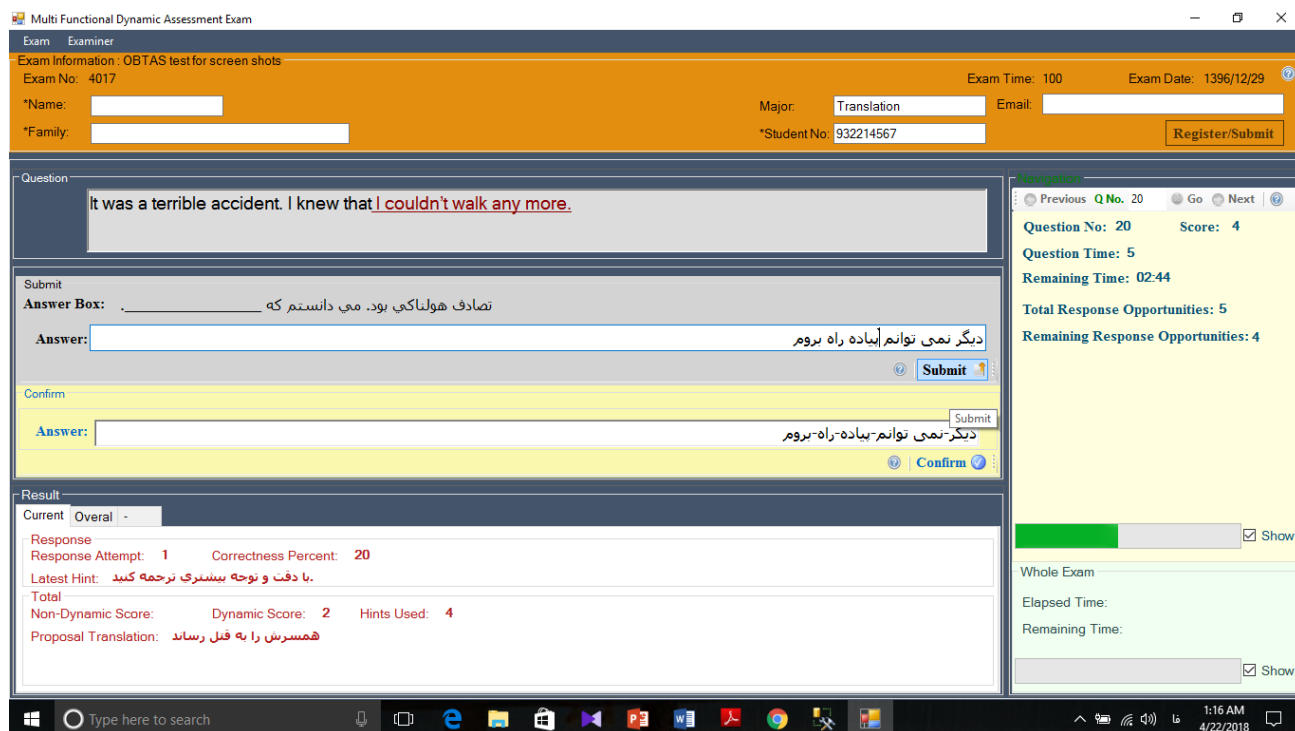


Figure 2: Screenshot of Submission of a Translation Question Item's First Response

