https://doi.org/10.17323/jle.2023.12262

On the Identifiability of Cognitive Diagnostic Models: Diagnosing Students' Translation Ability

Mona Tabatabaee-Yazdi, Aynaz Samir

Tabaran Institute of Higher Education, Mashhad, Iran

ABSTRACT

Background. In recent years Cognitive Diagnostic Models (CDMs) have attracted a great deal of attention from researchers in a variety of educational fields. However, they have not been taken into consideration in Translation Quality Assessment (TQA), in the aims of presenting fine-grained information about the strengths and weaknesses of translation students.

Purpose. The present study compares the ACDM, DINO, DINA, HO-DINA, and G-DINA models, in order to define the strengths and weaknesses of Iranian translation students and to examine whether the required translation attributes are compensatory, non-compensatory, additive, or hierarchical.

Method. 200 BA translation students translated a two-English-text translation, which was scored by three experienced translation raters using the Translation Quality Assessment Rubric (TQAR). The professional translators, established the relationships between the TQAR items and the nine proposed target translation attributes by constructing a Q-matrix.

Results. Based on the results, HO-DINA can be considered the best-fitting model. Bibliography and technical skills, together with work methodology skills, are shown to be the most difficult attributes for translation students.

Conclusion. HO-DINA is a non-compensatory model, thus the study findings assert that for a correct response to a test item, all measurable attributes need to be mastered.

KEYWORDS

attribute, diagnostic classification models, item response theory, Q-matrix, translation ability, test fairness

INTRODUCTION

According to William (2003), translation is a strategic conscious activity aimed at establishing communication between diverse cultural settings in a controlled way. With the rapid development of globalization and its impact on localization, an increasing volume of products are being imported and exported into other countries. Thus translation plays a significant role in cultural communication (Yan, 2013, p. 36). Indeed, «the need to deliver information quickly and efficiently has put translation at the heart of diverse international cultural, economic, and military enterprises» (Jimenez-Crespo, 2020, p. 375). Due to the importance of globalization and human communication needs, the role of translation has become more valuable. Translation, as an effective way of communication, permits people from different languages and cultures to learn about diverse aspects of the international community, as well as their cultures and ideologies. With the rapid development of translation technology and communication around the world, the ability to translate from or to other languages has become particularly important. According to Nord (1999), translator training programs at universities offer skills and knowledge, which enable students to enhance the required skills, abilities, and translation competences through training, guidance, practice, and getting experience. In this regard, Paradis et al. (1982) believe that the ability to translate foreign languages is a complex integrated cognitive task associated with underlying cognitive components, beyond the

Citation: Tabatabaee-Yazdi M., & Samir A. (2023). On the identifiability of cognitive diagnostic models: Diagnosing students' translation ability. *Journal of Language and Education, 9*(1), 138-157. https://doi.org/10.17323/jle.2023.12262

Correspondence:
Aynaz Samir
aynazsamir@yahoo.com

Received: April 03, 2021
Accepted: Mach 19, 2023
Published: March 31, 2023

ability to speak or understand words and linguistic structure in two languages. In the views of Kelly (2005) and Mackenzie (2004), the translation of different technical texts requires significant interaction among different translation competences and attributes.

In the field of Translation Studies, translation competence [TC] development has been approached from different perspectives. Domínguez Milanés (2015, p. 29) stated that professional competences are «the ensemble of knowledge, skills, personal traits, emotions, motivation, abilities, values, and attitudes that allow an individual to perform successfully not only in a given, situated activity but also in a number of unexpected scenarios in the current, deeply unstable market conditions» (as cited in Martínez-Carrasco, 2017, p. 152). Martínez-Carrasco (2017), studying the nature of translation competence, asserted that «the most common translation sub-competences cover the (inter)linguistic reality of translation, the extra-linguistic input, the role of ICT and other sources of information, [a number of] inter-personal and professional [elements], a psycho-physiological frame, the strategic component, and transferability» (p. 220). For PACTE, translation competence refers to «the underlying system of knowledge required to translate» (2011, p. 321). As «the degree of TC is reflected in both the process and the product of translating» (2014, p. 88), TC encompasses a mix of procedural and declarative knowledge represented in PACTE as a model with five sub-competences (2003): strategic sub-competence, bilingual sub-competence, instrumental sub-competence, extra-linguistic sub-competence, knowledge about translation sub-competence, and psycho-physiological components. Kiraly and Hofmann (2016) add that translation competence is acquired "in a step-by-step, cumulative and essentially linear manner» (p. 72).

Mackenzie (2004) states that translation competence consists of management skills, information technology (IT) skills, marketing ability, linguistic-cultural skills, and interpersonal skills. Kelly (2005, 2008) also suggests her own translation competence model comprising seven sub-competences: (a) interpersonal competence; (b) attitudinal (or psycho-physiological) competence; (c) subject-matter competence; (d) strategic competence; (e) cultural and intercultural competence; (f) professional and instrumental competence; (g) communicative and textual competence in at least two languages and cultures. Perez (2014, as cited in Kabát, 2020, p. 59) also finds that linguistic, intercultural, thematic, technological, info-mining, and translation service provision competences are the most important components of TC. Piecychna (2013) suggests psychological, thematic, textual, and linguistic sub-competences (p. 155). Other translation scholars introduce other important translation skills as part of TC, including grammar skills (Conde, 2013; Dewi, 2015; Hansen, 2010; Lee & Ronowick, 2014), terminological skills (Angelone, 2013; Goff-Kfouri, 2005; Lee & Ronowick, 2014), spelling skills (Beeby, 2000; Dewi, 2015; Doyle, 2003; Waddington, 2001), creativity (Dewi, 2015; Polliastri & Paulina, 2009), time management (Dewi, 2015; Doyle, 2003), problem-solving skills (Dewi, 2015; Nord, 2009), accuracy (Farahzad, 1992; Khanmohammad & Osanloo, 2009; Polliastri & Paulina, 2009), and fluency (Conde, 2011; Dewi, 2015; Farahzad, 1992).

For PACTE (2000) translation competence development can be defined as:

(1) A dynamic, spiral process, which, like all learning processes, evolves from novice knowledge (pre-translation competence) to expert knowledge (translation competence); it requires learning competence learning strategies). During the process, both declarative and procedural types of knowledge are integrated, developed, and restructured.

(2) A process in which the development of procedural knowledge - and consequently of the strategic sub-competence - is essential.

(3) A process in which the TC sub-competences are developed and restructured (as cited in Hurtado Albir, 2015, p. 260).

Thus, assessing student competence in translation programs is essential for the improvement of their professional skills (Beeby et al., 2011; Bonyadi, 2020). Yamaguchi and Okada (2018) mention that achievement tests assess a number of skills and the student's current knowledge. However, translation students with the same overall scores in a translation exam may differ widely in their competences, strengths, and weaknesses. Therefore, a comparison of the sum of the scores of students with the same knowledge differences, cannot really reveal their strengths and weaknesses (Tabatabaee-Yazdi, 2020). Since overall assessment scores cannot satisfy the expectations of the teachers or examiners, the differences in difficulty of the questions should not be ignored (Tabatabaee-Yazdi, 2020). It is crucial to collect detailed information on exam items and more precise data on the specific attributes or skills of students, rather than classifying students based on their scores (DiBello & Stout, 2007; Lee & Sawaki, 2009a).

Only a few empirical studies have been conducted on assessing translation quality with regard to translation process and products (Angelelli & Jacobson, 2009; Stobart & Gipps, 1997). In the words of Conde, "evaluation is still a field in which much remains to be explored" (2012, p. 68). This contrasts with Newmark's (1998) views on assessment: "translation quality assessment is a vital link among translation theory and its practice; it is also a pleasing and informative exercise, specifically, if the assessor evaluates two or more different translations of the same text based on translation standards" (p. 184). Although in some educational settings, statistical assessment techniques have been used to calculate overall skills (Genesee 2002; Genesee & Upshur 1996;

ReaDickins & Germaine, 1993; Tabatabaee-Yazdi, Motallebzadeh, Baghaei, 2021), these techniques have not been used to identify and diagnose the strengths and weaknesses of translation students.

Cognitive Diagnostic Models (CDMs) have emerged to model statistically the examinees' cognitive operations. Their aim is to provide diagnostic feedback, improve teaching and learning processes, and remove the shortcomings of some traditional statistical assessments including classical test theory and factor analysis used to construct tests, and interpret test results by focusing merely on the overall scores. CDMs are given greater attention in some assessment settings since they can increase the opportunity of learning by "pinpointing why students perform as they do" (Leighton & Gierl, 2007, p. 5). The CDM defines the examinees' strengths and weaknesses based on certain specific attributes (Chen & de la Torre, 2013; De la Torre & Lee, 2013; Von Davier, 2005). Rather than placing each examinee on a continuous ability scale based on their scores, CDMs yield categorical diagnostic information about examinees' strengths and weaknesses with different fluency profiles.

