

# Critical Analysis of Translation Tests in 18 Specialized Translation Courses for Undergraduate Students

Reima Al-Jarf

## ABSTRACT

The present study aimed to describe and evaluate the current assessment practices prevalent in the different translation courses offered at the College of Languages and Translation (COLT). A sample of specialized translation final exams in 18 translation subject areas was collected. Each final exam was analyzed in terms of the following: (1) # of English and Arabic source texts included on each final exam for each course (2) readability and difficulty level of texts included in the translation exams, (3) # of exams with a terminology subtest, (4) English and Arabic text length in words, (5) reliability, validity and discriminating power of final exams. Data analysis showed that 50% of the exams included one English text, 32% included 2 texts and 18% included 3 texts. Arabic texts were included in 73% of the exams. However, 59% of the exams included one Arabic text, 9% included two texts and 5% included 3 texts. In Addition to English and Arabic texts, 56% of the exams included a vocabulary subtest. 41% do not have any Arabic texts. The English text length ranged between 66-430 words with a median length of 181 words. The Arabic text length ranged between 26-180 with a median length of 97. The typical Flesch Reading Ease of English texts was 40 and the typical Flesch-Kincaid Grade level score was 11. There were no significant differences among the different college levels nor different subject areas in text length or text difficulty level. Translation exams currently used at COLT lack validity, reliability, and discriminating power. Some reasons for lack of reliability and validity are given. Students' views on translation exams are also reported. A model for more valid, reliable, and discriminating translation exams is given with students views of it as well.

**Keywords:** translation assessment, translation test content analysis, translation test discrimination test power, translation test reliability and validity

**Published Online:** June 11, 2021

**ISSN:** 2736-4534

**DOI :**10.24018/ejedu.2021.2.3.86

**Reima Al-Jarf**

King Saud University, Riyadh, Saudi Arabia.

(e-mail: reima.al.jarf@gmail.com)

## I. INTRODUCTION

Assessment of translation competence in performing translation tasks by student-translators is an integral part of translation pedagogy and translator training. To assess translators' competence, several types of standardized and teacher-made translation test are being used. However, the question remains how good those tests are in assessing translators' competence, what they consist of, which skills they focus on, in addition to psychometric issues such as reliability, and validity and discrimination power. A review of the literature has shown numerous studies that focused on standardized translation tests such as: The state German to Finnish translation test and certificate in Finland which is both a second language and translation examination [1]; and the GITIS (Graduate Institute of Translation and Interpretation Studies), a battery of admission tests used at Fu Jen Catholic University, Taiwan to screen out candidates with sufficient ability in their language pairs. The test battery is composed of a five-part written test, oral group-tests, and

individual tests before an international panel of experts [2]. Another test is the CATTI (China Accreditation Test for Translators and Interpreters) test battery), which is the most authoritative translation and interpretation proficiency qualification accreditation test in China for measuring competence in translation and interpreting between Chinese and English, Arabic, German, French, Spanish, Russian, or Japanese, in domains such as academia, business, media and government. CATTI is divided into four levels: Senior I, II, and III. The total test time is two hours for translation proficiency, one hour for interpreting proficiency, three hours for translation practice, one hour for interpreting practice for Levels I and II, and half an hour minutes for interpreting practice for Level III [3]; [4].

Further studies focused on identifying other types of translation assessment tasks some of which are use of controlled free-response test items to assess several aspects of translation at once [5]; a Listening Summary Translation Exam in Taiwanese to evaluate the summary translation ability of applicants who want to work as linguists in law

enforcement agencies in the US with focus on authenticity of task [6]; and an open-ended translation test from English to Japanese, and a multiple-choice translation test [7].

Few more studies used web-based testing. For example, [8] recommended using a professional portfolio to assess students' professional competences. This professional portfolio helps translation students define their own general translation competences and set future career goals, and become familiar with the translation market rates. [9] used the dynamic online system with automated scoring and intelligent feedback for non-English majors' translation exercises and self-tests. [10] conducted interpreting tests live synchronously and concurrently with multiple candidates, using web-based synchronous cyber classrooms. These tests are based on the accreditation test for professional interpreters utilized by the National Accreditation Authority of Translators and Interpreters in Australia. The tests are comprised of dialogue interpreting, consecutive interpreting, sight translation, and questions on ethical issues. [11] utilized the Calibrated Parsing Items Evaluation, that maximizes translators' performance through identifying the parsing items with an optimal p-docimology and item discrimination. This method checks all the possible parses (annotations) in the source text by means of the Brat Visualization Stanford CoreNLP software. The Calibrated Parsing Items Evaluation takes a step towards the objectification of translation assessment by allowing evaluators to assess impacts of the items in the source texts via docimologically justified parsing items.