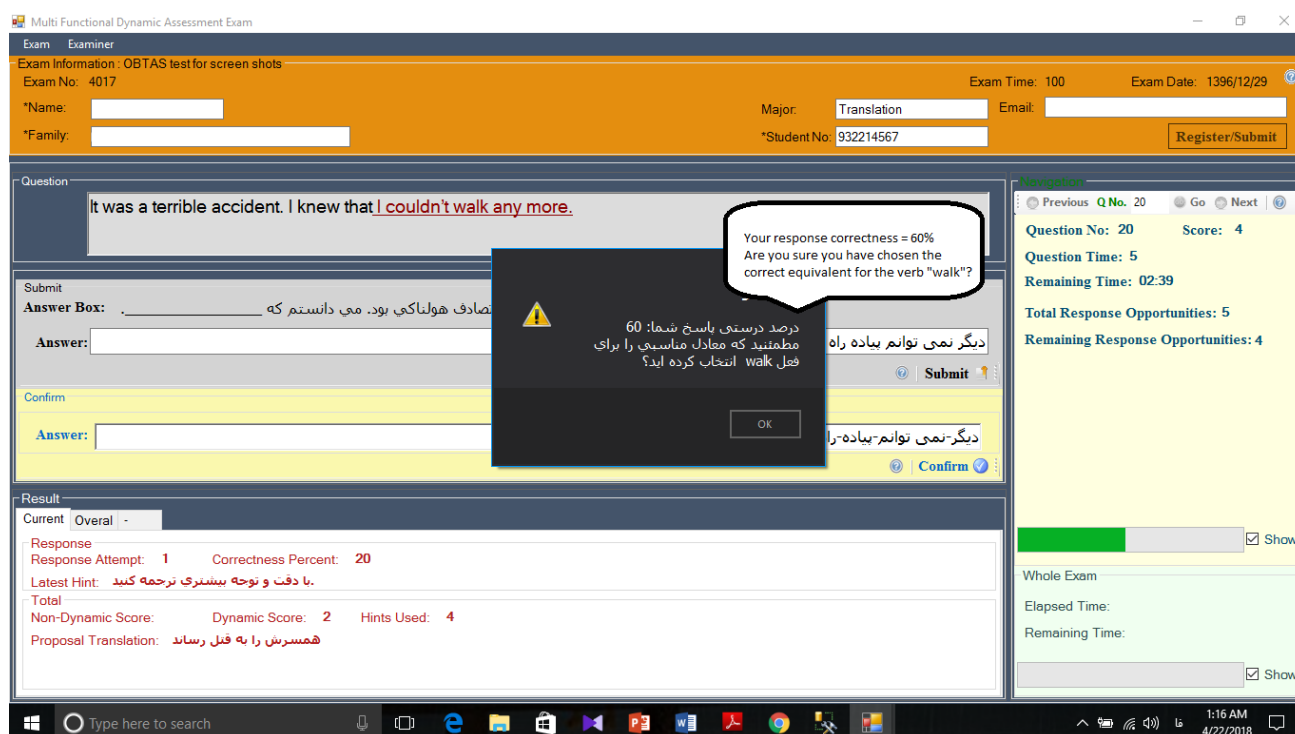


Figure 3: Screenshot of Providing an Implicit Mediating Prompt to the Learner

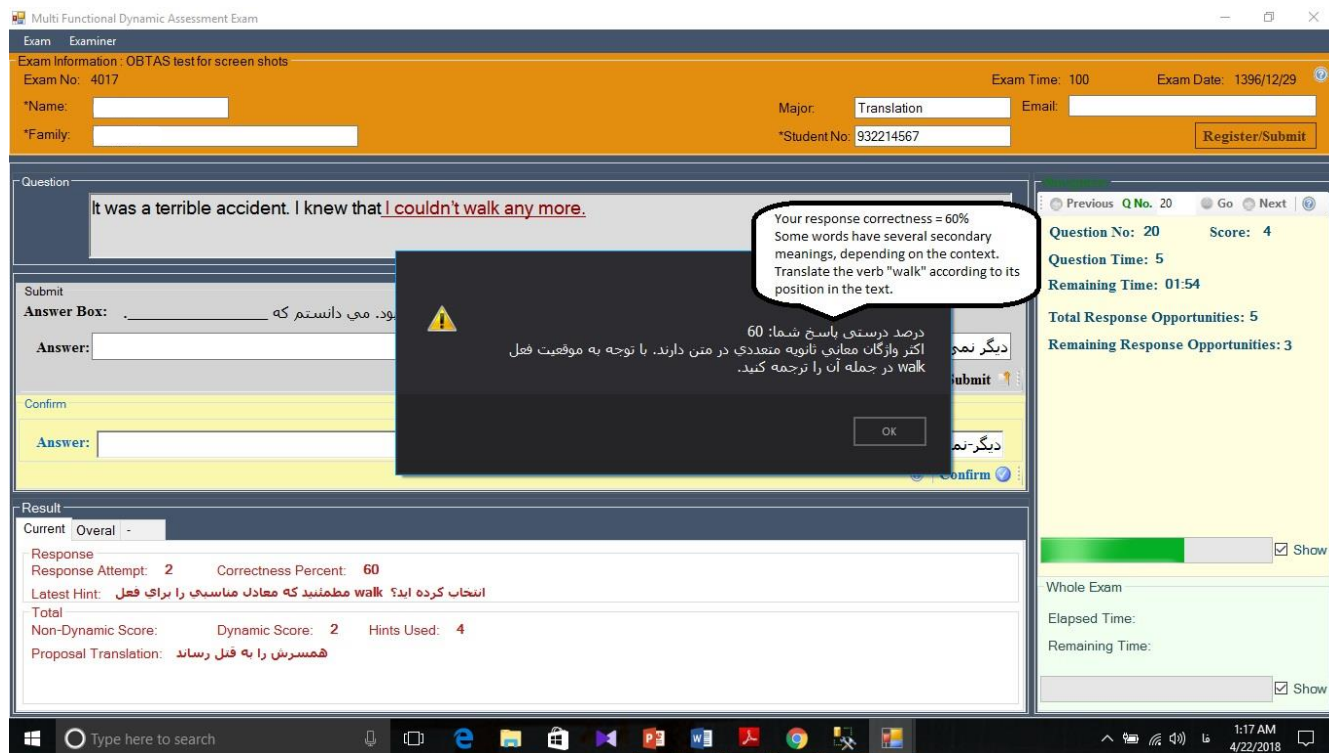


Figure 4: Screenshot of Providing a More Explicit Mediating Prompt to the Learner

3.3.2. Formulating the Process of Providing Mediating Prompts

For each anticipated incorrect response two relevant prompts can potentially be generated. If an examinee's first response to a translation task is incorrect (its correctness percentage is lower than the minimum correct response percentage set), then the first more implicit relevant prompt is displayed, depending on the number and type of translation errors as well as the priority of the related prompts. As a case in point, in Figure 2, the secondary meaning of the verb "work" is not recognized by the examinee, and so, s/he used its primary equivalent. An implicit prompt of "Are you sure you have chosen the correct equivalent for the verb "work"?" is then displayed (Figure 3), and the examinee is allowed another opportunity to respond. If the examinee's next response includes the same translation error, the second more explicit prompt is offered (see Figure 4). Based on whether the third response contains the same error or not, the first prompt of the next specified problematic aspect is displayed, provided that the occurring error is relevant. The process continues as such until either all the allocated response opportunities have been used, or the examinee has been able to translate the item with a higher score than the correctness threshold. Figure 1 clearly illustrates the

formulation of the process of providing mediating prompts during the assessment.

The information gathered from the design phase was, subsequently, submitted to the computer programmer to implement the *Design specification* in the form of a software program.

3.3.3. The Application of the Software

The opening page of the COBTAS provides examinees with a short description of the test and a set of simple instructions in English, as well as Persian. On the next page the examinees are required to register their personal information, including name, major, and student ID number. Then, by clicking the *register* icon, the first item appears along with its awarded total points, its allocated time, and the number of response opportunities available. The examinees enter their translations of the specified item in the allocated space. Confirming their first response, the examinees are then provided with the response correctness percentage, and the relevant mediating prompt, if needed. If the examinees do not succeed in producing the correct answer within the allocated time, or miss their assigned response opportunities, the software automatically moves to the next item, and the correctness percentage of the last response is displayed.

Upon completion of the test, a volume of information and a set of scores are displayed on the computer screen, including:

- *Total number of mediating prompts*
- *Non-dynamic score (N-DS)*

As representative of learners' independent performance/ static score, N-DS is computed, depending on the correctness percentage of the examinee's initial attempts to translate the items prior to any prompting and the total points awarded for each item (ranging from 3 to 5, based on the translation question items' level of difficulty).

- *Dynamic score (DS)*

DS is weighted in such a way that the total correctness percentage of the examinee's last attempt to translate each item is reduced by 5% for every prompt applied.

- *Gain score (GS)*

In accordance with DA research, in this assessment instrument, GS, which represents the observed changes in the students' performance in DA procedure, is weighted by subtracting N-DS from DS.

- *Examinees' profiles*

In order to represent a more accurate diagnosis of the examinees' performance during the procedure, upon the completion of the test, the examinees' individual profiles and their group profiles are generated. The profiles are designed based on the translation errors test takers have made in different translation constructs within the test context. Accessing examinees' profiles assists instructors to meticulously diagnose the translation difficulties each individual had in the marked categories/translation constructs and to identify the translation difficulties prevalent within the defined group of translation students, as well.