CDMs help students answer exam items properly by breaking down the questions/items into several strategies (Birenbaum et al., 1993). Accordingly this permits CDMs to produce "multidimensional diagnostic profiles based on statistically-driven multivariate classifications" of students with regard to performance levels on each of the required skills (Kunina-Habenicht et al., 2012, p. 64).

Ravand & Robitzsch (2018) determine two main purposes for using CDMs: (a) categorizing students into similar skill profiles regarding their answer patterns and (b) identifying compensatory or non-compensatory relations between the attributes of a given skill. There are different arrays of CDMs with various theoretical principles regarding relations between attributes (Ravand & Baghaei, 2020). This is a series of certain mental processes, knowledge, strategies, skills, and competences in which students must answer the items of a test correctly (Leighton & Gierl, 2007).

Types of CDMs

The main factor that distinguishes CDMs from each other is the way in which they model the association between the required attributes when performing a test item or a given task, as well as the probability of a correct response (Table 1, Ravand, 2016, p.3). According to the correlation between attributes, CDMs are classified into two types: (1) conjunctive or non-compensatory models, and (2) disjunctive or compensatory models. If the performance of one or more of the attributes can compensate for the non-performance of other attributes, *compensatory models (Disjunctive)* are used (Ravand & Baghaei, 2020; Tabatabaee-Yazdi, 2020). Conversely, if the performance of one or more of the attributes cannot compensate for the non-performance of other attributes, non-compensatory models (Conjunctive), are used. Thus, in order to achieve a high probability of a correct answer, performance of all the required attributes is needed (Ravand & Baghaei, 2020; Tabatabaee-Yazdi, 2020).

Additive CDMs (de la Torre, 2011) have been proposed as another classification of CDMs. Unlike compensatory models, respondents are credited for the number of attributes performed, signifying that the performance of any one of the attributes can increase the possibility of a correct answer (Ravand & Baghaei, 2020). Finally, the most recent extensions or versions of CDMs, proposed by Templin & Bradshaw (2013), are hierarchical and non-hierarchical CDMs. Hierarchical CDMs (HCDMs) model the structural relations between the required attributes and the impact which the order of teaching materials (where learning a skill is prioritized upon other skills) has on increasing the likelihood of obtaining a correct answer to an item (Tabatabaee-Yazdi, 2020).

More recently, CDMs have been categorized under two major categories: specific and general. In specific CDMs, just one type of association (disjunctive, conjunctive, and additive) can be possible within any assessment. However, general CDMs do not hypothesize any pre-specified relations between underlying sub-skills. Therefore, several kinds of interactions are possible within the same assessment, assuming different relationships between attributes across the items (Ravand & Baghaei, 2020; Tabatabaee-Yazdi, 2020).

Nearly all the models used in the context of the assessment show that they are effective for offering diagnostic feedback in the teaching and learning process (Nichols, 1994). Therefore, a number of CDMs have been used in language assessment studies such as the DINA (Chen & Chen, 2016), the General Diagnostic Model (GDM) (Von Davier, 2005), the G-DINA (Chen & Chen, 2016; Effatpanah et al., 2019; Ravand & Baghaei, 2020; Ravand et al., 2020).), the DINO (Chen & Chen, 2016), the reduced reparameterized unified model (RRUM) (Aryadoust, 2018; Kim, 2015; Lee & Sawaki, 2009a; Li, 2011), and the hierarchical diagnostic classification model (HCDM) (Tabatabaee-Yazdi, 2020).

The objective of this study was to apply and compare the DINO, DINA, ACDM, HO-DINA, and G-DINA models to recognize the strengths and weaknesses of the translation ability of Iranian BA students in a translation test. It also aimed to test whether the required translation attributes are compensatory, non-compensatory, additive, or hierarchical. Accordingly, DINO (as a specific disjunctive model), DINA (as a specific conjunctive model), ACDM (as a specific additive model), HO-DINA (as a specific hierarchical model), and G-DINA (as a general multifunctional model) were applied.

Table 1
CDM Categorization

	DCM Type	Examples	Author(s)
Specific	Disjunctive	• deterministic-input, noisy-or-gate model (DINO)	• Templin & Henson (2006)
		• noisy input, deterministic-or-gate (NIDO) model	• Templin (2006)[1]
	Conjunctive	• deterministic-input, noisy-and-gate model (DINA)	• Junker & Sijtsma (2001)
		• noisy inputs, deterministic "and-gate (NIDA)	• DiBello, Stout, and Roussos (1995); Hartz (2002)
	Additive	• Additive CDM (ACDM)	• de la Torre (2011)
		• compensatory reparameterized unified model (C-RUM)	• DiBello, Stout, and Roussos (1995); Hartz (2002)
		• non-compensatory reparameterized unified model (NC-RUM)	• Hartz (2002)
		• linear logistic model (LLM)	• Maris (1999)
	Hierarchical	• hierarchical DINA (HO-DINA) model	• de la Torre (2008)
General / Saturated	Disjunctive, Conjunctive, & Additive Hierarchical	• general diagnostic model (GDM)	• Von Davier (2005)[2]
		• log-linear CDM (LCDM)	• Henson, Templin & Willse (2009)
		• generalized DINA (G-DINA)	• de la Torre (2011)
		• hierarchical diagnostic classification model (HDCM)	• Templin & Bradshaw (2013)

Note. Adapted from "Application of a Cognitive Diagnostic Model to a High-Stakes Reading Comprehension Test", by H. Ravand, 2016, Journal of Psychoeducational Assessment, 34, p. 3, (https://doi.org/10.1177/0734282915623053). Copyright 2016 by SAGE

DINO

Developed by Templin and Henson (2006), DINO is considered the first in the line of CDMs. In this model, which is very similar in structure and composition to its counterpart DINA, the performance of any one of the attributes increases the chance of a right answer to the test item or the given task. Therefore, in the DINO model, students "are classified into two latent classes: those who have not mastered any of the required subskills and those who have mastered at least one of the subskills" (Aryadoust, 2018, p. 7). "The DINO model is often used in the application of psychiatric assessment, for which the positive response to a diagnostic question (item) could be due to the presence of one disorder (attributes) among several" (Fang, Liu, & Ying, 2019, p. 8).

DINA

DINA, as a non-compensatory model (Junker & Sijtsma, 2001), is known for its parsimony, understandability, and as an easy fit for the data (Fang et al., 2019). Local independence at the attribute level is one of its main features, indicating that the performance of one attribute does not affect that of another. Therefore, "measured attributes in an item are independent of one another" (Galeshi, 2012, p. 18). In other words, the item or task can be answered correctly, if all the required attributes have been mastered thoroughly. However, students who have not mastered any of the attributes may hazard a guess and answer the item accurately (Rupp, Templin & Henson, 2010).

ACDM

Additive CDM (ACDM; de la Torre, 2011), another compensatory model similar to G-DINA (de la Torre, 2011) as a general model, permits both disjunctive and conjunctive relations between attributes within the same test. The ACDM hypothesizes that the probability of producing a correct answer to an item or a task can be increased by mastering each of the required attributes. The absence of one attribute can be compensated for by the existence of other attributes (Effatpanah, 2019). Therefore, ACDM assumes that mastering each of the required attributes advances the likelihood of success, while the absence of one attribute can be compensated by the mastery of other attributes. Moreover, the existence of each sub-skill is independent of the other subskills (Galeshi, 2012).

HO-DINA

HO-DINA is an extension of DINA, which hypothesizes that the required attributes are hierarchically structured (De La

1 Templin, J. (2006). *CDM User's Guide*. University of Kansas.

2 Von Davier, M. (2005). *mdltm* [Computer software]. Educational Testing Service.

Torre & Douglas, 2004; De La Torre & Minchen, 2014), and fit the data better than DINA (De La Torre & Douglas, 2004). The first assumption of HO-DINA is that performance of the required attributes is interrelated to “a higher-order and broadly-defined ability parameter similar to the unidimensional θ parameter in the IRT models” (Aryadoust, 2018, p. 6); and secondly, the existence of each attribute is independent of the other attributes (De La Torre & Douglas, 2004). Therefore, HO-DINA can be considered a non-compensatory model which claims that the only essential condition for a correct response to a test item or a given task is performance of all the required attributes measured by that item (Aryadoust, 2018).

G-DINA

G-DINA (De la Torre, 2011) is a general CDM which allows for both compensatory and non-compensatory relationships within the same test, signifying a different model for each item on the same test (Ravand & Robitzsch, 2018). In the G-DINA model, unlike the DINA model, the non-performance of one, some, or all of the required attributes leads to an unequal probability of success for the students. This indicates that irrespective of how many attributes students have mastered, as long as they have mastered at least one of the required subskills (De La Torre & Minchen, 2014), they have the same probability of answering the items correctly. Consequently, knowing one or all the attributes does not necessarily lead to a higher chance of giving a correct response (De La Torre & Minchen, 2014).