A second line of research focused on criteria that have to be taken into consideration in translation assessment. Those criteria included: (i) employing a psychometrically based approach to the development of translation tests [2]; (ii) the source texts should be authentic, self-contained, comprehensible, and not previously covered [5]; (iii) authenticity of the translation task (Wu and Stansfield, 2001) [6]; (iv) specifying the types of knowledge required in text comprehension and translation which include linguistic, encyclopaedic, interactive, metacommunicative, and global textual knowledge [1]; (v) reliability, validity, practicality, and fairness of the translation test [4]; [12], [13]; [14]; [15]; [16]; [17]; [7]; [18], construct, criterion-related and concurrent validity [19]; [12]; [20]; (vi) objectivity and scorability [20]; (vii) accuracy of translation tests as competence evaluation rather than source/target text comparisons [21]; and (viii) allowing the students to bring two bilingual paper dictionaries, but no electronic devices [3].

A third line of research investigated translation assessment weaknesses. [22] pointed out problems of setting criteria for translation quality control and assessment of translations from Russian, which combines different approaches to translation assessment. [23] asserted that the translation profession has not achieved reliability, validity, objectivity, practicability, or consensus in defining and evaluating adequacy in translation and that it is impossible to have a framework for assessing translations. He called for international standards for translation adequacy. The translation testing procedure has been criticized for its subjective character. No real steps have been made so far towards developing an objective translation test [20]. [24] asserted that the content of translation tests should depend

primarily on the aim of the course. For example, tests that measure students in metalinguistic-awareness-oriented courses should focus on items attesting to metalinguistic knowledge rather than a mere competence in the language per se.

Despite the multitude of studies in the literature that have investigated numerous aspects of translation assessment, there is lack of studies in Saudi Arabia that explore how teacher-made specialized achievement translation exams at colleges of languages and translation are constructed. The author did not find any studies that analyzed the linguistic aspects of specialized translation exams such as the source text length, its difficulty level, and translation speed required. She did not find any studies that analyze the psychometric aspects of specialized translation tests such as reliability, validity, and discrimination power. Therefore, the current study aims to analyze, describe and evaluate the specialized translation final exams developed by translation instructors at the College of Languages and Translation (COLT), King Saud University, Riyadh, Saudi Arabia in terms of: (i) the linguistic aspects of translation final exams such as the number of English and Arabic source texts on the exams, English and Arabic source text length in words and the total exam length in words, percentage of exams with English and Arabic terminology subtests, the difficulty level of the English source texts, and translation speed in word per hour required; (ii) the psychometric aspects which include reliability, validity, and discrimination power of the translation exams used; (iii) availability of test instructions and what they tell the students; (iv) whether the translation exams comply with the objectives of the translation program at COLT, the skills they need to acquire during the program and tasks they need to be able to perform after graduation.

In addition, the current study aims to answer the following questions: Are there significant differences among the different college levels and exams for the different subject areas in: (a) the English and Arabic source text length, (b) the Flesch Ease Scores for the English source texts, (d) the Flesch-Kincaid Readability Grade Level for the English source texts and (f) translation speed.

Results of the present study will be based on a content and statistical analysis of specialized final exams in 18 subject areas at COLT. The ease scores and readability grade levels were not computed for the Arabic texts as they are not available.

This study is significant as it provides a framework for assessing specialized translation courses in numerous subject areas, taking into consideration psychometric standards for developing translation tests and essential elements of translation tests. It shows translation instructors at COLT how to create authentic translation tests that mirror the translation reality and objectives of teaching translation at COLT, which aspects of the translation final exams deserve more attention from translation instructors, whether there is agreement between the teacher-made translation final exams and the translation program goals, and the role of authenticity in translation test development in general.