- *Learning potential score (LPS)*

To provide a more accurate quantitative measurement of examinees' learning potential, LPS, which was originally proposed by Kozulin and Garb (2002), is computed as a complementary score. Kozulin and Garb propose an LPS formula as follows:

$$\text{LPS} = (2 * \text{DS} - \text{N-DS}) / \text{MaxS}$$

MaxS= the maximum score gained in the test

Kozulin and Garb argue that the learners' ZPD and their learning potential cannot be fully understood relying only on the students' ultimate gain scores (GS) yielded from the DA process. They include maximum possible score of DA tests in the LPS calculation and argue that the presence of a maximum score in LPS considerably helps diagnose the students' learning potential. Despite the considerable merits of LPS, consistent with recently computerized DA (C-DA) research (Poehner et al. 2015), the study did not exclusively rely on examining LPS as the representative of the students' potential abilities. Rather, DS and GS

were examined as complementary to LPS so as to provide a more satisfactory diagnosis of students' abilities and the degree of instructional support they require.

In Kozulin and Garb's empirical study, students' LPSs ranged from 0.47 (suggesting very low level of learning potential) to 1.21 (demonstrating high learning potential). Assigning students to three different categories, depending on their LPSs, Kozulin and Garb contended that learners falling under different categories need different degrees of instructional support. In the present study, LPS of under 0.7 was considered low, from 0.7 to 0.99 was regarded as mid, and over 1 was regarded as high degree of learning potential.

3.3.4. COBTAS piloting

In order to get informative feedback on the test and implement some modifications in both the test content and the software package, COBTAS was piloted on 5 translation trainees. During this phase, the translation experts were asked to ensure the content validity⁶ of the test. Based on the feedback received on the software layout, some minor changes were also made in displaying the test items.

During the last phase, the newly designed software was debugged in terms of examining validity and verification, which will be further discussed in the result section.

4. Results

4.1. Descriptive Statistics of COBTAS

Table 2 demonstrates the descriptive statistics of the COBTAS results, including maximum score, mean N-DSs, DSs, and GSs. Considering this information, the distinction between the examinees' independent performance and their mediated performance is clearly illustrated. That is, the remarkable enhancement in the students' scores during the procedure (from 52.6 to 62.5) demonstrates the usefulness of the mediating prompts on the individuals' performance.

⁶ The extent to which an assessment reflects all facets of a specified construct

Table 2: Descriptive Statistics of COBTAS Pilot

	COBTAS results
Examinees` No.	57
No. of Items	20
Maximum score	76
Mean N-DS	52.6
Mean DS	62.5
Mean GS	10

DS: Dynamic Score, N-DS: Non-Dynamic Score

4.2. Ensuring Test Reliability

In order to evaluate the internal consistency of this new test, Cronbach Alpha Coefficient⁷ was applied. Cronbach Alpha estimated the internal reliability of the non-dynamic and dynamic tests at 0.92 and 0.90 respectively, suggesting very strong internal consistency of the COBTAS scale (see Table 3).

4.3. Ensuring Test Validity

The content validity of the newly-made test was established through a number of phases mentioned in the method section of the present study. In view of the fact that the higher progress in the learners` performance, the higher construct validity is expected to be for the relevant DA test (Poehner 2007; Pishghadam and Barabadi 2012), a paired samples *t* test was used to establish the construct validity⁸. The results of the *t* test between mean DSs and N-DSs (Table 4) clearly reveal a remarkable statistical difference between these two sets of scores ($t= 8.61$, $p<.0005$, Eta squared = .58), suggesting a statistically significant increase from N-DSs ($M = 52.7$, $SD =7.0$) to DSs ($M =62.05$, $SD=5.44$), denoting the remarkable benefits of mediations on the participants` performance. As previously mentioned, in computing DS, for each mediating prompt the individuals receive, they are penalized by a

⁷ The commonly used indicator of internal consistency ideally should be above 0.7 (Devellis 2003, cited in Pallant 2010: 97)

⁸ The degree to which a test measures what it claims (Brown, 2000)

reduction of 5% on the items' total score. Thus, for instance, using four mediations for an item, an individual receives a penalty of 20% reduction on the item's total score. The significant progress made in the students' scores, then, confirms the usefulness of the instructional mediations on the individuals' translation performance, at least within the test context (see also Appendix 2).

Table 3: Reliability Statistics of the Constructed Test

COBTAS test	Cronbach's Alpha	N of Items
N-DA	.92	20
DA	.90	20

N-DA: Non-Dynamic Assessment

4.4. *The Examination of COBTAS Considering DA Functions*

The strong positive correlation between NDSs and DSs substantiated the positive effect of mediation on the learners' performance ($r = .812, p < .0005$). In contrast, the correlation between N-DSs and GSs ($r = -.576, p < .0005$) indicated a drastically negative relationship between two sets of scores, denoting that learners with lower N-DSs generally benefited more from mediation than those with higher N-DSs (Table 5). The results are also in line with what was reported by the SLA DA research (e.g. Kozulin and Garb 2002; Poehner et al. 2015).

The findings also indicated that learners with similar N-DSs achieved different DSs, and GSs, denoting that COBTAS succeeded in distinguishing learners' initial performance and their potentiality for learning. Considering Table 6, examinees 8 and 9, for instance, both received the same N-DSs (38). However, with mediation, examinee 8 could achieve 10.5 points with DS (49.2) and examinee 9 gained 29.2 points with DS (68.1). This would suggest that examinees with similar initial performance might require a radically different level of subsequent instructional support.

The results also indicated that in order to provide a more accurate quantitative evaluation of learning potential, not only DS and GSs, but also the examinees' LPS should be taken into consideration. The findings demonstrated that in a number of cases, mediation did not help the examinees with high N-DSs (e.g. learners 56 and 57) gain as high scores as the examinees

with lower N-DS (e.g. learner 1). As might be expected, there might be no room for the examinees with high levels of independent performance to enhance their level of translation competence within the test context. With reference to Table 6, these examinees (learners 56 and 57) with low GSs still have high level of LPS (LPS=1). It can be contended that owing to LPS relation to maximum score, the examinees` low GSs were automatically disregarded in calculating their LPSs. Nevertheless, this does not mean that the examinees` progress in the test process and their responsiveness to mediation are undervalued in the LPS. As Table 6 demonstrates, examinees 9 and 10 both earned relatively low N-DS but then improved substantially with mediation, and consequently, achieved the topmost-range LPS of 1.3.