METHODS

Participants and Settings

The data set of the present study comprises 200 Iranian junior and senior university students studying *English Translation* at undergraduate level in various national and private universities in Iran. The sample consisted of 51 males (25.5%) and 149 females (74.5%) whose ages ranged between 20 to 44 (M=24.79 years, SD= 3.89). Of the total sample, 157 (78.5%) were studied at national universities and 43 (21.5%) at private universities (Table 2).

Three raters assessed the translations. All are assistant professors and were native speakers of Persian and proficient in English as their foreign language. They had at least 8 years of experiencing in teaching and translating. Three possessed an MA in *Translation Studies* and had been teaching translation courses in different universities of Iran for more than seven years. They were one male and two females between the ages of 27 to 38 years old ($mean_{age}$=36, SD= 1.2). Each rater was provided with a Translation Quality Assessment Rubric, consisting of 23 items to score the translations (Samir & Tabatabaee-Yazdi, 2020).

Instruments

Two instruments were used to collect the data required for the specific objectives. The first instrument was a translation test comprising two texts (a political text of 304 words and a journalistic text of 303 words) extracted from two textbooks on political translation, and news articles taught to BA students at Iranian universities. The political text was about the cabinet government system in different countries. The journalistic text examined terrorism in Asian countries. The two texts were classified as advanced. The text were to be translated into Persian in class within one hour. Students were allowed to use any kind of resources including the internet, online dictionaries, Computer-Aided Translation (CAT) tools, and/or any other software.

The second instrument was Translation Quality Assessment Rubric (TQAR) developed and validated by Samir and Tabatabaee-Yazdi (2020; see Appendix B). The rubric consists of 23 items on a four-point Likert scale, ranging from 4 (*superior*), 3 (*advanced*), 2 (*fair*), to 1 (*poor*). The rubric can be used to identify and score the translation abilities of non-native

Table 2

Demographic Profile of Respondents (n=200)

	Category	Frequency	Percentage
Gender	Male	51	25.5
	Female	149	74.5
Age	20-24	97	48.5
	25-29	57	28.5
	30-34	25	12.5
	35-39	15	7.5
	Above 40	6	3
University	National	157	78.5
	Private	43	21.5

English-speaking translation students in the university context.

For the sake of CDM analysis, and constructing the required Q-matrix, the four-point Likert scale was merged into *mastered* and *non-mastered* . This was achieved by categorizing "poor" and "fair" scales into "non-mastered= 0", and "advance" and "superior" into "mastered= 1". Thus, the TQAR turned into a dichotomous response scale of yes=1 and no=0 to assess the nine proposed translation attributes.

Procedure

Data collection began in October 2019 and lasted for about four months. A consent form was distributed beforehand to the fifty-five English Translation departments currently offering translation programs in Iranian universities, in order to seek their permission and willingness to be included in the study. Only 18 signed consent forms were sent back to the researchers. Thus, only 18 universities, representing National and Azad Universities, participated in this study. In total, 200 translation students were included. Three one-hour training sessions were held for translation raters, in order to increase their awareness about the research objectives, translation attributes, TQAR, Q-matrix construction, and the rating scale.

Analysis

In order to identify accurate correlations between target attributes and test items, a Q-matrix was hypothesized (Table 3, Appendix A). According to Li (2011), Tabatabaee-Yazdy, (2020), and Tatsuoka (1983) a Q-matrix illustrates rows[3] and columns[4] which symbolize test items and required underlying traits to answer each item correctly. In order to identify the required underlying traits or subskills, the present study used Samir and Tabatabaee-Yazdi's Translation Quality Assessment Rubric (2020; see Appendix B) as the Q-matrix items and sub-skills. In a joint session, the researchers and the three raters agreed on the required attributes. Then, the translation raters created a Q-matrix by organizing the 23 items of the TQAR into nine translation attributes, as follows:

(1) linguistic skills (lexical) in the source language (LSSL: L), used to assess students' ability to recognize the meaning of a wide variety of terminology and lexical items in the source text (ST) accurately and appropriately,

(2) linguistic skills (grammatical) in the source language (LSSL: G), aiming at evaluating students' ability in recognizing the ST grammatical rules such as the relative order of subject, verb, modifiers, clauses, and syntactic elements,

(3) linguistic skills (lexical) in the target language (LSTL: L), used to assess students' ability to use specific terms in the translation of the technical text,

(4) linguistic skills (grammatical) in the target language (LSTL: G), used to identify students' ability to follow target language (TL) grammatical rules, such as subject and verb agreement or arrange the words according to the TL rules,

(5) organizational knowledge (ORG), used to assess students' ability to translate and organize ideas cohesively and coherently in order to convey all sections (sentences, titles, headlines…) of the ST to TT,

(6) cultural knowledge (CUL), used to assess students' ability to produce correct and idiomatic use of the target language and to preserve an appropriate register in the translation,

(7) translation skills (TRN,. aiming to study students' ability to produce the target text (TT) at an acceptable level of fluency and avoid words and expressions of ambiguous meanings,

(8) work methodology skills (WM), used to assess students' ability to manage time and respect the deadline,

(9) bibliography and technical skills (BT), used to assess students' ability to use a relevant terminological database or bibliography.

The final Q-matrix presented in Table 3 was proposed after several rounds of revisions and the necessary modifications. Of the total items, 11 of them were associated with attribute 1 (LSSL: L), 11 items, with attribute 2 (LSSL: G); 13 items, with attribute 3 (LSTL: L); 13, with attribute 4 (LSTL: G); 11, with attribute 5 (ORG); 8, with attribute 6 (CUL); 13, with attribute 7 (TRN); 4, with attribute 8 (WM); and 3, with attribute 9 (BT). 1s in Table 3 indicate that the probability of producing a correct answer on each item is conditional on the mastery of the attributes, whereas 0s show that the item does not need the attributes. For example, in order for students to get the correct answer to item 7 (to deal with terminological terms correctly), they should have mastered the first, third, and seventh attributes. Therefore, in this item, students are not required to master attributes 2, 4, 5, 6, 8, and 9.

[3] Rows in this matrix signify the number of items on the test.

[4] Columns in this matrix show the number of test's underlying attributes.

RESULTS

R statistical software and CDM package version 6.1-10 (Robitzsch, Kiefer, George & Uenlue, 2018[5]) were used to analyze and compare the fit of the five selected CDMs, including DINO, DINA, ACDM, HO-DINA, and G-DINA. The CDM package illustrates the various fit indices which can be applied, in order to identify the best model among the competing models (relative fit indices), It can also be used to check the fit of a model to the observed response data (absolute fit indices) (Effatpanah, Baghaei, & Boori, 2019; Rupp, Templin, & Henson, 2010; Tabatabaee-Yazdi, 2020). The models were compared to the relative and absolute fit indices using AIC[6], BIC[7], CAIC[8], Mx2[9], MADcor[10], SRMSR[11], MADQ3 [12], and MADRESCOV[13].

Optimal Model Fit

The fit of the data to CDMs identifies the accuracy of the correlation between attributes and items. Table 4 shows the Absolute and Relative fit indices of the five models. Although there are no agreed-upon cut-off values for the absolute fit indices of CDMs, the smaller the effect size (values), the better a model fits (Robitzsch et al., 2015[14]).

As can be seen in Table 4, the Npars (number of parameters) column shows that the G-DINA model estimated 1094 item parameters: ACDM, 156 parameters; DINA and DINO 92, parameters; and HO-DINA, 64 parameters. This shows that HO-DINA is a parsimonious model and G-DINA is the most complicated model. Moreover, there is a non-significant value (> 0) for MaxX2 for all the models. As to the MADcor, SRMSR, and MADRESIDCOV, the G-DINA had the lowest values. However, with respect to AIC, BIC, and CAIC, the value of HO-DINA was the lowest when compared to the other models. Therefore, since BIC carries a great penalty for more highly parameterized models, it can be anticipated that the G-DINA model has the worst value (Li, Hunter, & Lei, 2016).

As a result, after considering all the indices, the HO-DINA was selected as the best-fitting specific CDMs for further examination, in order to study whether the model can precisely detect the students' translation skills.

HO-DINA Analysis

HO-DINA Parameters

Table 5 describes the study's model (HO-DINA) parameters. Ttwo items (items 7 and 8) are shown in Table 5 as examples. The successful performance of item 7 (Terminology and False Friends) requires students to have mastered attributes 1 (LSSL: L), 3 (LSSL: L), and 7 (TS). Students who have mastered none of the required attributes only have a 30% chance to respond correctly. However, students who have mastered only LSSL: L have a 31% probability of guessing the correct item. Students who master attribute 1 had a 0.30 + 0.31 = 0.61% probability of success on item 7. In the same vein, if students have mastered only attribute 3, there is a 0.30% chance of answering the item correctly. Thus, those who have mastered attribute 3 have a 0.30 + 0.30 = 0.60% probability of success on this item. Finally, those respondents who have mastered the three attributes have a 56% probability of getting the right answer and, consequently, 86% (0.56 + 0.30 = 0.86) probability of success on this item. Another example would be *spelling* (item 8), where students are required to know attribute 3 (LSSL: L). If the students have not mastered the attribute, they only have a 0.74% chance to perform well on this item. However, by mastering the attribute, the probability of performing item eight increases to 99%.