## II. METHODOLOGY

### A. Curriculum, Material and Tasks

The translation program at COLT is a B.A. program that is 5 years or 10 semesters long. Each semester is called level. In the first 2 years (4 semesters), the students take Listening, Speaking, Reading, Writing, Grammar, and Vocabulary Building courses (20 hours per week per semester). In semesters 5-10, they take 6 Linguistics, 6 Interpreting courses, 2 Computer Applications in Translation courses, a Problems in Translation course, and 18 specialized translation courses distributed as follows:

- 1) Level 5: Natural Sciences, Humanities.
- 2) Level 6: Islamic, Medical, Media, Administrative, Engineering, Military.
- 3) Level 7: Sociology, Politics, Educational.
- 4) Level 8: Security, Computer Science.
- 5) Level 9: Petroleum, Legal, Agricultural, Literary.

In addition, the students complete a translation project that is 25,000 words long (100 pages) from English to Arabic or Arabic to English.

There is no textbook for the translation courses. Each course instructor is free to choose the texts to be used for in-class translation practice by the students. The students choose the book that they would like to translate for their Project.

As for assessment at COLT, the students take 2 written interm tests and a final exam. 50% of the total course mark is allocated to semester work and 50% to the final exam. The passmark is 60%. Two hours are allocated to the final exam. The students are allowed to use specialized and general monolingual and bilingual paper dictionaries. In each course, the translation interm tests and final exams are developed by the course instructors. The tests usually consist of 1 or more English and/or Arabic source texts. Instructors are free to choose the source texts to be included on the tests in terms of topic, length, difficulty level and tasks. Instructors are also free to include a terminology subtest or not. No source text selection or scoring criteria are imposed on the instructors by the college.

### B. Study Samples

- 1) Eighteen final exams for 18 specialized translation courses (Natural Sciences, Humanities, Islamic, Medical, Media, Engineering, Military, Sociology, Administrative, Politics, Educational, Security, Computer Science, Petroleum, Legal, Agricultural, Literary) were collected from translation instructors at COLT.
- 2) The final exam test scores for students enrolled in those 18 specialized courses were obtained from the course instructors, and the course letter grades and number of students who got an A, B, C, D and F in those 18 translation courses were obtained from the Registration Department at KSU.
- 3) A sample of 90 students was randomly selected from the 18 translation courses and was surveyed to find out what they think of the translation tests at COLT, their strengths, and shortcomings.

### C. Data Analysis

The English and Arabic source texts included in the final exams were entered into MS WORD. The Microsoft WORD

readability statistics were used. For each source text included on each final exam, the author computed the following:

- 1) the total number of English and Arabic source texts on each final.
- 2) the number of exams that include a terminology subtest.
- 3) the English, Arabic and total source text length in each final.
- 4) the Flesch Reading Ease score and the Flesch-Kincaid Grade level score for each English source text only as the formulas cannot be applied to Arabic texts.
- 5) the translation speed in words per hour for each exam by dividing the total number of words in the texts by 120 minutes which is the total test time.
- 6) analysis of Variance (ANOVA) to find out whether there are significant differences among the different college levels and different subject areas in the source text length, source text ease score and readability grade level.
- 7) the mean, median, standard deviation, standard error, range, and internal consistency reliability coefficient for each final exam.
- 8) the percentage of students who got an A, B, C, D, E, and F in each translation course in order to show the discrimination power of the exams.

In addition, the author analyzed the content of the source texts on each final exam and the instructions and compared the translation tasks with the program goals. Students' responses to the survey questions were analyzed and are reported qualitatively.

## III. RESULTS

### A. Description of test content

Data analysis showed that 56% of the translation exams contain 2 texts, 11% contain 1 text only, 22% contain 3 texts and 11% contain 4 texts regardless of the language of the source text. Specifically, 56% contain 2 English texts, 39% contain 1 English text, and 5% contain 3 English texts. As for Arabic, 41% of the exams have no Arabic texts, 56% have 1 Arabic texts, and 3% have 2 Arabic texts. 44% of the final exams do not contain a terminology subtest, 56% contain a terminology subtest (11% English only, 22% Arabic only, and 23% both English and Arabic).

Regarding the text length, Table 1 indicates that the English source text length ranged between 66-430 words, with a median length of 181 words. The Arabic source text length ranged between 26-180, with a median length of 97. There is no gradation in the total source text length according to the college level, and no uniformity in total text length within the same level.