4.5. *Examinee Profiles*

Figure 6 and 7 present the individual profiles of examinees 32 and 33. Considering the similar scores these individuals received during the procedure (see Table 6), the examinees` profiles were selected and analysed to explore their translation problems, and to see whether there are any similarities and/or differences in the translation errors these participants made dealing with different translation constructs of the test. The results suggested that while examinee 33 was primarily struggling with the translation of pronouns with regard to fluency, examinee 32 had issues with translation of collocation with regard to adequacy. The profiles evidently illustrate that these examinees had different problems responding to each translation item, and by extension needed radically different instructional supports. Examining the group profile (Figure 5) also reveals that generally the participants ran into serious difficulties in the translation of collocations, particularly with regard to the level of fluency.

Table 4: Validity Statistics of the Constructed Test

Two types of test	Mean	N	t value	Std. Deviation	Sig. (2-tailed)	Effect size (Eta squared)
N-DA	52.70	57	8.61	7.00	.000	.58
DA	62.05	57		5.44		

Table 5: N-DS and its Correlation with DS and GS

N-DS and DS	N-DS and GS	Sig. (2-tailed)
.812**	-. 576**	.000

Table 6: Scores for Selected Translation Trainees in COBTAS Test

Student ID	N-DS	DS	GS	LPS
1	19.0	43.8	۲۴/۸	0.9
2	20.2	29.1	8.90	0.5
8	38.7	49.2	۱۰,۵	0.8
9	38.9	68.1	۲۹,۲	1.3
۱۰	40.1	71.0	30.90	1.3
2۱	51.7	70.0	18.3	1.2
2۲	51.8	57.1	۵,۳	0.8
3۲	57.8	71.9	14.10	1.1
3۳	58.1	70.4	۱۲,۳	1.1
56	72.1	74.8	2.70	1.0
57	74.4	75.2	.8	1.0

