



Validating a Test to Measure Translation Teachers' Assessment Literacy

Ghasem Modarresi

Department of English, Quchan Branch, Islamic Azad University, Quchan, Iran

Kaveh Jalilzadeh

Department of Foreign Languages, Istanbul University, Cerrahpaşa, Turkey

Christine Coombe

Higher Colleges of Technology, Dubai Men's College, UAE

Asieh Nooshab

Department of English, Quchan Branch, Islamic Azad University, Quchan, Iran

Background

Teachers typically allocate at least one-third of their class time to testing and assessing students (Bachman, 2014). The term assessment literacy implies what language teachers and instructors are expected to know about assessment matters (Stoynoff & Chapelle, 2005). In this regard, Malone (2008) emphasized the application of assessment knowledge to classroom practices. Assessment literacy is increasingly being regarded as an indispensable part of teacher education and professionalism (Coombe et al., 2012). According to Inbar-Lourie (2008), assessment literacy refers to teachers' abilities to realize the social role of assessment. Meanwhile, Scarino (2010) highlighted the role of teachers' conceptions of assessment literacy.

In their systematic review, Gotch and French (2014) listed 36 instruments designed for evaluating assessment literacy. However, the number of teachers who lack sufficient knowledge of or training in assessment are many (Lam, 2015). For this reason, Stiggins (2010) maintained, "assessment illiteracy abounds" (p. 233). To lessen the problem, Xu and Brown (2016), in their seminal work, reviewed 100 reliable publications on assessment literacy and demonstrated the need for reconceptualizing teacher assessment literacy because of its complexity.

The call for systematic and objective translation evaluation traces back to the 1959 FIT conference, where the theme was translation quality (Wilss, 1998). From then on, there has been much development in Translation Quality Assessment (TQA) to come up with more systematic and objective assessment scenarios. Research on TQA shows that teachers run into problems while assessing a translation product (Williams 2004; Modarresi & Ghoreyshi 2018), and there is not much accord among scholars and practitioners on the assessment criteria (Hatim & Mason, 1997). Scholars in Translation Studies have not



sufficiently addressed the concept of assessment literacy. Recently, the necessity of research in translation assessment has been emphasized (Angelelli & Jacobson, 2009; Martínez Melis & Hurtado Albir, 2001).

To address this problem, new ways of exploring teachers' knowledge of assessment are needed. Indeed, that teachers have adequate knowledge of assessment in translation studies can be determined by means of validated tools in the form of a standardized test not based on observation, interviews or questionnaires since none of them could evaluate translation teachers' knowledge comprehensively and accurately. According to Clifford (2007), in measuring teachers' knowledge of translation assessment, one should define and validate the constructs or traits being measured. Although different models have been proposed by scholars in translation (e.g., Beeby, 2000; Kim, 2009), they were developed to measure translation ability objectively. For example, Angelelli (2009) developed and validated a scale to allow for assessment of interactional competence ranging from Advanced to Poor. Or, Martínez-Mateo et al. (2016), adopting a mixed bottom-up top-down approach to quality assessment, presented a new approach for TQA. More recently, in the Iranian context, Khorami and Modarresi (2019), employing sophisticated statistical analyses, developed and validated a rubric for consecutive interpreting performance, and Ashrafi et al. (2013) developed and validated a test for translation teacher competency to measure the translation competence focusing on the teacher.

Taken together, the present study adopted the theoretical framework suggested by Angelelli (2009) to design a test using a multiple-choice format. For Angelelli (2009), a multiple-choice format is appropriate since it can (1) cover a variety of topics to be tested effectively, (2) provide opportunities for different translation teachers since most of them are acquainted with the multiple-choice format, (3) cope with item difficulty, and (4) provide reliable scores that are easily calculated. Therefore, to achieve this goal, the present study, as an initial attempt, intended to develop and validate a standardized instrument for measuring the assessment literacy of translation teachers.

Method

Participants

The study opted for the criterion sampling method to select the participants for the study and in doing so specified the following criteria: a) teaching English at the university level, b) Holding a PhD degree or being a PhD student in Translation Studies, English Language Teaching, English Literature, or Linguistics, and c) Having experience of training translation courses for at least five years at the university level. As such, a pool of 20 translation teachers (males: $n = 13$, 65%; females: $n = 7$, 35%; Mean age = 47.25, SD = 8.24) were selected to participate in the interview phase, and moreover, a total of 159 translation teachers (males: $n = 95$, 59.7%; females: $n = 64$, 40.30%; Mean age = 43.79, SD = 7.09) participated in the validation process based on convenience sampling from a number of Iranian universities.