It is evident in Table 1 that the Ease Score for all the exams ranged between 26.9 to 78.2, with a typical Ease Score of 47.5. 56% of the exams are difficult or very difficult to read, 27% are fairly difficult to read, and 17% are fairly easy to read (See Table 2). The higher the score, the easier the text and the lower the score, the more difficult the text. There is no gradation in the source text ease score from one college level to the next, nor among the different exams for the same level. The ease score for a low college level can be low (texts are difficulty) and for a high college level can be high (texts are easy).

As for the text difficulty level, the Flesch-Kincaid Readability Grade Level Score for the texts in Table 1 showed that the text difficulty level ranged between 5.5-12,

with a median of 11. 28% of the exams are very easy (Grades 5-9 readability), and 82% are fairly difficult to read (Grades 10-12 readability) (See Table 3). Even for higher college levels, the texts included on the exams do not match the expected college readability level, especially because the students are being prepared to be professional translators and they are expected to read and translate long authentic texts and not short, simplified ones.

Regarding translation speed, Table 1 shows that the typical final exam requires the students to translate 139 words per hour, with a range of 66 to 430 words per hour which is not satisfactory at all. Here again, there is no gradation in the required translation speed between the college levels and even within the courses for the same college level. This means that the translation exams do not test students' ability to translate fast. Developing students' ability to translate fast is necessary for enabling them to handle the bulk of texts they need to translate within a limited amount of time when they work as professional translators in the future.

TABLE 1: ENGLISH AND ARABIC TEXT LENGTH, FLESCH READING EASE SCORE, FLESCH-KINCAID READABILITY GRADE LEVEL AND TRANSLATION SPEED IN WORDS PER HOUR

College Levels	Translation Courses	Text Length			English Source Text Readability		Trans. Speed in Words Per Hour
		English	Arabic	All	Flesch Reading Ease Score	Flesch-Kincaid Read. Grade Level	
5	Humanities	87	-	87	50.1	9.8	44
5	Natural Sci	430	-	430	73.4	6.9	215
6	Admin.	156	176	332	26.9	12	166
6	Military	181	26	207	47.3	10.1	104
6	Medical	280	-	280	51.4	10.3	140
6	Engineer.	204	176	380	53.4	9.5	190
6	Media	370	180	550	53.7	11	275
6	Islamic	105	40	145	78.2	5.5	123
7	Sociology	335	132	467	53.5	10.4	234
8	Security	203	56	259	29.9	12	130
8	Ed.	66	85	151	30.6	12	126
8	Computer	338	-	338	30.6	12	169
8	Political	164	97	261	42.4	12	131
8	Commerc.	115	97	212	72.7	7	106
9	Legal	175	81	256	37.8	12	128
9	Petrol.	333	142	475	42	11.5	238
9	Agri.	169	98	267	44.3	11.7	134
9	Literary	229	116	345	47.6	12	173

TABLE 2: INTERPRETATION OF THE FLESCH READING EASE SCORE<sup>1</sup>

Score	Interpretation
90-100	The text is very easy to read, easily understood by an average 11-year-old student
80-90	The text is easy to read
70-80	The text is fairly easy to read
60-70	The text is easily understood by 13- to 15-year-old students
50-60	The text is fairly difficult to read
30-50	The text is difficult to read, best understood by college graduates
0-30	The text is very difficult to read, best understood by university graduates

TABLE 3: INTERPRETATION OF THE FLESCH-KINCAID GRADE LEVEL SCORE<sup>2</sup>

Score	School level (US)	Notes
100-90	Grade 5	The text is very easy to read. It is easily understood by an average 11-year-old student.
90-80	Grade 6	The text is Easy to read. It is written in conversational English for consumers.
80-70	Grade 7	The text is Fairly easy to read.
70-60	Grade 8 & 9	The text is in easily understood by 13- to 15-year-old students.
60-50	Grade 10-12	The text is fairly difficult to read.
50-30	College	The text is difficult to read.
30-10	College graduate	The text is very difficult to read. It is best understood by university graduates.
10-0	Professional	The text is extremely difficult to read. It is best understood by university graduates.