Figure 5: Group Profile of 57 Examinees Participating in CMTA Pilot test

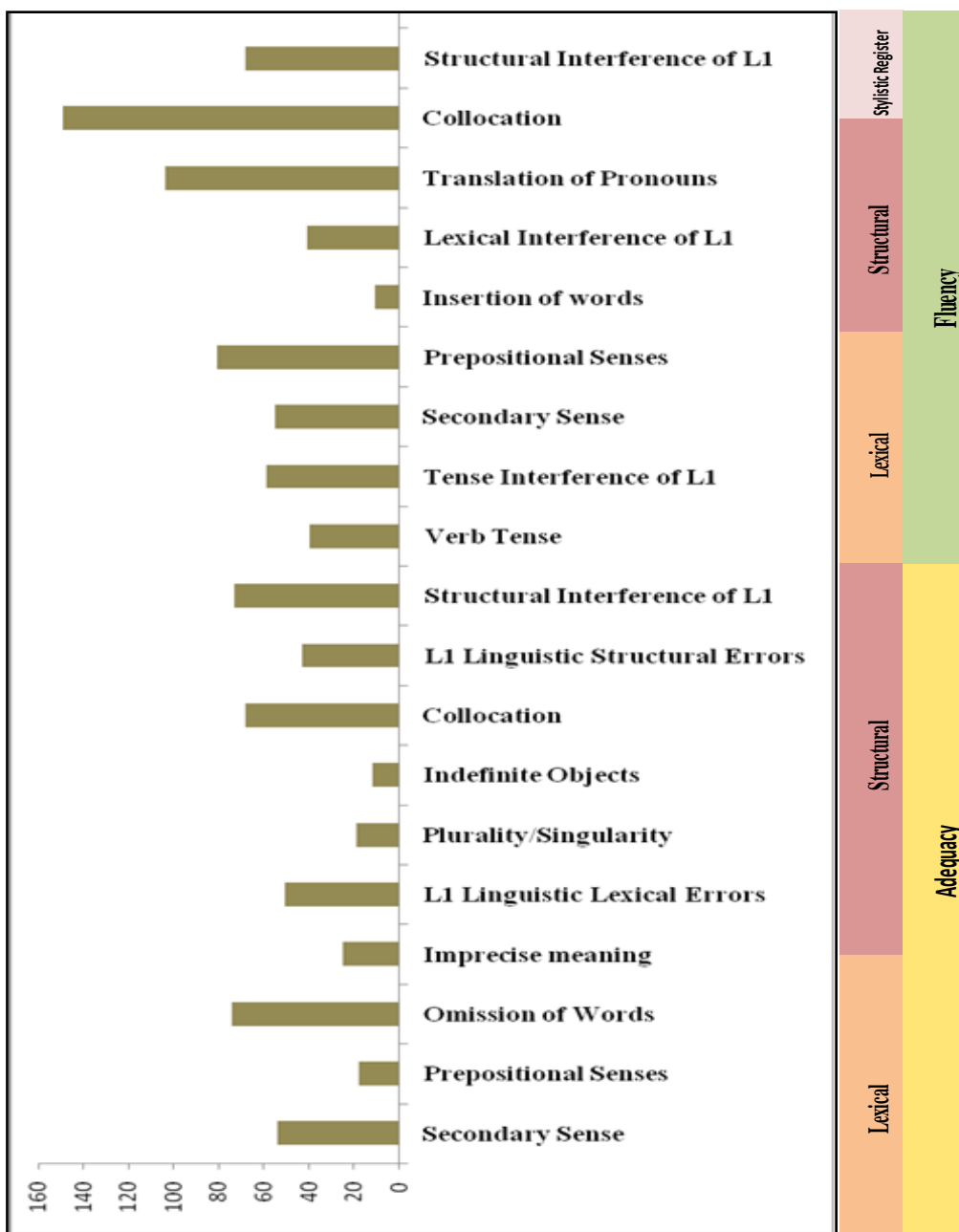


Figure 6: Individual Profile of Examinee 32

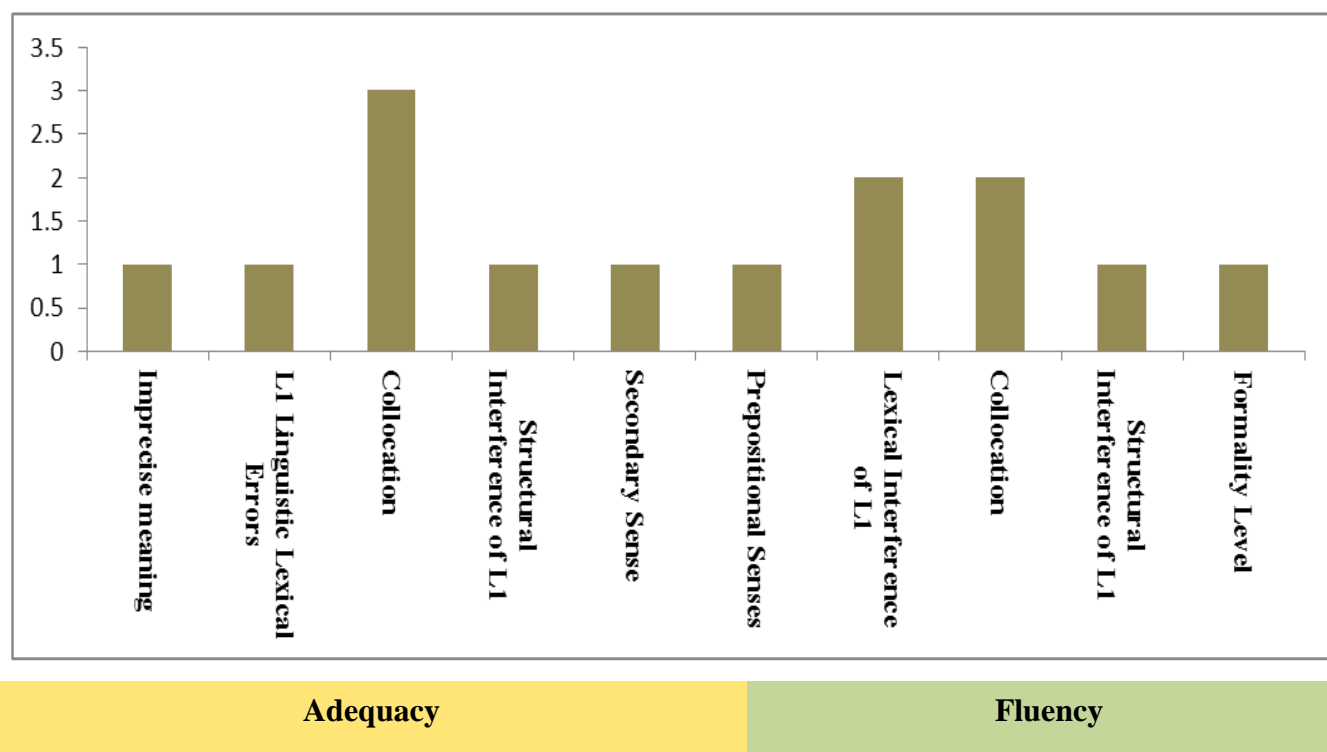
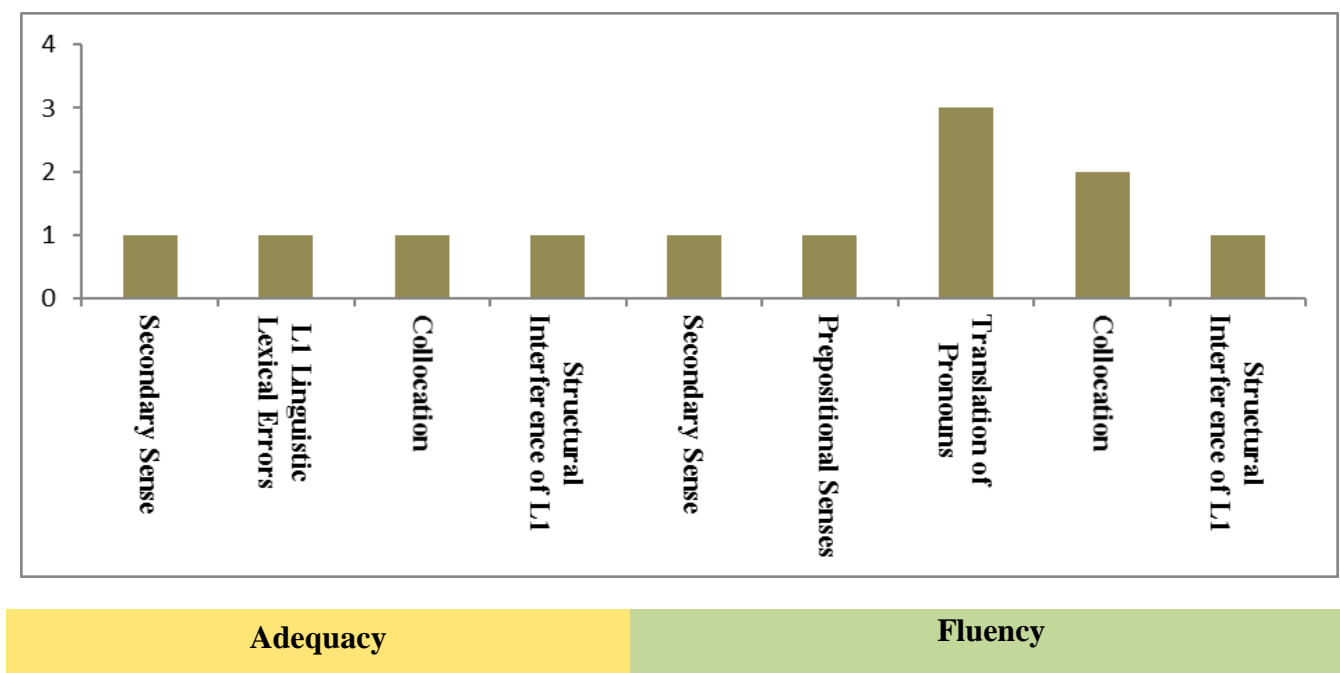


Figure 6: Individual Profile of Examinee 33



Discussion

The present study aimed at constructing and validating an open-ended translation assessment instrument following DA purposes (i.e. instructional purposes and diagnostic assessment purposes). The statistical analysis substantiates the validity and reliability of the constructed test. The high reliability of both N-DSs and DSs in our study demonstrates a great internal consistency of the scale. Consistent with what has been reported by Poehner et al (2015), the t test between DSs and N-DSs indicates a remarkable statistical improvement in the examinees' scores. Accordingly, the impact of mediation on the examinees' performance can be validated decisively.

A review of translation assessment literature indicates that almost all translation assessments in the educational settings are based on subjective evaluation, and, hence, cannot safeguard validity and reliability. In this project, the corpus generated and used for constructing the test instrument contributes to a more objective assessment of students' translation abilities and ensures test content validity. The existence of certain criteria and an automatic scoring process in the constructed translation test (COBTAS) also considerably minimize subjectivity and enhance the validity and reliability of the test.