5 Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, C. (2018). CDM: Cognitive diagnosis modeling (Rpackage version 6.1-10). https://cran.r project.org/web/packages/CDM/index.html

6 Akaike Information Criterion (AIC; Akaike, 1974) is one of the relative fit indices used to select between non-nested models. Models with lower AIC are more preferable.

7 Bayesian Information Criteria (BIC; Schwarz, 1978) is another relative fit index used to choose between non-nested models. Models with lower BIC are more preferable.

8 Consistent AIC (Akaike, 1974).

9 Mx2 (Chen & Thissen, 1997) is the test of global model fit. It is the mean difference between the model-predicted and observed response frequencies. If CDM fits the data well, "the x2 test statistic is expected to be 0 within each latent class as the attribute profile of the respondents would perfectly predict the observed response patterns" (Rupp et al., 2010, p. 269).

10 The mean absolute difference for the item-pair correlations (MADcor) statistic (DiBello, Roussos, & Stout, 2006) shows the mean of absolute discrepancy between observed and predicted pairwise item correlations across all item pairs. A MADcor value of 0.049 in Jang (2005) was suggested as a good fit for the DCM to the data.

11 The standardized root mean square residual (SRMSR) is the square root of the difference between the observed correlation and the model covariance matrix. Models with SRMSR values below 0.05 are considered models with "substantively negligible amount of misfit" (Maydeu-Olivares, 2013, p. 84) and models with values below 0.08 as good fit (Hu & Bentler, 1999).

12 The $MADQ^3$ is "the Pearson product-moment correlation of a set of residuals from the IRT model" (Chen & Thissen, 1997, p. 280; Yen, 1984). This index is to some extend less sensitive than the Mx^2 (Aryadoust, 2018).

13 The average of absolute values of pairwise item covariance residuals (MADRESCOV; McDonald & Mok, 1995) illustrates the average deviations between matrices of observed and reproduced item correlations.

14 Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2015). CDM: Cognitive Diagnosis Modeling. R package version 4.5-0. http://CRAN.R-project.org/package=CDM

Table 3

The Final Q-Matrix

Attribute Item	Att. 1 LSSL: L	Att. 2 LSSL: G	Att. 3 LSTL: L	Att. 4 LSTL: G	Att. 5 ORG	Att. 6 CUL	Att. 7 TRN	Att. 8 WM	Att. 9 BT
1	0	1	0	1	0	0	1	0	0
2	1	1	1	1	1	1	0	0	0
3	1	1	1	1	1	1	1	0	0
4	1	1	1	1	1	1	1	0	0
5	0	0	0	0	1	0	1	0	0
6	0	1	0	1	0	0	0	0	0
7	1	0	1	0	0	0	1	0	0
8	0	0	1	0	0	0	0	0	0
9	1	1	1	1	0	0	1	0	0
10	1	1	1	1	1	1	1	0	0
11	1	1	1	1	1	1	1	0	0
12	1	1	1	1	1	0	1	0	0
13	0	0	1	1	1	0	1	0	0
14	1	1	1	1	1	1	1	0	0
15	1	1	1	1	1	1	1	0	0
16	1	0	1	1	1	1	1	0	0
17	0	0	0	0	0	0	0	1	0
18	0	0	0	0	0	0	0	1	0
19	0	0	0	0	0	0	0	1	0
20	0	0	0	0	0	0	0	1	0
21	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	1
23	0	0	0	0	0	0	0	0	1

Table 4

Fit Indices

	Relative Fit Indices			.	Absolute Fit Indices				.
Model	Npars	AIC	BIC	CAIC	MaxX2 (p)	MADcor	SRMSR	MADQ3	MADRESIDCOV (MADRCOV)
DINO	92	3936.19	4239.64	4331.64	178.314(0)	0.06378	0.10439	0.07110	0.98298
DINA	92	3935.68	4239.12	4331.12	178.313(0)	0.06347	0.10411	0.06874	0.98105
ACDM	156	3924.42	4438.96	4594.96	**174.389(0)**	0.05037	0.09266	**0.05990**	0.77066
HO-DINA	**64**	**3923.23**	**4134.33**	**4198.33**	178.357(0)	0.06577	0.10502	0.07812	1.00438
G-DINA	1094	5567.13	9175.49	10269.49	175.756(0)	**0.04445**	**0.08931**	0.06571	**0.70469**

Table 5

HO-DINA Parameters

Item No.	Required Attributes	Mastery Patterns	Probability
7	Att.1–3–7	A000	0.30
7	Att.1–3–7	A100	0.31
7	Att.1–3–7	A010	0.30
7	Att.1–3–7	A001	0.31
7	Att.1–3–7	A110	0.30
7	Att.1–3–7	A101	0.31
7	Att.1–3–7	A011	0.31
7	Att.1–3–7	A111	0.56
8	Att.3	A0	0.74
8	Att.3	A1	0.99

Table 6

Attribute Difficulty

Attributes	Attribute probability 1
Linguistic skills (lexical) in the source language	0.814
Linguistic skills (grammatical) in the source language	0.808
Linguistic skills (lexical) in the target language	**0.932**
Linguistic skills (grammatical) in the target language	0.808
Organizational knowledge	0.818
Cultural knowledge	0.815
Translation skills	0.814
Work methodology skills	0.086
Bibliography and technical skills	**0.074**

Attribute Difficulty

Table 6 shows the performance status of the nine translation attributes. As the table shows, bibliography and technical skills, and work methodology skills, with probabilities of 7% and 8%, respectively, are the most difficult attributes. Only 7% of the students have mastered and can mobilize their bibliography and technical skills while translating a text. Conversely, linguistic skills (lexical) in the target language were shown to be the easiest attribute to master (93% probability), indicating that 93% of the students mastered and are able to mobilize linguistic skills (lexical) in the target language satisfactorily.

Attribute Correlations

CDMs are also used to calculate the correlation between the attributes to show the extent of similarities between them. Values larger than 0.70 reflect a strong correlation; 0.50 and 0.70 are considered as moderate, and less than 0.50 as weak (Henson, Templin, & Douglas, 2007; Kunina-Habenicht, Rupp, & Wilhelm, 2012).

As Table 7 shows, all the values show a strong correlation between the attributes, signifying a non-compensatory, and complementary nature of the HO-DINA model (Aryadoust, 2018; Effatpanah, 2019; Lee & Sawaki, 2009b; Ozaki, 2015; Stone & Zhang, 2003; SU, 2013; Wang, Zheng & Chang, 2014), as wll as the existence of a relationship between the translation attributes. This indicates that these attributes require almost the same underlying cognitive processes. For example, if a respondent performed well on attribute one, he/she would probably perform well on attributes 2, 4, 5, 6, and 7 as well. However, “hierarchy relationships of attributes differ from pre-requisite relationships of attributes in which attributes are ordered in difficulty” (Lim, 2015).

Class Probabilities

As Table 8 illustrates, class probabilities show how students are ranked in different latent classes. The number of latent classes in CDMs, is calculated through (2^k), where k is the number of attributes. Therefore, there are 2^9=512 latent classes according to the present study's Q-matrix configuration. The first column shows some of the latent classes (2^9 = 512) and the second column signifies the related attribute patterns.

For space considerations, the table shows only some of the latent classes. The data shows that the attribute profile α128= [111111100] was the most populated class with 68% probabilities, This class include approximately 135 students, followed by attribute profiles α1=[000000000] and α512 = [111111111] with 7% probabilities. These classes included approximately 13 students in each, respectively. Therefore, regarding α128, the data indicates that 135 students are expected to have mastered the first seven attributes, while not the eighth and ninth ones. Moreover, 13 students have mastered all the nine attributes, and 13 others have mastered none of the attributes.

Class Probabilities for the Respondents

The class probability of the response pattern (11111111011100001000000) is presented in Table 9. The table illustrates the probabilities that each student belonged to each of the 512 latent classes. The first column shows some of the latent classes (2^9 = 512) and the second column shows the response patterns. Table 9 shows that, for example, there are 23% and 17% probabilities for student number 1, belonging to latent classes 128 and 95, respectively. Thus, this student has a 23% chance of mastering all attributes with the exception of the eighth and ninth, while there is only a 7% probability that he/she has mastered all the attributes or belongs to latent class 512.