*B. Differences in Exam Text Length, Difficulty Level and Translation Speed*

Analysis of Variance of (ANOVA) indicated no significant differences among the different college levels in the total test length nor the English and Arabic text length separately (F =1.25; DF=17; Mean = 267). Similarly, ANOVA showed significant differences among the different levels and the different subject areas in the Flesch Reading Ease Score (F=1.87; df =17; Mean = 47) and in the Flesch-Kincaid Grade Level Score (F = 2.3; Df = 17; Mean = 11). Finally, Analysis of Variance of (ANOVA) showed no significant differences among the different college levels and different subject areas in translation test speed (F = 2.3; Df = 17; Mean = 134 words).

*C. Reliability of the Translation Exams*

Table 4 reveals that the internal consistency reliability coefficient for the 18 translation exams ranges between .08 to .64, with a median reliability coefficient of .30. It also reveals that 76% of the courses have a reliability coefficient below .50, and 41% have a reliability coefficient below .20. This means that the reliability coefficient of the 18 exams is low because the exam texts are short, the number of texts on the exam is small, the score variance is small (See Table 4), and the exam score range (difference between the highest and lowest score) is small. The latter reflects variability of the scores among students in a particular course.

<sup>1</sup> <https://yoast.com/flesch-reading-ease-score/>

<sup>2</sup>[https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid\\_readability\\_tests](https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests)

TABLE 4: THE MAN, MEDIAN, STANDARD DEVIATION RANGE, VARIANCE AND RELIABILITY COEFFICIENT FOR THE TRANSLATION FINAL EXAM SCORES ONLY\*

College Levels	Translation Courses	Mean	Median	Standard Deviation	Range	Variance	Reliability Coefficient
5	Human.	88.39	89	5.14	22	23.65	0.18
7	Commerc.	77.53	77	6.67	24	44.47	0.13
9	Computer	84.56	83	6.67	24	44.47	0.14
7	Political	86.59	86	5.44	26	29.56	0.08
6	Engineer.	77.05	80	5.84	27	34.16	0.48
8	Ed.	74.48	74	8.21	28	67.33	0.55
7	Socio.	74.31	75	6.21	29	38.55	0.30
6	Admin.	71.61	70	7.54	30	56.81	0.21
8	Security	70.59	72	9.26	30	56.81	0.47
9	Legal	71.69	70	9.39	30	88.19	0.56
6	Islamic	79.02	78	6.05	31	36.63	0.18
9	Agri.	70.45	70	9.7	31	94.09	0.56
9	Petrol.	74.14	74	6.53	31	42.59	0.17
6	Military	73.02	73	10.17	32	103.45	0.64
6	Medical	80.09	78	8.23	34	67.73	0.34
6	Media	78.69	75	9.57	34	91.54	0.44
9	Literary	77.05	78	7.9	34	62.91	0.14
5	Natural	78.77	79	9.62	38	92.48	0.44

\* Interm I and interm II marks are not included

#### D. Validity of the Translation Tests

Analysis of the source texts in the 18 final exams showed that some translation texts lack face, content validity, authenticity and naturalness of the translation tasks chosen as in the following examples: (i) Some exams contained a terminology subtest which required the students to give the equivalent of each terms in isolation not in context. (ii) The text given was a dictionary definition. (iii) Some exams required the translation of single sentences, not a long text. (iv) One test contained true/false questions about information given in the course. (v) Some exams gave the students 2 texts to choose from. The two texts were not comparable in genre, length, ease score and readability grade level. (vi) There is overlap among subject area exams in text genre. The text given on the Natural Science translation exam and Petroleum exam were from biology. (vii) The Literary translation exam text was about how poetry should be translated, not a literary excerpt such as a poem, play or novel to be translated by the students. (viii) The translation tasks required by the exams do not match the tasks required for the translation project in which the students translate a long authentic text (a whole book), or the authentic texts to be translated when on the job after graduation. (ix) On some exams, the students' background knowledge affects the translation of some texts as in the Intifada text which the students can translate because the topic is familiar, not because they understand the ideas in the text. (x) The texts lack variety in difficulty level and length. (xi) Lack of adequate content coverage. For example, in the Natural Sciences translation course, the students translate chemistry, physics, biology, meteorology and astronomy texts. But the final exam contains a biology text only and the other areas are ignored. (xiii) The test length and difficulty levels do not increase from one college level to the next. (xiv) 8 English texts and 1 Arabic text do not have a title to help the students understand the overall topic of the text.