The statistical evidence of the study demonstrates the effectiveness of COBTAS in improving the students' translation performance and diagnosing their translation abilities. From an instructional perspective, the strong positive correlation between N-DSs and DSs and the statistically significant improvement in the examinees' scores demonstrate the considerable enhancement in their translation ability within the test context. More precisely, despite the penalties triggered by the need for mediation, the overall mean scores of students' mediated-performance (DS) were remarkably higher than the mean scores yielded from their independent performance (N-DS). Therefore, it seems, students generally benefited from instructional mediations. Apparently, mediation in COBTAS assists learners identify the problems underlying poor translation performance and improve their translation abilities. At the same time, it may help examinees advance their cognitive abilities and arrive at a great level of independence. As Ellis (2000) states, the provision of mediation facilitates learning and helps examinees expand their cognitive capacities. The examinees then internalize the

task and are able to perform it independently (Ellis, 2000). Owing to space constraints, we could not expand our discussion of how learning becomes internalized in the COBTAS process any further. This is something that should be addressed in future research. Unlike traditional/static assessment which focuses only on gathering information about students' existing abilities, instruction in DA is an integral part of assessment. In Sternberg and Grigorenko's view, (2002), a test is dynamic only if it modifies learners' performance during assessment.

Our findings are consistent with previous DA research within Applied Linguistics (e.g. Kozulin and Garb 2002; 2004; Poehner et al. 2015). As an example, Kozulin and Garb (2004), succeeded in demonstrating students' L2 development in their work through the application of DA approaches and relevant L2 learning strategies. Similarly, in constructing online multiple-choice tests of L2 listening and reading, Poehner et al. (2015) demonstrated that the principle of mediation in L2 assessment can maximally help students to develop their listening and reading skills.

Within the translation teaching context, there is almost no record of using an interventionist DA approach that helps translation teaching practices. Common translation teaching practices are based on student production and teacher correction, the restrictions involved (e.g. time limitations) impede the possibility of one-to-one learner-teacher (mediator) interactions in translation classes. In such a situation, only a few translation versions produced by students receive teacher feedback. Unlike the current translation teaching system, utilizing an interventionist DA approach in COBTAS provides the possibility for simultaneous one-to-one learner-mediator interactions which considerably accelerates the learning process. In addition, offering, firstly, implicit prompts and, subsequently, more explicit prompts enhances learners' cognitive abilities and their level of translation performance.

In view of the crucial role of mediation in DA and Vygotsky's emphasis on the importance of the quality of mediation, the researchers attempted to design a set of applicable and appropriate prompts to be used in COBTAS. Designing the prompts based on analysing one-to-one interactionist test administration greatly assisted researchers to align the mediations with students' requirements and to ensure the usefulness and validity of the prompts' content. However, the question as to whether other types of instructional prompts might be more helpful in probing students' problems in translation C-DA requires further research. Overall,

by providing mediation, COBTAS not only improves the examinees' performance but also offers them comprehensive diagnostic information to the benefit of their subsequent instructional support.

Within the field of TA and in accordance with Vygotsky's theory and DA researchers' contentions, it is thought that conventional static tests reflect only some predetermined aspects of students' translation abilities. However, DA provides the opportunity to determine students' understanding and potential capacity in translation learning. In the present study, while the students' initial response to each item (N-DS) represents the students' actual translation abilities or static test results, analysing DA results (DS, GS, and LPS) reflects the students' potential capacities and progress. The researchers found statistically significant evidence that interpreting traditional translation tests' results cannot fully reflect the translation trainees' abilities. Despite a similar initial performance, some examinees responded differently to mediation and accordingly gained different DSs, GSs, and LPSs (e.g. learners 8, and 9). Thus, it seems that different subsequent instructional support is required for such cases. This suggests that assessing students' abilities based solely on the results obtained from traditional tests can be misleading (Anton 2009).

In conformity with the recent SLA C-DA research (Poehner et al. 2015), our findings proved that we cannot exclusively rely on reporting any single score yielded by the DA procedure for diagnosing students' potential abilities. While DS reflects the overall performance of learners after treatment (receiving mediations), both GS and LPS provide the degree of observed changes in the learners' performance. That is, the students whose GSs are substantial, commonly achieve high-range LPS as well (e.g. learners 9 and 10). However, in conformity with what has been reported by Kozulin and Garb (2002; 2004), the findings indicated that learners with high N-DS who could not improve their scores substantially during the DA process and accordingly received low GS might still have high LPS (learners 56 and 57). It can be argued that both examinees' independent level of performance (N-DS) as well as their responsiveness to mediation are taken into account in LPS calculation. Nonetheless, the question as to whether LPS can be considered as the fixed attribute of learning potential and whether it is efficient for classing students in different groups for subsequent instructional support remain to be addressed in future research. Based on the aforementioned accounts, our findings support the idea of the efficacy of the statistical information generated by COBTAS

in diagnosing overall capabilities of the examinees and distinguishing their *zone of proximal development* (potential abilities) from their *zone of actual development* (existent actual abilities).

To elicit more detailed information about the examinees' abilities within various translation constructs addressed by the test, COBTAS generated the students' *individual* and *group profiles*, as a supplement to statistical information. The concept of utilizing learners' profiles in COBTAS was grounded on research conducted by Feuerstein et al (2010). Introducing learners' profiles, they provided a thorough description of the difficulties students encountered during the tasks, particular mediations they required and their responsiveness to external mediations. The analysis of individual profiles suggested that examinees with nearly equal scores performed differently to the various translation items within the test context (See Figure 6, and 7), and therefore, often needed different types of translation instructional support. Thus, analysing individual profiles, translation instructors can be provided with nuanced information regarding learners' translation errors and the amount of effort they need to progress. But, the question is how useful individual profiles are in designing curricula or planning future teaching for students who want to participate in COBTAS. With an aim to applying the individuals' diagnostic information in designing subsequent translation curricula, COBTAS generated the group profile (Figure 5). Analysing this profile provides the possibility for assessors to diagnose the frequency and kind of particular translation errors occurring among the group of participants. Scrutinizing both group and individual profiles can be a great help in designing thorough translation curricula and teaching plans.

A distinguishing feature of the study, in our view, was the systematic evaluation of students' *productive skills* through the design of an *open-ended* computerized translation test. This property marks a clear departure from other Applied Linguistics C-DA research which chiefly assesses learners' receptive skills (e.g. Poehner et al. 2015). Apparently, employing selective methods of assessment (e.g. multiple-choice methods) can be appropriate for evaluating students' receptive skills (e.g. reading). As it was argued, however, the productive nature of translation requires translation assessment to move beyond the scope of assessing students' perception of appropriate translation and thus, necessitates evaluating the examinees' generated translations. Designing COBTAS as an *open-ended* computerized translation test is expected to help instructors more accurately evaluate students' translation abilities. In

addition, this quality allows for better understanding of any change in the students' performance through the instructional supports they received during the DA procedure. Consequently, not only the examinees' translation products, but also the process of producing translation, and the strategies they use in response to mediations are comprehensively analysed. In COBTAS, these changes in students' translation performance were indicated via both statistical and descriptive information presented. However, considering the plethora of responses offered for each item, a huge volume of data is yielded from administering such tests which needs more detailed analysis. Applying novel scientific methods such as *data mining*⁹ may be useful for such future research. We note that employing an open-ended method of assessment in COBTAS itself involved a number of restrictions in terms of selection of genre and length of source items. Using more advanced models and algorithms (e.g. *machine learning models*) may help test developers to considerably reduce these limitations.