Instruments

To collect data for the qualitative phase of the study, a semi-structured interview protocol guide with four questions was developed by the researchers to seek out the conceptions of translation teachers regarding what shape translation assessment needed to take as well as the knowledge and skills needed. They were asked to respond to the questions in a face-to-face interview. After that, a test of translation teacher assessment literacy was designed and validated by the researchers. The validated test included two major sub-components with 20 items in a multiple-choice format.

Data Collection

To develop the test, the researchers initially collected data from 20 translation teachers to solicit their ideas about translation assessment literacy, and this information was collected over a three-week period. The contents of the questions mainly centered on participants' knowledge and conceptions of testing and assessment, their attitudes towards objective and subjective evaluation of translations and the role of social factors and teacher identity in assessment literacy in practice, following the conceptualization developed by Xu and Brown (2016).

Initially, the question items were constructed based on the information collected from the interviews and the related literature on the issue. Second, for the initial piloting of the items, the researchers asked two individuals specialized in testing and assessment, to pinpoint the problems with the content and clarity of the items. Third, to obtain feedback from the initial pilot group, the researchers conducted the final piloting phase, during which the relevant test was administered to 159 translation teachers over a three-month period. The first section of the test included demographic information. The question items were written in the English language in a multiple-choice format. The following is an example of the test items intended to test the teachers' language assessment literacy (the asterisk * indicates the correct response):

- As a form of assessment, portfolios in assessment are.....
- a) considered as an important self-correction technique
- b) mostly encouraged in process-oriented translation
- c) a purposeful selection of students' works*
- d) a procedure for cooperative test construction

The following is an example of the test items designed to assess translation assessment literacy:

- To assess objectively the translation competency of translation teachers working at the university level, the only rubric that has been developed in the Iranian context was developed by Ashrafi et al. (2013). The test is mainly based on the theoretical framework suggested by
- a) Kelly's (2008) facets of translator competency*
- b) PACTE's (2005) translation competence
- c) Bell's (1991) description of sub-competencies
- d) Campbell's (1991) model of translation competence

Data Analysis

To analyze the qualitative data, the researchers opted for a theme-based procedure (Dörnyei, 2007) so as to classify the commonalities obtained from the interviews. The inter-coder agreement was taken care of, and the inter-coder reliability for coded transcripts was calculated. To assess the inter-coder agreement, the researcher followed a negotiated agreement approach to compare their codings and discuss their disagreement so as to reconcile and arrive at a final version (Campbell et al., 2013). As explained by Krippendorff (2004), the inter-coder reliability refers to reproducibility across coders in an effort to determine if coders would code the same data the same way. It is worth mentioning that the inter-coder agreement requires that the two coders reconcile through discussion (Garrison et al., 2006), and the inter-coder reliability requires that the two coders choose the same code for the same unit of text (Krippendorff, 2004). As for the construct validity of the relevant test, first, the Rasch measurement model, as an appropriate model for dealing with categorical data like answers to multiple-choice questions, was used to see if the test of translation teacher assessment literacy fits into a Rasch model of test performance (Bond & Fox, 2007). Moreover, Factor Analysis and SEM were implemented to validate the newly developed translation assessment literacy.

Results

Initially, the researchers outlined the major common themes that emerged from the responses regarding translation teachers' needs for assessment literacy in the Iranian context. To assure the inter-coder reliability, the researchers coded the responses by eliciting the commonalities, and they also formulated rather similar findings with minor differences. For example, the three common themes elicited from the interviews, including teacher knowledge, teacher thought, and teacher belief, were coded as teacher cognition by the researchers. Following the guidelines suggested by Campbell et al. (2013), the number of coding agreements were divided by the number of agreements and disagreements combined, and they achieved 68 per cent inter-coder reliability. There were 45 common themes that at least one of the coders invoked a code, and of these, there were 31 cases that all of coders had invoked the code. Therefore, the level of inter-coder reliability was 68 per cent ($31/45 = .68$). However, after negotiating discrepancies, the study reached 80 per cent inter-coder reliability ($25/31 = .80$). Therefore, following the coding reliability and agreement, the number of common themes was reduced to 24 codes which are as follows: Construct validity, reliability, assessment rubric, assessment technique, error analysis, translation competence, translation mistakes, House's (2015) functional-pragmatic model, teacher cognition, syntactic errors, Kim's (2009) meaning-oriented assessment, lexical errors, language mistakes, fluency, contextual factors, Waddington's (2001) model of translation quality, interpretation assessment criteria, think-aloud protocol, raters, PACTE's (2005) translation competence, Beeby's (2000) translation assessment rubric, assessment of localized software products, Wilss' (1998) translation assessment criteria, and screen capture technology.