#### E. Test Discrimination Index

One of the important characteristics of a good is that it should discriminate between students who have mastered and

those who have not mastered the skills under study. A good translation test should also have a high discrimination power especially because students at COLT are going to translate a book in the translation project and will be translating authentic texts when they work as translators after graduation. In this respect, the translation exams under study do not have a high discriminating power. The distribution of the final exam course marks in Table 4 shows that the typical student scored 75% on the final exam and the typical difference between the highest and the lowest scores in the course is 30%. In 96% of the courses, all the students passed the course. In the other 4%, very few students failed the final exam. Moreover, the distribution of letter grades for the 18 translation courses displayed in Table 5 indicate that 6% of the students got an A (range 0 to 20); 31% got a B (range 17 to 77%); 31% got a C (range 4.5% to 45%); 20% got a D (range 0 to 44%); 1% got an F, i.e., failed the course (range 0 to 13%). In 35% of the courses, nobody failed the course. In other words, the percentage of students who got an A and B combined is more than 50% in 6 courses, more than 60% in 4 courses; more than 70% in 3 courses; and between 80% to 96% in 4 courses. This means that the translation exams in the different courses are too easy and exams are skewed right and do not sort out students properly.

TABLE 5: DISTRIBUTION OF LETTER GRADES IN THE TRANSLATION COURSES\*

Translation Courses	N	A	B	C	D	F
		90-100 Exc.	80-89 Very Good	70-79 Good	60-69 Below Average	Below 60 Failing
Natural	163	0%	17%	33%	35.5%	13%
Human.	164	4%	77%	19%	0%	0%
Media	101	2%	22%	4.5%	26.5%	4%
Engineer.	106	1%	35%	24.5%	32%	7.5%
Islamic	98	3%	24%	42%	20%	10%
Military	111	5%	41%	27%	25%	2%
admin	101	13%	47%	29%	12%	0%
Medical	105	12%	62%	20%	4%	2%
Commerc.	95	11%	47%	32%	10%	0%
Political	93	9%	27%	35%	20%	9%
Socio.	111	8%	46%	41%	5%	0%
Security	75	4%	19%	33%	44%	0%
Computer	89	6%	31%	45%	18%	0%
Ed.	106	2%	70%	26%	2%	0%
Literary	90	12%	24%	32%	31%	1%
Legal	87	7%	28%	31%	26%	8%
Petrol.	101	9%	30%	35%	26%	1%
Agri.	80	20%	63%	11%	5%	1%

\* The course grade is the sum of interm I, interm II and final exams marks

#### F. Test Instructions

As revealed by the data analysis, 67% of the exams do not have any instructions to the students to define the type and characteristics of the translation output. 33% have brief instructions that tell the students in which direction the translation should be (from English to Arabic or Arabic to English), to translate the underlined sentences in the text only, or choose one text form 2 or 3 given texts.

#### G. Exam Tasks and Translation Program Objectives

The translation program objectives do not specify what the students should be able to translate in each level in terms of text length, difficulty level and translation speed. That is why, translation instructors have no guide to follow. Each designs her course exams at her own discretion.

### H. Students' Views

Students' responses to the surveys revealed several shortcomings of the translation assessment procedures. The students indicated that the translation tests at COLT help them pass the translation courses with good grades but do not help them acquire good translation skills. They never learnt to translate fast. They learnt to translate slowly even when the exam texts are easy. Sometimes they spend 2 hours translating 8 lines, as they spend a lot of time checking the meanings of the words in the text in the dictionary, even those they already know. They pointed out that the test texts lacked variety in content and difficulty level. On some tests, the texts were a lot harder than those discussed in class. Some other exams contained few texts and sometimes they only translated few underlined sentences in a text, not the whole text. Sometimes the text on the exam is similar to a text discussed in class. Sometimes the exam in a higher level is easier than an exam in a lower level. The translation program does not tell them the length of the texts that they need to translate in each college level. Some students commented:

Ameera: I relied heavily on the dictionary and since we have plenty of time, I look up even the words that I already know.

Sana: At the beginning of the test session, I feel that I have plenty of time, so I spend too much time looking up most of the words in the text in the dictionary. Thus, I run short of time.

Sara: Having plenty of time encouraged me to revise my translation several times and make too many corrections. This way I ended up with a target text that deviated from the source text.