5. Conclusion

Building upon Vygotsky's theory and DA research, this study attempted to construct a translation assessment software (COBTAS) for instruction and diagnostic assessment purposes. By designing COBTAS, we endeavoured to expand our understanding of *translation assessment* and *translation teaching* as interrelated. Taken together, the evidence in this project substantiated that including instructional mediations in the present translation test considerably assisted in enhancing students' translation abilities and providing accurate diagnosis of their translation competence as well.

This project highlighted the insufficiency of traditional static translation test results for interpreting students' translation abilities. Apparently, traditional assessment is based on the notion that measuring exclusively students' past development (ZAD) will indicate their subsequent progress and learning requirements. Vygotsky (1978) rejected this idea, stating that a learner's independent performance only reveals a partial picture of their capacity, and for a better understanding of a learner's latent ability we need to explore their ZPD. In this project, combining the individuals' DA scores (DS, GS, and LPS) with static assessment scores (N-DS), and supplementing the students' profiles with statistical information seems to

⁹ It is an interdisciplinary subfield of computer science which is defined as the process of extracting useful patterns in a large volume of dataset (Chakrabarti, Ester, Fayyad, Gehrke, Han, Morishita, & Wang, 2006).

provide a clear picture of students' actual and potential translation abilities. However, to shed light onto the usability of the information generated by COBTAS, it appears that a more systematic investigation (e.g. interview with students and teachers) is required.

This project is novel in that it applies C-DA approaches in the translator training sphere. Therefore, the existence of deficiencies in designing such a test is inevitable. It is hoped that this study will pave the way for researchers in Translation Studies to investigate this challenging area of research and to create more advanced generations of such translation C-DA instruments.

Email address: azi.akrami@gmail.com

6. References

- American Translators Association (ATA) Certification Program (2015).
(http://www.atanet.org/certification/aboutcert_overview.php)
- Angelelli, Claudia V. (2009) 'Using a Rubric to Assess Translation Ability' in C. Angelelli, and H. Jacobson (eds), *Testing and Assessment in Translation and Interpreting Studies*, John Benjamins: Philadelphia, 13-47.
- Angelelli, Claudia V., and Holly E. Jacobson (eds) (2009) *Testing and Assessment in Translation and Interpreting Studies: A call for dialogue between research and practice*, John Benjamins Publishing.
- Anton, Marta (2009) 'Dynamic assessment of advanced second language learners', *Foreign Language Annals*, 42(3): 576-598.
- Bassnett-McGuire, Susan (1991) *Translation Studies* (Revised edition), London: Routledge.
- Bowker, Lynne (2000) 'A corpus-based approach to evaluating student translations' *The Translator*, 6(2): 183-210.
- Brown, A. L., and R. A. Ferrara (1985) 'Diagnosing zones of proximal development' in J. Wertsch (ed.), *Culture, Communication and Cognition: Vygotskian perspectives*, Cambridge: Cambridge University Press, 273-305.
- Brown, J. D. (2000) 'What is construct validity', *JALT Testing & Evaluation SIG Newsletter*, 4(2).
- Campione, Joseph C., and Ann L. Brown (1990) 'Guided Learning and Transfer: Implications
- Azam Akrami, Behzad Ghonsooly, Mohammad Yazdani and Pouria Mohammad Alami, Construction and Validation of a Computerized Open-ended Bi-functional Translation Assessment System, 1 – 37.*

- for Approaches to Assessment' in Frederiksen, R. Glaser, A. Lesgold and M. G. Shafto (eds), *Diagnostic Monitoring of Skill and Knowledge Acquisition*, Hillsdale, NJ: Lawrence Erlbaum, 141-172.
- Chakrabarti, S., Ester, M., Fayyad, U., Gehrke, J., Han, J., Morishita, S., ... & Wang, W. (2006). Data mining curriculum: A proposal (Version 1.0). *Intensive Working Group of ACM SIGKDD Curriculum Committee*, 140.
- Cohen, Andrew. (1994) 'Assessing Language Ability in the Classroom', 2nd edition. Boston.
- De Sutter, G., Cappelle, B., De Clercq, O., Loock, R., & Plevoets, K. (2018) 'Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translations', *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, 16.
- DeVellis, R. F. (2003) *Scale Development: Theory and applications*, Thousand Oaks, California: Sage.
- Dunne, Keiran (2009) 'Assessing software localization for a valid approach' in C. Angelelli, and H. Jacobson (eds), *Testing and Assessment in Translation and Interpreting Studies: A call for dialogue between research and practice*, Amsterdam: John Benjamins Publishing Company, 13-47.
- Duvall, E. D. (2008) *No Secrets To Conceal: Dynamic assessment and a state mandated, standardized 3rd grade reading test for children with learning disabilities*, The Pennsylvania State University.
- Ellis, Rod (2000) 'Task-based research and language pedagogy', *Language Teaching Research* 4(3):193-220.
- EMT (2009) *Competences for Professional Translators, Experts in Multilingual and Multimedia Communication*.
- EMT (2017) https://ec.europa.eu/info/sites/info/files/emt_competence_fw_k_2017_en_web.pdf (accessed: 14 April 2018)
- Eyckmans, J., Anckaert, P., & Segers, W. (2009) 'The perks of norm-referenced translation evaluation', *Testing and Assessment in Translation and Interpreting Studies*, 73-93.
- Faber, Pamela (1998) 'Translation Competence and Language Awareness', *Language Awareness*, 7(1):9-21
- Farahzad, Farzane (2006) *Tarjomeye pishrafte 1* [Translation advanced 1], Tehran: Payame-e-Noor Press.
- Feuerstein, Reuven, Refael Feuerstein, and Louis H. Falik (2010) *Beyond Smarter: Mediated*
- Azam Akrami, Behzad Ghonsooly, Mohammad Yazdani and Pouria Mohammad Alami, Construction and Validation of a Computerized Open-ended Bi-functional Translation Assessment System*, 1 – 37.