Skill Mastery Probabilities

Table 10 shows the skill performance probabilities of respondents on each of the required attributes for each of the test items or tasks. Due to space restrictions, the attribute performance probability of only five randomly chosen students is illustrated in Table 10. The values above 0.50 show

Table 7
Attribute Correlations

Attribute Item	Att. 1	Att. 2	Att. 3	Att. 4	Att. 5	Att. 6	Att. 7	Att. 8	Att. 9
Att. 1 (LSSL: L)	1.00								
Att. 2 (LSSL: G)	0.99	1.00							
Att. 3 (LSTL: L)	0.97	0.97	1.00						
Att. 4 (LSTL: G)	0.99	0.99	0.97	1.00					
Att. 5 (ORG)	0.99	0.99	0.97	0.99	1.00				
Att. 6 (CUL)	0.99	0.99	0.97	0.99	0.99	1.00			
Att. 7 (TRN)	0.99	0.99	0.97	0.99	0.99	0.99	1.00		
Att. 8 (WM)	0.96	0.96	0.94	0.96	0.97	0.96	0.97	1.00	
Att. 9 (BT)	0.95	0.96	0.93	0.96	0.95	0.95	0.95	0.97	1.00

Table 8
Class Probabilities

Latent Class	Attribute Pattern	Class probability	Class Expected Frequency
1	000000000	0.07	13.37
64	111111000	0.01	1.17
126	101111100	0.01	2.00
127	011111100	0.01	1.21
128	111111100	**0.68**	135.02
256	111111110	0.02	4.24
384	111111101	0.01	0.73
512	111111111	0.07	13.11

Table 9

Class Probabilities for the Respondents

Latent Class		Response Patterns
		11111111011100001000000
5	001000000	0.07
14	101100000	0.04
49	000011000	0.02
79	011100100	0.08
95	011110100	**0.17**
96	111110100	0.10
123	010111100	0.09
128	111111100	**0.23**
298	100101001	0.00
511	011111111	0.02
512	111111111	0.07

Table 10

Skill Mastery Probabilities

Test Takers	Response Pattern	Attribute Profile	Probability	LSSL: L	LSSL: G	LSTL: L	LSTL: G	ORG	CUL	TRN	WM	BT
3	00000101101101100000101	111110000	0.60	0.03	0.02	0.69	0.02	0.03	0.03	0.02	0.00	0.00
18	10111101111111110011001	011110100	0.81	0.88	1.00	0.99	1.00	0.99	0.92	1.00	0.00	0.00
83	11111111111110100100001	111111110	0.73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.27	0.02
98	11111101111111110111111	111111111	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99
131	00111101100111000001100	011011100	0.34	0.72	0.38	0.92	0.36	0.77	0.74	0.77	0.00	0.00

that the respondents have mastered the attributes with high confidence (Hu, Miller, Huggins-Manley, & Chen, 2016).

As an illustration, the probabilities that student 18 with the skill profile of [011110100] has mastered the attributes are: 0.88, 1.00, 0.99, 1.00, 0.99, 0.92, 1.00, 0.00, and 0.00, respectively. In other words, there is an 88% chance that he/she has mastered LSSL: L and 0% probability for mastering WM and BT.

DISCUSSION

Nearly all educational tests require students to become involved in different forms of cognitive processing. While trying to validate the inferences drawn from the students (Embretson, 1983, 1998; Messick, 1989; Snow & Lohman, 1989), the students' knowledge and strategies in applying any cognitive processing and problem-solving strategies during exams must be taken into account. Most measurement specialists believe that cognitive theory has a significant role in educational assessment (Frederiksen et al., 1993; Irvine & Kyllonen, 2002; Nichols et al., 1995), They consider that it can help scholars investigate the internal features of tests, assess the rules and develop innovative psychometric models, and describe the psychology which underpins students' test performance (Gierl et al., 1999; Hattie et al., 1999; Nichols, 1994; Nichols & Sugrue, 1999). In this regard, Embretson (1983) believes that cognitive theory advances psychometric practices by postulating the construct representation of a test through students' mental processes, knowledge, and the strategies they use to answer the test items. When these cognitive needs are appropriately explained, they can be gathered into cognitive models to design items which can elicit particular mental processes and problem-solving strategies. Thus, a test score "anchored to a cognitive model is more explainable, and perhaps more meaningful, to a diverse group of users, since performance is described using a set of cognitive competences in a well-defined content area" (Leighton et al., 2004, p. 205). Cognitive models, as a "simplified description of human problem solving on standardized educational tasks which helps to characterize the knowledge and skills students at different levels of learning have acquired and to facilitate the explanation and prediction of students' performance" (Leighton & Gierl, 2007, p. 6), provide an explicit framework for linking cognitively based inferences with specific, fine-grained test score interpretations" (Gierl & Leighton, 2007; as cited in Gierl et al., 2010, p. 319).

In the field of Translation Studies, the student's ability to translate from a foreign language is a complex, integrated cognitive task related to certain underlying cognitive components (Paradis et al., 1982, as cited in Lorenzen & Murray, 2008). Translation students are expected to perform a set of translation competences, in order to translate appropriately in a test. However, an overall assessment score in such tests cannot satisfy the expectations of the teachers or examiners. Overall scores cannot distinguish between learners in terms of their skills and competences. Thus, evaluation of strengths and weaknesses requires more specific information from tests, as well as more precise data on specific attributes or skills, rather than classifying students based on their scores (Lee & Sawaki, 2009b).

CDMs can be an effective tool in measuring how attributes underlying translation interact to produce an appropriate translation, hence the decision to apply the DINO, DINA, ACDM, HO-DINA, and G-DINA models. The fit of the data to CDMs reveals the level of accuracy in the correlation between attributes and items. The results of fit statistics showed that the HO-DINA was the best-fitting model. This also follows Su's (2013) findings, who asserted that Hierarchal DINA is more applicable in small sample sizes. Moreover, HO-DINA as the best fitting model supports the assumption that "the attribute hierarchy provides an interpretative framework to guide both the development of items and the interpretation of examinees' scores, in such a way that test performance can be linked to specific cognitive inferences about examinees' knowledge and skills (Gierl, Alves, & Majeau, 2010, p. 319)**.** Therefore, the study's proposed translation attributes can form a hierarchy which outlines the psychological ordering of the attributes or cognitive skills needed to answer the test items.

The results specify that there is a strong correlation between the translation attributes. This means that some of the attributes required almost the same underlying cognitive processes. This could serve to claim that there is a non-compensatory relation between the translation attributes, and thus translation skills are complementary and interdependent.

The findings also indicate that the two *flat-mastery profile*s called "master of all attributes" α512= [111111111] and "non-master of all attribute" α1= [000000000] were two of the most prevalent skill profiles. Approximately, 7% of the students were classified as master of none of the skills, and 7% were classified as master of all skills. The latent class 128 with an attribute profile of α128= [111111100] has a class probability of about 0.68, indicating that approximately 68% of the students are expected to have mastered the first seven attributes. According to Lee and Sawaki (2009b) and Rupp et al. (2010), flat skill profiles are caused by a high positive correlation among the attributes or unidimensionality of the evaluated scale. Thus a learner who has mastered a skill could be a master of the other skills as well.

Furthermore, the results revealed that LSTL: L is the easiest. It showed that a large number (93%) of the translation students can use specific terms in the translation of the political and journalistic texts, avoid spelling errors which cause misunderstanding about the intended meaning, translate the message and the structure of the ST expression close to the TT, and demonstrate content and meaning at a good level of accuracy in the TT (Abdi, 2019; Alibabaei, 2020; Davaninezhad, 2016; Kazemian & Vasheghani Farahani, 2020; Khatibzadeh & Sameri, 2013; Samir & Tabatabaee-Yazdi, 2020).

The analysis has also shown that BT, TRN, and WM are the most difficult translation attributes to master. Budianto and Fardhani (2010) declare that "the primary difficulty when translating a text into a second language is to produce a natural-sounding target text" (p. 5). Klimkowski (2015) also states that "42.19 percent of the translation students had insufficient skills of time and work management in successful professional functioning" (p. 79). According to Marczak and Krajka (2016), two important skills in order of importance for the translation students to learn are CAT tools as well as time management skills. The results are inconsistent with the findings by Molanazar and Kamyab (2015) who emphasize revising and editing as necessary skills that all students need to acquire at university. Khoshsaligheh et al. (2019) and Sharif (2016) state that adding more practical workshops and courses to the curriculum, such as translation of technical texts and revising and self-assessment skills, prepare students to work as qualified translators for the real workplace. Mossop (2003), additionally arrived at the same conclusion. He has suggested that, in order to overcome these deficiencies and problems, translation training programs should contain certain subjects which use advanced forms of computer software, including word processing. Today, translation students consider that word processing programs are the standard means of transforming a source text into their target text (Taghizadeh & Azizi, 2017). A similar conclusion was reached by Abdi (2020) and Austermuhl (2014). They assert that technology is currently of paramount importance in the translation market, and as such translators must have the relevant knowledge and skills to work with national and international clients and access different data resources. Molanazar and Kamyab (2015) have noted that different computer technologies can be used in the translation production process. In view of that, in order "to survive in the Iranian translation market, it is necessary for translation students to acquire CAT tools, such as translation memories, word processing programs, terminology management systems, multilingual dictionaries, or even at times raw machine translation output" (Khazaeefar & Khoshsaligheh, 2014, p. 147).