Alia: Since we could use all kinds of dictionaries during the test session, I never worried about learning new terms.

Maha: Having too much time encouraged some students to cheat.

Dalal: Having a terminology test on the interm tests and even on the final exam helped me get a good mark in the course.

Maryam: I feel nervous and confused if I am given a long text to translate and do not know how to handle long texts.

Hala: We translate all kinds of texts in the same way which literal translation. We have to translate every single word in the text even in a medical and scientific texts and lose marks if we do not.

## IV. DISCUSSION AND RECOMMENDATIONS

Findings of the present study indicated that the specialized translation exams currently used at COLT have many shortcomings and they do not meet the criteria of a good test in terms of reliability, validity, discrimination power, authenticity, variability, and other test features mentioned by prior researchers such as [4]; [12]; [19]; [7]; [13]; [14]; [6]; [16]; [15]; [2]; [5]; [1]; [17]; [18]; [20]; [21]; [3].

To improve the quality of translation exams at COLT, the study recommends the following: In designing translation tests for COLT students, instructors should take into consideration what students are expected to do in the program (translation project), and after they graduate. A translation test should be both a power test and a speed test. It should discriminate among those who have and have not acquired the translation skills. The final exam length should increase from one college level to the next in the number of texts and source text length. The test should consist of at least 3 pages

(4 different texts) covering several areas of the subject field studied in the course. The test should contain several texts that vary in topic and difficulty level. For example, the Engineering translation exam should include a variety of texts of different lengths from different engineering specialties: mechanical, electrical, civil, chemical, aerospace, architectural engineering, and others. The students should be trained to translate at least a page (500 words) an hour, with minimal reliance on the dictionary. Translation instructors teaching different translation courses need to coordinate with each other to avoid any overlap in the material covered on the tests and to control the length and difficulty level of the tests for the different college levels, so that exams are graded in difficulty. Each text should have a title. The test instructions must specify the direction of translation, the kind of translation the students are supposed to render (summary, full, conceptual, free translation... etc.), and the aspects of the translation output emphasized such as organization, layout, grammar, cohesion, punctuation, spelling, not only meaning. To help the students to produce a cohesive and coherent translation output, and to help them infer the meaning of difficult words from context, focus should be on the meaning, not the exact words of the source text, especially in scientific and technical texts. The students need to learn how to infer the meanings of unknown words from context to reduce their heavy reliance of the dictionary, because in a real translation situation, sometimes the translator has no access to a dictionary. In addition, consulting the dictionary every now and then is time consuming, the student loses focus on the overall meaning and will pay attention to single words. When the students have to translate a long text in a limited time, this will reduce dictionary use.

Following the above guidelines will help the students learn to read the text quickly, identify the difficult words quickly and look up few words only in the dictionary. The students will develop inferential comprehension skills. They will learn to translate quickly and efficiently and finish the translation on time, no matter how long the test is. They will learn to focus on the meaning, not words. Sentences in the translation output will be cohesive and organized. They will learn to write more carefully and efficiently making fewer spelling and grammatical mistakes. Above all, they will be able to handle the tasks they are expected to perform in the translation project and in their future job.

## REFERENCES

- [1] L. Piitulainen, "The State Translation Examination in Finland: Language or Translation Test?" *Intercultural German Studies*, no. 24, pp. 291-302. 1998.
- [2] E. A. Arjona-Tseng, "Psychometric Approach to the Selection of Translation and Interpreting Students in Taiwan," in *Bridging the Gap: Empirical Research in Simultaneous Interpretation*, S. Lambert, and B. Moser-Mercer Eds, Amsterdam: John Benjamins, 1994, pp. 69-86.
- [3] H. Zhao, and X. Gu, "China Accreditation Test for Translators and Interpreters (CATTI): Test Review Based on the Language Pairing of English and Chinese," *Language Testing*, 33, 3, pp. 439-446. 2016.
- [4] M. Zou, and W. Wu, "Review on China Accreditation Test for Translators and Interpreters (CATTI)" *English Language Teaching*, vo. 8, no. 7, pp. 152-156. 2015.
- [5] F. Farahzad, "Testing Achievement in Translation Classes," in *Teaching Translation and Interpreting: Training, Talent and Experience*, C. Dollerup and A. Loddegaard, Eds. Amsterdam: John Benjamins Publishing, 1992, pp. 271-278.
- [5] W. M. Wu, and C. W. Stansfield, "Towards Authenticity of Task in Test Development," *Language Testing*, vol. 18, no. 2, pp. 187-206. 2001.