Following this, to determine the underlying traits of the test, the researchers opted for Rasch analysis to validate the test of translation teacher assessment literacy. To this end, the researchers employed Rasch analysis to confirm the unidimensionality of the test. In so doing, WINSTEPS Version 3.63 was used by the researchers (Linacre, 2009). Moreover, Exploratory Factor Analysis (EFA) was used to reveal the sub-components of the major trait since Rasch measurement just reveals the major trait. Following this, Confirmatory Factor Analysis was also performed to confirm the results obtained by means of EFA. The results of Rasch measurement displayed an item separation index of 3.18 with item reliability of .91, and a person separation index of 3.11 with person reliability of .73, which indicated a quite precise measurement.

As illustrated in Table 1, all items fit the Rasch model, except for items 12 and 14 since their infit mean square indices were outside the acceptable boundary of 0.70-1.30, following the guidelines proposed by Bond and Fox (2007). To be more exact, these items were signs of multi-dimensionality and model deviance because their indices were .68 and .64, respectively. Consequently, after running Rasch analysis, the test was reduced to 23 items. After that, to determine the hidden factors of the test, Factor Analysis was run via SPSS Software Version 22.

The researchers first checked the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) value, ranging from 0 to 1, and a minimum value of .6 is needed for good Factor Analysis (Tabachnick, & Fidell, 2001). The KMO value was .351, and the Barlett's Test of Sphericity value was significant ($p = .000 < .05$) so that there was no problem to run factor analysis. After that, the pilot study was done to validate the test through which the newly-made test, consisting of 23 items, was administered to 159 participants. Principal component analysis extracted six factors that had eigenvalues of more than 1.0, and accounted for 68% of the variance. Twenty-one items had loadings of 0.30 or greater on any factor. In other words, items 1 and 21 were not found to have loadings of 0.30 or higher on any factor. Therefore, these items were removed from the test. Thus, the validated test included 21 items. The researchers then opted for the Scree test to discover the number of factors for rotation, and the Scree test represented a two-factor solution as a more parsimonious grouping of items. Finally, oblique rotation represented the underlying factor structure based on which the first factor included 13 items, and the second factor included eight items. Indeed, after factor rotation, the whole bank of items consisted of 21 items. As the final step of the construct validation process, the results of Amos 20 displayed a good fit for the data. Some modifications were carried out on the model because some measurement models did not display adequacy to the data (see Figure 1). Altogether, one item

was removed, including one item from factor one because of low loading. Indeed, after modifications were made, the goodness-of-fit of the model enhanced significantly. χ^2/df was 2.54, less than the cut-off point of 3; RMSEA was .066, less than .08; and GFI, CFI, and TLI were .92, .91, and .91, respectively, which were above the suggested cut-off point of .90 (Tseng & Schmitt, 2008).

TABLE 1
Item Statistics and Fit Statistics

Entry number	Total score	Total count	Measure	Model S.E.	Infit MNSQ	ZSTD
23	13	75	1.88	.31	1.13	.7
25	24	75	1.05	.25	1.18	1.9
22	53	75	-.64	.26	1.15	1.3
8	37	75	.30	.24	1.17	3.8
18	50	75	-.45	.25	1.10	1.2
19	31	75	.63	.24	1.08	1.3
7	36	75	.35	.24	1.09	2.1
3	39	75	.19	.24	1.09	2.0
9	51	75	-.51	.25	1.03	.3
4	43	75	-.04	.24	1.00	0
10	28	75	.81	.24	.95	-.7
13	12	75	1.98	.32	.97	-.1
5	49	75	-.39	.25	.96	-.5
16	57	75	-.92	.27	.96	-.2
17	54	75	-.71	.26	.94	-.5
20	54	75	-.71	.26	.94	-.5
6	66	75	-1.78	.36	.93	-.2
15	36	75	.35	.24	.93	-1.7
11	50	75	-.45	.25	.92	-.9
2	54	75	-.71	.26	.91	-.7
21	23	75	1.11	.25	.91	-.9
24	57	75	-.92	.27	.86	-.6
1	49	75	-.93	.25	.75	-1.2
12	32	75	.58	.24	.68	-2.1
14	53	75	-.64	.26	.64	-1.2
MEAN	42.0	75.0	.00	.26	1.00	.1
P. SD	14.1	.0	.89	.03	.09	1.3

As shown in Figure 1, the associations between the factors and the items were satisfactory since they were more than .30 and less than .90. As for the association for main factors, the two factors had an acceptable correlation of .68. Consequently, the final validated test comprised two factors and 20 items. Pondering the content of the items, the researcher named these new factors as follows: Factor one: *Translation Assessment Literacy*, and Factor two: *Testing and Assessment Knowledge* (see Appendix A).