The findings of this study can have implications for translation students and teachers. Making translation students aware of their strengths and weaknesses in translation can encourage them to concentrate on their problems and enhance their translation competences. It also enables translation teachers to have a better understanding of the competences where translation students face more difficulties.. Building on the results from the students and the weaknesses encountered in TRN, WM, and BT skills, teachers can aid and support students by offering practical workshops in producing a target text (TT) to an acceptable level of fluency, managing the time, and using relevant bibliography, or electronic terminology instruments. These workshops could help students deal with difficult times in translation performance (Kafi et al., 2018; Taghizadeh & Azizi, 2017; Samir et al. 2018). Besides, the proper identification of translation students' competences leads teachers to apply effective teaching techniques and construct translation tests suitable for students' performance and levels.

CONCLUSION

The present study, aimed at investigating the strengths and weaknesses of Iranian Translation students at BA level in translation and diagnosing their translation competences, has confirmed the usefulness of the TQAR in assessing the quality of translation in different genres. The findings indicate that HO-DINA cannot only classify the translation student's skill mastery/non-mastery reliably and properly but also offer informative and valid information about the learning status of translation students. The findings also advance the improvement of translation assessment by offering diagnostic information regarding cognitive processes and the relation between the translation's attributes. Accordingly, more concern must be taken toward educational assessment in translation programs, which requires the help and support of different specialists in the field of translation studies.

Importantly, this study is unique in that, for the first time, CDMs have been implemented in Translation Studies and more specifically in Translation Quality Assessment. However, the current study has a significant limitation that needs to be discussed. The present study analyzed the results of only 18 English Translation departments in a number of universities in Iran. Thus, future research could evaluate the strengths and generalizability of the findings in other countries and universities. Despite the findings recorded, further research is required to discover, test, and compare the existence of different probable hierarchies among the study's proposed attributes.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Mona Tabatabaee-Yazdi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Aynaz Samir: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

REFERENCES

Abdi, H. (2020). Translation and technology: Investigating the employment of computer-aided translation (CAT) tools among Iranian freelance translators. *Theory & Practice in Language Studies*, *10*(7), 811-818. http://doi.org/10.17507/tpls.1007.13

Abdi, H. (2019). A skopos-theoretical analysis of political texts from English into Persian: The case of Donald Trump's executive orders. *Theory and Practice in Language Studies*, *9*(7), 884-890. http://doi.org/10.17507/tpls.0907.20

Alibabaei, F. (2020). English translation of Persian metaphors in Iranian political news discourse: A relevance theory approach. *Journal of Language and Discourse Practice*, *1*(2), 1-10.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on nAutomatic Control, AC-19*, 716–723. http://doi.org/10.1109/TAC.1974.1100705

Angelelli, C. V., & Jacobson, H. E. (2009). Introduction. Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice. In *Testing and assessment in Translation and Interpreting Studies* (pp. 1-10). John Benjamins. http://doi.org/10.1075/ata.xiv

Angelone, E. (2013). The impact of process protocol self-analysis on errors in the translation product. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, *8*(2), 253-271. http://doi.org/10.1075/tis.8.2.07ang

Aryadoust, V. (2018). A cognitive diagnostic assessment study of the listening test of the Singapore–Cambridge general certificate of education O-level: Application of DINA, DINO, G-DINA, HO-DINA, and RRUM. *International Journal of Listening*, 1-24. https://doi.org/10.1080/10904018.2018.1500915

Austermuhl, F. (2014). *Electronic tools for translators*. Routledge. http://doi.org/10.4324/9781315760353

Beeby, A. (2000). *Teaching translation from Spanish to English.* University of Ottawa Press. https://doi.org/10.2307/j.ctt1cn6sgb.10

Beeby, A., Fernández Rodríguez, M., Fox, O., Kuznik, A., Neunzig, W., Rodríguez-Inés, P., & Wimmer, S. (2011). Results of the validation of the PACTE translation competence model: Translation project and dynamic translation index. In S. O'Brien (Ed.), *Cognitive explorations of translation* (pp. 30-53). Bloomsbury. http://doi.org/10.1075/btl.94.22pac

Birenbaum, M., Kelly, A. E., & Tatsuoka, K. K. (1993). Diagnosing knowledge states in algebra using the rule-space model. *Journal for Research in Mathematics Education*, *24*(5), 442-459. http://doi.org/10.2307/749153

Bonyadi, A. (2020). A skill-based approach for assessing EFL students' translation performance. *Journal of Language and Translation*, *10*(2), 35-46.

Budianto, L., & Fardhani, A. E. (2010). *A practical guide for translation skills*. UIN-Maliki Press.

Chen, B. (2013, December 6-8). *Globalization and film titles translation* [Paper presentation]. Proceedings of the international symposium on globalization: Challenges for translators and interpreters, Zhuhai, China.

Chen, H. L., & Chen, J. S. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly, 13,* 218–230. https://doi.org/10.1080/15434303.2016.1210610

Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement, 37*(6), 419-437. https://doi.org/10.1177/0146621613479818

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265-289. http://doi.org/10.2307/1165285

Conde, T. (2013). Translation versus language errors in translation evaluation. In D. Tsagari, & R. Vam Deemter (Eds.), Assessment issues in language translation and interpretation (pp. 97-112). Peter Lang AG.

Conde, T. (2012). Quality and quantity in translation evaluation: A starting point. *Across Languages and Cultures*, *13*(1), 67-80. http://doi.org/10.1556/Acr.13.2012.1.4

Davaninezhad, F. K. (2016). Supplementary practices and decreasing translation problems in translating political English news texts into Persian. *Theory and Practice in Language Studies*, *6*(1), 184-193. http://doi.org/10.17507/tpls.0601.24

De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76* (2), 179-199. http://doi.org/10.1007/s11336-011-9207-7

De la Torre, J. (2008). An empirically-based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45*(4), 343-362. http://doi.org/10.1111/j.1745-3984.2008.00069.x

De la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333-353. http://doi.org/10.1007/BF02295640

De la Torre, J., & Lee, Y. S. (2013). Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, *50*(4), 355-373. http://doi.org/10.1111/jedm.12022

De la Torre, J., & Lee, Y. S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement, 47*(1), 115-127. https://doi.org/10.1111/j.1745-3984.2009.00102.x

De la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa, 20*, 89–97. https://doi.org/10.1016/j.pse.2014.11.001

Dewi, H. D. (2015). *Comparing two translation assessment models: Correlating student revisions and perspectives* [Unpublished doctoral dissertation]. Kent State University.

DiBello, L. V., & Stout, W. (2007). Guest editor's introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement, 44*(4), 285-291. http://doi.org/10.1111/j.1745-3984.2007.00039.x

DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of Statistics, Psychometrics* (vol. 26, pp. 979-1030). Elsevier.

DiBello, L. V., Roussos, L. A., & Stout, W. (2006). 31a review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of statistics, 26*, 979-1030. http://doi.org/10.1016/S0169-7161(06)26031-0

DiBello, L. V., Stout, W. F., & Roussos, L. (1995). Unified cognitive psychometric assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-390). Erlbaum.

Doyle, M. S. (2003). Translation pedagogy and assessment: Adopting ATA's framework for standard error marking. *The ATA Chronicle*, *32*(11), 21-28.

Effatpanah, F. (2019). Application of cognitive diagnostic models to the listening section of the international English language Testing System (IELTS). *International Journal of Language Testing, 9*, 1-28. http://doi.org/10.1186/s40468-019-0090-y

Effatpanah, F., Baghaei, P., Boori, A. (2019). Diagnosing EFL learners' writing ability: A diagnostic classification modeling analysis. *Language Testing in Asia, 9*(12). https://doi.org/10.1186/s40468-019-0090-y

Embretson, S. E. (1999). Cognitive psychology applied to testing. In F. T. Durso (Ed.), *Handbook of applied cognition* (pp. 629–660). Wiley & Sons.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. [Empirical Study]. *Psychological Methods, 3*(3), 380-396. https://doi.org/10.1037/1082-989x.3.3.380

Embretson, S. E. (1983). Construct validity: Construct representation vs. nomothetic span. *Psychological Bulletin*, *93*(1), 179-197. https://doi.org/10.1037//0033-2909.93.1.179

Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika*, *84*(1), 19-40. http://doi.org/10.1007/s11336-018-09658-x

Farahzad, F. (1992). Testing achievement in translation classes. In C. Dollerup, & A. Loddegaard (Eds.), *Teaching translation and interpreting: training, talent, and experience* (pp. 271-278). John Benjamins. http://doi.org/10.1075/z.56.41far

Fišer, D. (2008). The teaching and learning of ICT skills for translators. In R. Dimitru & K. Freigang (Eds.), *Translation Technology in Translation Classes* (pp. 103-12). Institutul European.