- [7] A. A. Ito, "Two Types of Translation Tests: Their Reliability and Validity," *System: An International Journal of Educational Technology and Applied Linguistics*, vol. 32, no. 3, 395-405. 2004.
- [8] A. Galán-Mañas, "Professional Portfolio in Translator Training: Professional Competence Development and Assessment," *Interpreter and Translator Trainer*, vol. 13, no. 1, pp. 44-63. 2019.
- [9] Y. A. Tian, "A Dynamic Online System for Translation Learning and Testing," EUROCALL Conference, Southampton, United Kingdom, Aug 23-26, 2017.
- [10] N. S. Chen, and L. Ko, "An Online Synchronous Test for Professional Interpreters," *Educational Technology & Society*, vol. 13, no. 2, pp. 153-165. 2010.
- [11] A. Akbari, and M. Shahnazari, "Calibrated parsing items evaluation: A Step towards Objectifying the Translation Assessment." *Language Testing in Asia*, vol. 9, Article 8, 2019.
- [12] R. S. Al-Jarf, "What Teachers Should Know about vocabulary testing," International Conference on Language Testing and Assessment. Guangzhou, China. November 27-30. 2015.
- [13] R. S. Al-Jarf, "Linguistic and measurement considerations in Translation tests," 13<sup>th</sup> World Congress of the Association Internationale de Linguistique Appliquee (AILA). Singapore, December 16-21, 2002a.
- [14] R. S. Al-Jarf, "Reflections on Translation Assessment," American Association of Applied Linguistics (AAAL) Conference. Salt Lake City, Utah, April 6-9. 2002b.
- [15] R. S. Al-Jarf, "Issues in Translation Assessment," 5<sup>th</sup> CTELT Annual Conference on Teaching, Learning and Assessment, Dubai, United Arab Emirates, May 9-10, 2001.
- [16] C. Waddington, "Different Methods of Evaluating Student Translations: The Question of Validity," *Meta*, vol. 48, no. 2. 2001.
- [17] G. Turover, "Criteria for Evaluating Translation," *Sendebur*, no. 7, pp. 281-286. 1996.
- [18] C. W. Stansfield, W. Wu and C. C. Liu, "Listening Summary Translation Exam (LSTE) in Taiwanese (Also Known As) Minnan, Southern Fukienese, Southern Min, Xiamen, Amoy," Final Project Report. ERIC No. ED413788. 1997.
- [19] S. M. Alavi and H. Ghaemi, "Reliability assessment and construct validation of translation competence questionnaire (TCQ) in Iran," *Language Testing in Asia*, vol. 3, Article 18. 2013.
- [20] B. Ghonsooly, "Development and Validation of a Translation Test," *Edinburgh Working Papers in Applied Linguistics*, vol. 4, pp. 54-62. 1993.
- [21] S. J. Campbell, "Towards a Model of Translation Competence," *Meta*, vol. 36, np. 2-3, 329-343. 1991.
- [22] A. A. Aubakirova, "Nurturing and Testing Translation Competence for Text-Translating," *International Journal of Environmental and Science Education*, 11, 11, 4639-4649. 2016.
- [23] G. McAlester, "The Evaluation of Translation into a Foreign Language," in *Developing Translation Competence*, C. Schaffner and B. Adab, Eds, Amsterdam: John Benjamins, 2000, pp. 229-241.
- [24] M. Shlesinger, "What Are We Testing When We Test Translation Students?" *English Teachers' Journal*, no. 49, pp. 38-40. 1996.



**Reima Al-Jarf** is professor of ESL, ESP, linguistics and translation at King Saud University, Riyadh, Saudi Arabia. She has 700 publications and conference presentations in 70 countries. Some of her articles are published in ISI and Scopus journals. She reviews Ph.D. theses, promotion works, conference and grant proposals, and articles for numerous peer-reviewed

international journals including some ISI and Scopus journals. She won 3 Excellence in Teaching Awards, and the Best Faculty Website Award at her university. Her areas of interest are: Foreign language teaching and learning, technology integration in education and translation studies.