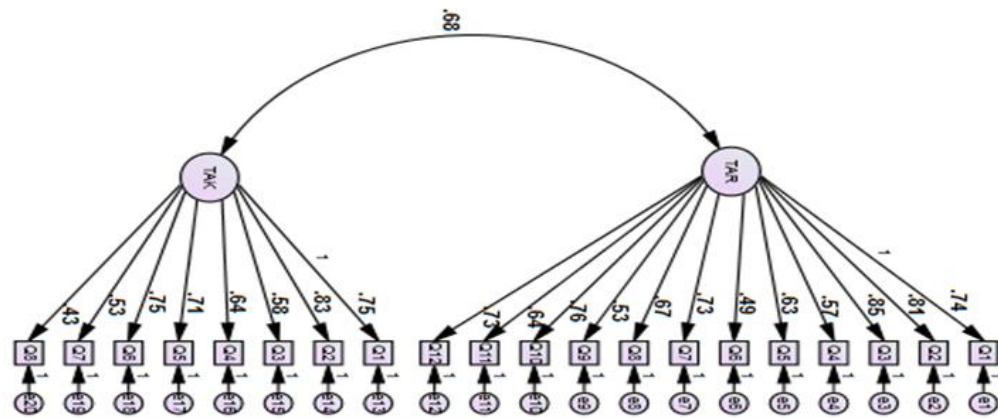


Figure 1. The results of SEM (Translation assessment literacy, and testing and assessment knowledge).

Discussion and Conclusion

As an embarking endeavor, the present study tried to design and validate a test of assessment literacy for translation teachers through the Rasch Model and Confirmatory Factor Analysis. The study is in line with the previous informative suggestions proposed by scholars in the field of translation studies such as Muñoz (2012) who stated, “we gained much insights into translators’ mental life, but there has been very little construct-validating research” (p. 170). The SEM analysis results also revealed a large positive association between knowledge of testing and assessment and translation assessment literacy ($r = .68$), indicating that those who were more knowledgeable in assessment literacy performed better on translation assessment literacy. However, previous research on translation assessment quality concentrated on evaluation rubrics (e.g., Farahzad, 1992; Goff-Kfourri, 2004) with no standardized test to measure teachers’ knowledge and skills in assessment literacy. Now that the teacher performs a key role in enhancing the educational quality of translation programs, teachers’ knowledge of translation assessment could be evaluated by means of the validated test. Precise measurement would enhance students’ translation quality since the teacher can pinpoint their mistakes, which would increase their motivation and lead to their active participation. Previous literature emphasizes the role that knowledge of assessment and testing could play in educational success (Earl, 2013; Popham, 2014).

The present study reinforces the conclusion that the provision of translation assessment tests is a fruitful task since the movement in evaluating translation products from a subjective, intuitive perspective to an objective, logical perspective has led the researchers to focus on evaluating translation students based on assessment models and rubrics. However, research has unfortunately indicated that teachers tend to make use of traditional assessment methods rather than more innovative assessment models that focus on contextual factors (Tsgari, 2021). As Waddington (2001) concludes, translation evaluators can find the quality of translation in quantifying the quality, and the researchers of the present study concluded that providing students with knowledge of assessment rubrics helps them enhance their translation competence since they would know exactly how their works was being judged and evaluated by the teacher.

Major implications are drawn from the present study for translation teachers and students. Translation teachers are recommended to acquire the major concepts of testing and assessment and try to evaluate students’ translations based on standards rather than relying on their plausibility and intuitions. It seems that teachers’ inadequate knowledge of assessment in translation is due to the fact that the number of sources on testing and assessment in translation and interpreting studies is too few (e.g., Angelelli, 2009; House, 2015), and to the best of our knowledge, there is no comprehensive textbook that provides translation evaluation models and rubrics for teachers in the field. They can indeed assess students’ translations both qualitatively and quantitatively, taking a mixed bottom-up, top-down approach to quality assessment (Martínez-Mateo et al., 2016). Successful teachers reflect on assessment literacy in practice. As for the translation students, they are expected to ask their teachers to clarify how their translations are being assessed, that is, based on what rubrics or standards. Indeed, as Modarresi (2019) concluded, “getting involved in translation has a positive relationship with translation ability, and this, in turn, would reinforce translation students’ tendency to become competent and professional in translation” (p. 240).