Frederiksen, N., Mislevy, R. J., Bejar, I. I. (1993). *Test theory for a new generation of tests.* Erlbaum. https://doi.org/10.2307/2348147

Galeshi, R. (2012). *Cognitive Diagnostic model, a simulated-based study: Understanding Compensatory Reparameterized Unified Model (CRUM)* [Unpublished doctoral dissertation]. Virginia Tech.

Genesee, F. (2002). Evaluation. In R. Carter, & D. Nunan (Eds.), *The Cambridge guide to teaching english to speakers of other languages* (4th ed., pp. 144-150). Cambridge University Press.

Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education*. Cambridge University Press.

Gierl, M. J., Alves, C., & Majeau, R. T. (2010). Using the attribute hierarchy method to make diagnostic inferences about examinees' knowledge and skills in mathematics: An operational implementation of cognitive diagnostic assessment. *International Journal of Testing*, *10*(4), 318-341. http://doi.org/10.1080/15305058.2010.509554

Goff-Kfouri, C. A. (2005). Testing and evaluation in the translation classroom. *Translation Journal, 9*(2), 75-99.

Hansen, G. (2010). Translation 'errors'. *Handbook of translation studies*, *1*, 385-388. http://doi.org/10.1075/hts.1.tra3

Hartz, S. M. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practicality* [Unpublished doctoral dissertation]. University of Illinois.

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

Hattie, J., Jaeger, R. M., & Bond, L. (1999). Persistent methodological questions in educational testing. *Review of Research in Education, 24,* 393-446. http://doi.org/10.2307/1167275

Henson, R., Templin, J. L., & Douglas, J. (2007). Using efficient model-based sum-scores for conducting skills diagnoses. *Journal of Educational Measurement, 44*(4), 361–376. http://doi.org/10.1111/j.1745-3984.2007.00044.x

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using Log-Linear Models with latent variables. *Psychometrika, 74*(2), 191-210. http://doi.org/10.1007/s11336-008-9089-5

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Hu, J., Miller, M. D., Huggins-Manley, A. C., & Chen, Y.-H. (2016). Evaluation of model fit in cognitive diagnosis models. *International Journal of Testing, 16*(2), 119-141. https://doi.org/10.1080/15305058.2015.1133627

Hurtado Albir, A. (2015). The acquisition of translation competence. Competences, tasks, and assessment in translator training. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, *60*(2), 256-280. http://doi.org/10.7202/1032857ar

Irvine, S. H., & Kyllonen, P. C. (Eds.). (2002). *Item generation for test development.* Erlbaum.

Jang, E. E. (2005). *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL* [Unpublished doctoral dissertation]. University of Illinois.

Jiménez-Crespo, M. A. (2020). Localization. In E. Bielsa & D. Kapsaskis (Eds.), The Routledge handbook of translation and globalization (pp. 375-390). Routledge. http://doi.org/10.4324/9781003121848-31

Jiménez-Crespo, M. A. (2013). *Translation and web localization*. Routledge. http://doi.org/10.4324/9780203520208

Jun-song, W. A. N. G. (2017). Translation quality assessment: Bridge the gap between theory and practice. *Journal of Literature and Art Studies*, *7*(10), 1701-1706. http://doi.org/10.17265/2159-5836/2017.10.023

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272. https://doi.org/10.1177/01466210122032064

Kabát, M. (2020). The competencies of software translators from the point of view of Slovak translation studies. *Bridge: Trends and Traditions in Translation and Interpreting Studies*, *1*(1), 57-67.

Kafi, M., Khoshsaligheh, M., & Hashemi, M. R. (2018). Translation profession in Iran: current challenges and future prospects. *The Translator*, *24*(1), 89-103. http://doi.org/10.1080/13556509.2017.1297693

Kazemian, R., & Vasheghani Farahani, M. (2020). Translating metadiscourse features in TED political talks: A parallel corpus-based investigation of semantic underspecification in English and Persian. *Language and Translation Studies (JLTS)*, *53*(2), 127-154. https://doi.org/10.22067/lts.v53i2.85499

Kelly, D. (2005). *A handbook for translator trainers: A guide to reflective practice*. Sent Jerome Publishing.

Khanmohammad, H. O. M., & Osanloo, M. (2009). Moving toward objective scoring: A rubric for translation assessment. *Journal of English language Studies, 1*, 131-153.

Khatibzadeh, P., & Sameri, M. (2013). Translation of binomials in political speeches and reports; A contrastive study of English and Persian. *Journal of Translation and Interpretation, 6*(1), 18-33.

Khazaeefar, A., & Khoshsaligheh, M. (2014). Insights from EMT for translator training in Iran. *Ferdowsi Review: An Iranian Journal of TESL, Literature and Translation Studies*, *1*(1), 135-152.

Khoshsaligheh, M., Moghaddas, M., & Ameri, S. (2019). English translator training curriculum revisited: Iranian trainees' perspectives. *Teaching English Language*, *13*(2), 181-212. https://doi.org/10.22132/tel.2019.96238

Kim, A. Y. A. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, *32*(2), 227-258. http://doi.org/10.1177/0265532214558457

Kiraly, D., & Hofmann, S. (2016). Towards a postpositivist curriculum development model for translator education. In D. Kiraly, L. Rüth, C. Canfora, & A. Cnyrim (Eds.), *Towards authentic experiential learning in translator education* (pp. 67–87). Mainz University Press at V&R Unipress. https://doi.org/10.14220/9783737004954.67

Klimkowski, K. (2015). Entrepreneurial training in translator and interpreter education. *Roczniki humanistyczne*, *63*(11), 67-83. http://doi.org/10.18290/rh.2015.63.11-5

Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models. *Journal of Educational Measurement, 49*(1), 59–81. https://doi.org/10.1111/j.1745-3984.2011.00160.x.

Lauscher, S. (2000). Translation quality assessment: Where can theory and practice meet? *The translator*, *6*(2), 149-168. http://doi.org/10.1080/13556509.2000.10799063

Lee, Y. W., & Sawaki, Y. (2009a). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6*, 239–263. https://doi.org/10.1080/15434300903079562

Lee, Y. W., & Sawaki, Y. (2009b). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, *6*(3), 172–189. http://doi.org/10.1080/15434300902985108

Leighton, J. P., & Gierl, M. J. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and practices*. Cambridge University Press. http://doi.org/10.1017/CBO9780511611186

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's Rule-Space Approach. *Journal of Educational Measurement, 41*(3), 205-237. http://doi.org/10.1111/j.1745-3984.2004.tb01163.x

Leighton, J., Gierl, M. J., & Hunka, S. (1999, April). *Attributes in Tatsuoka's rule-space model.* Poster presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec, Canada.

Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 9*, 17-46.

Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, *33*(3), 391-409. https://doi.org/10265532215590848.

Lim, Y. (2015). *Cognitive diagnostic model comparisons* [Unpublished doctoral dissertation]. Georgia Institute of Technology.

Lorenzen, B., & Murray, L. L. (2008). Bilingual aphasia: A theoretical and clinical review. *American Journal of Speech-Language Pathology, 17*(3). https://doi.org/10.1044/1058-0360(2008/026)

Mackenzie, R. (2004). The competencies required by the translator's roles as a professional. In Malmkjær, K. (Ed.), *Translation in undergraduate degree programs* (pp. 32-38). John Benjamins Publishing Company. http://doi.org/10.1075/btl.59.04mac

Marczak, M., & Krajka, J. (2016). Translator education in the cloud: Students' perceptions of telecollaborative experiences. In E. Smyrnova-Trybulska (Ed.), *E-learning methodology-implementation and evaluation* (pp. 369-388). Uniwersytet Slaski.

Maris, E. (1999). Estimating multiple classification latent class models, *Psychometrika*, *64*(2), 197-212. http://doi.org/10.1007/BF02294535

Martínez-Carrasco, R. (2017). *Epistemological approaches to legal translation education: A situated account* [Unpublished doctoral dissertation]. Universitat Jaume I.

Maydeu-Olivares, M. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives, 11*, 71-137. https://doi.org/10.1080/15366367.2013.831680

McDonald, R. P., & Mok, M. M. C. (1995). The goodness of fit in item response models. *Multivariate Behavioral Research*, *30*, 23-40. https://doi.org/10.1207/s15327906mbr32-001

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Macmillan.

Molanazar, H., & Kamyab, S. (2015). Political and journalistic translation courses and the market demand in Iran. *European Online Journal of Natural and Social Sciences,4*(2), 285-293.

Mossop, B. (2003). What should be taught at translation school? In Pym, A., C. Fallada, J. Ramon Biau, & J. Orenstein (Eds.). *Innovation and E-learning in translator training* (pp. 20-22). Universitat Rovira i Virgili.

Newmark, P. (1988). *A Textbook of Translation*. Prentice Hall.