- learning and the brain's capacity for change*, New York: Teachers College, Columbia University.
- Ghonsooly, Behzad (2017) *Sanjesh-e tarjomeh* [translation assessment], Tehran: Allameh Tabataba'i University Press.
- Hatim, Basil and Ian Mason (1997) *The translator as communicator*, London & New York: Routledge.
- Honig, Hans (1998) 'Positions, Power and Practice: Functionalist Approaches and Translation Quality Assessment', *Translation and Quality*, Christina Schäffner (ed.), 6–34. Clevedon: Multilingual Matters.
- House, Juliane (1997) *Translation Quality Assessment: A model revisited*, Vol. 410, England: Gunter Narr Verlag.
- Kim, M (2009) 'Meaning-oriented assessment of translations', *Testing and Assessment in Translation and Interpreting Studies*, Amsterdam/Philadelphia: John Benjamins, 123-157.
- Kozulin, Alex (1990) *Vygotsky's Psychology: A biography of ideas*, New York: Harvester-Wheatsheaf.
- Kozulin, Alex, and Erica Garb (2002) 'Dynamic assessment of EFL text comprehension of at-risk students', *School Psychology International*, 23 (1): 112–127.
- Kozulin, Alex and Erica Garb (2004) 'Dynamic assessment of literacy: English as a third language', *European Journal of Psychology of Education*, 19(1): 65-77.
- Lantolf, James P., and Matthew E. Poehner (2004) 'Dynamic assessment of L2 development: Bringing the past into the future', *Journal of Applied Linguistics*, 1(1).
- Luria, Aleksandr Romanovich (1961) 'Study of the abnormal child', *The American Journal of Orthopsychiatry: A Journal of Human Behavior*, 31: 1-16.
- Martínez Melis, Nicole, and Amparo Hurtado Albir (2001) 'Assessment in translation studies: Research needs', *Meta: Journal des traducteurs Meta:/Translators' Journal*, 46(2): 272-287.
- PACTE (2003) 'Building a translation-competence model', in F. Alves (ed.), *Triangulating Translation: Perspectives in process-oriented research*, Amsterdam: John Benjamins, 43-66.
- Pallant, Julie (2010) *SPSS Survival Manual, 4th*. England: McGraw-Hill Education.
- Pishghadam, Reza, and Elyas Barabadi (2012) 'Constructing and validating computerized dynamic assessment of L2 reading comprehension', *Iranian Journal of Applied Linguistics*, 15(1): 73-95.
- Azam Akrami, Behzad Ghonsooly, Mohammad Yazdani and Pouria Mohammad Alami, Construction and Validation of a Computerized Open-ended Bi-functional Translation Assessment System, 1 – 37.*

- Poehner, Matthew E. (2007) 'Beyond the test: L2 dynamic assessment and the transcendence of mediated learning', *The Modern Language Journal*, 91(3), 323–340.
- Poehner, Matthew E., and James P. Lantolf (2010) 'Vygotsky's teaching-assessment dialectic and L2 education: the case for dynamic assessment', *Mind, Culture, and Activity. An International Journal*, 17: 312-330.
- Poehner, Matthew E., Jie Zhang, and Xiaofei Lu (2015) 'Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation', *Language Testing*, 32(3): 337-357.
- Pym, Anthony (2003) 'Redefining translation competence in an electronic age: In defense of a minimalist approach', *Meta: Translators' Journal*, 48(4): 481-497.
- Rashidi, Gholam Reza (2015) *Tarjome va motarjem* [Translation and translator], Vol. 13, Mashhad: Behnashr.
- Sager, Juan C. (1989) 'Quality and Standards – the Evaluation of Translations' In *The Translator's Handbook* (2nd edition), Catriona Picken (ed.), 91–102. London: Aslib.
- Secară, Alina (2005) 'Translation Evaluation—A State of the Art Survey', In *eCoLoRe / MeLLANGE Workshop: Resources and Tools for e-Learning in Translation and Localisation*, 39–44. Manchester: St. Jerome. <https://www.stjerome.co.uk/tsa/edited-volume/2079/#summary>.
- Sternberg, Robert J., and Elena L. Grigorenko (2002) *Dynamic Testing: The nature and measurement of learning potential*, Cambridge University Press.
- Vygotsky, Lev Semenovitch (1978) *Mind in Society. The development of higher psychological processes*, Cambridge, MA: Harvard University Press.
- Vygotsky, Lev Semenovitch (1998) *The Collected Works of LS Vygotsky*, Vol. 5: *The Problem of Age*. 187-206.
- Williams, M. (2009) 'Translation quality assessment. *Mutatis Mutandis*', 2(1): 3-23.
- Zehnalová, J (2013) 'Tradition and Trends in Translation Quality Assessment', *Tradition and Trends in Trans-Language Communication*, 41-58.
- Zequan, Liu (2003) 'Register Analysis as a Tool for Translation Quality Assessment', *Translation Journal* 7 (3).

7. Appendices

Appendix 1: Screenshot of Examinee COBTAS Statistical Test Results (NDA, DA, and LPS)

The screenshot displays the COBTAS exam interface. At the top, it shows 'Multi Functional Dynamic Assessment Exam' and 'Exam Examiner'. The exam information includes 'Exam No: 4017', 'Exam Time: 100', and 'Exam Date: 1396/12/29'. The examinee's details are: Name, Family, Major: Translation, Email, and Student No: 932214567. A 'Register/Submit' button is visible.

The question section shows the text: 'It was a terrible accident. I knew that I couldn't walk any more.' Below the question, there is an 'Answer Box' with Persian text: 'تصادف هولناکی بود. می دانستم که ...' and an answer: 'دیگر نمی توانم راه پیمایی کنم'. There is also a 'Confirm' section with the same Persian text and answer.

The 'Result' section shows the following scores: Non-Dynamic Score: 32, Dynamic Score: 55.3, Hints Used: 43, Total Exam Score: 76, and Learning Potential Score: 1.

On the right side, there is a 'Navigation' panel with 'Previous Q.No. 20', 'Go', and 'Next' buttons. It also shows 'Question No: 20', 'Score: 4', 'Question Time: 5', 'Remaining Time: 01:24', 'Total Response Opportunities: 5', and 'Remaining Response Opportunities: 0'. There are also 'Show' buttons for 'Whole Exam', 'Elapsed Time', and 'Remaining Time'.

The Windows taskbar at the bottom shows the system tray with the date and time: 1:17 AM, 4/22/2018.

COBTAS

Appendix 2: Comparing the Examinees` N-DS and DS