Nichols, P. (1994). A framework of developing cognitively diagnostic assessments. *Review of Educational Research, 64*, 575-603. http://dx.doi.org/10.3102/00346543064004575

Nichols, P. D., Chipman, S. F., & Brennan, R. L. (1995). *Cognitively diagnostic assessment*. Erlbaum. https://doi.org/10.4324/9780203052969

Nichols, P., & Sugrue, B. (1999). The lack of fidelity between cognitively complex constructs and conventional test development practice. *Educational Measurement: Issues and Practice, 18*, 18-29. https://doi.org/10.1111/j.1745-3992.1999.tb00011.x

Nord, C. (1999). Translating as a text-production activity. *On-Line Symposium on Innovation in Translator and Interpreter Training, Interpreter Training*, 3(4), 11-19.

Ozaki, K. (2015). DINA models for multiple-choice items with few parameters: Considering incorrect answers. *Applied Pychological Measurement*, *39*(6), 431-447. http://doi.org/10.1177/0146621615574693

PACTE. (2011). Results of the validation of the PACTE translation competence model: Translation problems and translation competence. In C. Alvstad, A. Hild, & E. Tiselius (Eds.). *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 317–343). John Benjamins. https://doi.org/10.1075/btl.94.22pac

PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation: Perspectives in process oriented research* (pp. 43-66). John Benjamins. https://doi.org/10.1075/btl.45.06pac

Paradis, M., Goldblum, M. C., & Abidi, R. (1982). Alternate antagonism with paradoxical translation behavior in two bilingual aphasic patients. *Brain and Language*, *15*(1), 55-69. http://doi.org/10.1016/0093-934X(82)90046-3

Piecychna, B. (2013). Legal translation competence in the light of translational hermeneutics. *Studies in Logic, Grammar, and Rhetoric, 34*(47), 141–159. http://doi.org/10.2478/slgr-2013-0027

Polliastri, P., & Paulina, A. (2009). Evaluation criteria for the improvement of translation quality. In M. Forstner, H. Lee-Jahnke, & P. A. Schmitt (Eds.). *Enhancing translation quality: Ways, means, methods* (pp. 239-261). Peter Lang AG.

Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, *34*, 782-799. https://doi.org/10.1177/0734282915623053.

Ravand, H., & Baghaei, P. (2020). Diagnostic classification models: Recent development, practical issues, and prospects. *International Journal of Testing*, *20*, 24-56. https://doi.org/10.1080/15305058.2019.1588278

Ravand, H., Baghaei, P., & Doebler, P. (2020). Examining parameter invariance in a general diagnostic classification model. *Frontiers in Psychology*, *10*, 2930. https://doi.org/10.3389/fpsyg.2019.02930

Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology, 38*(10), 1255–1277. https://doi.org/10.1080/01443410.2018.1489524.

ReaDickins, P., & Germaine, K. (1993). Evaluation. Oxford University Press.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications.* Guilford Press. https://doi.org/10.1080/15366367.2018.1434349

Samir, A., Motallebzadeh, K., Ashraf, H., & Ghonsooly, B. (2018). Development and validation of an MA English translation curriculum effectiveness questionnaire: Using the Rasch Model. *North American Journal of Psychology*, *20*(3), 601-618.

Samir, A., Tabatabaee-Yazdi, M. (2020). Translation quality assessment rubric: A Rasch model-based validation. *International Journal of Language Testing, 10*(2), 101-128.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464. https://doi.org/10.1214/aos/1176344136

Sharif, D. F. (2016). Evaluating the translation competency areas in the Iranian national translators' training curriculum*. International Journal of Innovation and Research in Educational Sciences, 5*(1), 340-349.

Sidiropoulou, M. (2004). *Linguistic identified through translation*. Rodopi. http://doi.org/10.1163/9789004486652

Snell-Hornby, M., Pochhacker, F., & Kaindl, K., (Eds). (1994). *Translation studies: An interdiscipline (Benjamins Translation Library 2)*. John Benjamins Publishing. http://doi.org/10.1075/btl.2

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263-331). American Council on Education, Macmillan.

Stobart, G., & Gipps, C. (1997). *Assessment: A teacher's guide to the issues* (3rd ed.). Hodder & Stoughton.

Stone, C. A., & Zhang, B. (2003). Assessing goodness of fit of item response theory models: A comparison of traditional and alternative procedures. *Journal of Educational Measurement*, *40*(4), 331-352. http://doi.org/10.1111/j.1745-3984.2003.tb01150.x

Su, Y. L. (2013). *Cognitive diagnostic analysis using hierarchically structured skills* [Unpublished doctoral dissertation]. University of Lowa. https://doi.org/10.17077/etd.5xs1u621

Tabatabaee-Yazdi, M., Motallebzadeh, K., Baghaei, P. (2021). A Mokken scale analysis of an English Reading Comprehension Test. *International Journal of Language Testing, 11*(1), 132-143.

Tabatabaee-Yazdi, M. (2020). Hierarchical Diagnostic Classification Modeling of reading comprehension. *SAGE Open*, *10*(2). http://doi.org/10.1177/2158244020931068

Taghizadeh, M., & Azizi, M. (2017). Exploring computer-aided translation competencies of some Iranian translators. *International Journal of English Language & Translation Studies*, *5*(1), 78-87.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345-354. http://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, *30*, 251-275. https://doi.org/10.1007/s00357-013-9129-4

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287-305. https://doi.org/10.1037/1082-989x.11.3.287

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*, 125–145. https://doi.org/10.1177/014662168400800201

Waddington, C. (2001). Different methods of evaluating student translations: The question of validity. *Meta: Journal Des Traducteurs/Meta: Translators' Journal*, *46*(2), 311-325. https://doi.org/10.7202/004583ar

Wang, C., Zheng, C., & Chang, H. H. (2014). An enhanced approach to combine item response theory with cognitive diagnosis in adaptive testing. *Journal of Educational Measurement*, *51*(4), 358-380. https://doi.org/10.1111/jedm.12057

William, K. (2003). *Understanding media theory*. Arnold.

Yamaguchi, K., & Okada, K. (2018). Comparison among cognitive diagnostic models for the TIMSS 2007 fourth grade mathematics assessment. *PIOS ONE*, *13*(2). http://doi.org/10.1371/journal.pone.0188691

APPENDIX A

Assessment Criteria	Attributes								
	linguistic skills in the SL		Linguistic Skills in TL						
	1. SL Lexical	2. SL Grammatical	3. TL Lexical	4. TL Grammatical	5. Organizational Knowledge	6. Cultural Knowledge	7. Translation skills	8. Work methodology skills	9. Bibliography and technical skills
Grammar (Word Form/Part of Speech, Word Order, Syntax)	0	1	0	1	0	0	1	0	0
Usage	1	1	1	1	1	1	0	0	0
(No) Addition	1	1	1	1	1	1	1	0	0
(No) Omission	1	1	1	1	1	1	1	0	0
Completeness	0	0	0	0	1	0	1	0	0
Punctuation	0	1	0	1	0	0	0	0	0
Terminology/False Friend Terminology	1	0	1	0	0	0	1	0	0
Spelling	0	0	1	0	0	0	0	0	0
Capitalization/ Italicization Rules	1	1	1	1	0	0	1	0	0
Faithfulness/literalness	1	1	1	1	1	1	1	0	0
Register/Tone	1	1	1	1	1	1	1	0	0
Genre (Text Style, Text type)	1	1	1	1	1	0	1	0	0
Cohesion/Coherence, Consistency	0	0	1	1	1	0	1	0	0
Accuracy	1	1	1	1	1	1	1	0	0
Fluency)Naturalness, Readability, No Ambiguity)	1	1	1	1	1	1	1	0	0
Creativity/Problem Solving (No Indecision)	1	0	1	1	1	1	1	0	0
17. Organization/time management	0	0	0	0	0	0	0	1	0
18. Initiative	0	0	0	0	0	0	0	1	0
19. Pace of work	0	0	0	0	0	0	0	1	0
20. Revision file, self-assessment	0	0	0	0	0	0	0	1	0
21. Quality of terminological data base	0	0	0	0	0	0	0	0	1
22. CAT skills	0	0	0	0	0	0	0	0	1
23. Relevance of bibliography	0	0	0	0	0	0	0	0	1

APPENDIX B

Items	Rate by the Raters				Rate by the Researchers	
	0		1		0	1
	1 (poor)	2 (Fair)	3 (Advanced)	4 (Superior)	Mastered	Non-Mastered
Grammar (Word Form/ Part of Speech, Word Order, Syntax …)						
Usage						
(No) Addition						
(No) Omission						
Completeness						
Punctuation						
Terminology/False Friend Terminology						
Spelling						
Capitalization/ Italicization Rules						
Faithfulness/literalness						
Register/Tone						
Genre (Text Style, Text type)						
Cohesion/Coherence, Consistency						
Accuracy						
Fluency)Naturalness, Readability, No Ambiguity)						
Creativity/Problem Solving (No Indecision)						
17. Organization/Time management						
18. Initiative						
19. Pace of work						
20. Revision file/Self-assessment						
21. Quality of terminological data base						
22. CAT skills						
23. Relevance of bibliography						
Total						