Overall, the integration of assessment principles and practice could mediate students’ progress not only at the end of the course but also during the training courses. Indeed, as for the translation teachers’ assessment literacy, in line with the suggestion by Khorami and Modarresi (2019), we are in the beginning stages of a challenging process to design and validate required measurement tools in the field of translation, and now the door is open for carrying out further research in this regard since assessment literacy could include broader topics such as the mediating role of social factors and knowledge of emerging new techniques in assessment such as computer-aided assessment and dynamic assessment (see Modarresi & Alavi, 2014) in order to establish a more globally inclusive picture of this area in relation to translation and interpreting.

Acknowledgement

We are grateful to the participants who took part in our study.

The Authors

Ghasem Modarresi (corresponding author) is Assistant Professor of Applied Linguistics at IAU of Quchan. His main interests include Psychology of Education, Testing and Assessment, and Applied Translation Studies.

Department of English, Quchan Branch,
Islamic Azad University, Quchan, Iran
Islamic Azad University of Quchan, Quchan-Mashhad km 4, Quchan, Iran
Postal Code: 9479176135
Tel: +98 51 47011001-9
Email: qasem.modarresi@gmail.com

Kaveh Jalilzadeh is a Lecturer in TESOL at the University of Cerrahpaşa, Istanbul. His research interests are Language Assessment, Research Methodology and Teacher Development.

Department of Foreign Languages,
Istanbul University-Cerrahpaşa, Turkey
İstanbul University-Cerrahpaşa, Avcılar Campus, Turkey.
Postal Code: 34320
Tel: +90 212 4040701
Email: kjalilzadeh1983@gmail.com

Christine Coombe is Associate Professor of General Studies at Dubai Men's College, Higher Colleges of Technology, UAE. Her main interests are Testing and Assessment, Research Methods, TBLT, Teacher Effectiveness, Teacher Leadership and Management and Positive Psychology and Well-being.

Higher Colleges of Technology, Dubai Men's College, UAE
Dubai - United Arab Emirates
Tel: +971 50-6194796
Email: ccoombe@hct.ac.ae

Asieh Nooshab has an MA in Translation Studies from IAU of Quchan. Her main interests include Translation and Interpreting Studies.

Department of English, Quchan Branch,
Islamic Azad University, Quchan, Iran
Islamic Azad University of Quchan, Quchan-Mashhad km 4, Quchan, Iran
Postal Code: 9479176135
Tel: +98 51 47011001-9
Email: nooshab.1989@yahoo.com

References

Angelelli, C. V. (2009). Using a rubric to assess translation ability. In C.V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice* (pp. 13-47). John Benjamins.

- Angelelli, V. C., & Jacobson, E. H. (2009). *Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice*. John Benjamins.
- Ashrafi, N., Pishghadam, R., & Hashemi, M. R. (2013). Construction and validation of the translation teacher competency test. *Translation Studies*, 11(43), 66-81.
- Bachman, L. F. (2014). Ongoing challenges in language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 3, pp. 1586-1603). John Wiley and Sons.
- Beeby, A. (2000). *Teaching translation from Spanish to English*. University of Ottawa Press.
- Bell, R. T. (1991). *Translation and translating*. Longman.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum.
- Campbell, S. (1991). Towards a model of translation competence. *Meta* 36(2), 329-343.
- Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semi-structured interviews: Problems of unitization and inter-coder reliability and agreement. *Sociological Methods & Research*, 42(3), 294-320.
- Clifford, A. (2007). Grading scientific translation: What's a new teacher to do? *Meta* 52(2), 376-389.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges and recommendations. In C. Coombe, P. Davidson, O'Sullivan, & S. Stoyonoff (Eds.), *The Cambridge guide to second language assessment* (pp. 20-29). Cambridge University Press.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
- Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning* (2nd ed.). Corwin.
- Farahzad, F. (1992). Testing achievement in translation classes. In C. Dollerup & A. Loddergard (Eds.), *Teaching translation and interpreting* (pp. 271-278). John Benjamins.
- Garrison, D. R., Cleveland-Innes, M., Koole, M., & Kappelman, J. (2006). Revisiting methodological issues in transcript analysis: Negotiated coding and reliability. *The Internet and Higher Education*, 9, 1-8.
- Goff-Kfoury, C. A. (2004). Testing and evaluation in the translation classroom. *Translation Journal*, 8(3), 7-13.
- Gotch, C. M., & French, B. F. (2013). Elementary teachers' knowledge and self-efficacy for measurement concepts. *The Teacher Educator* 48(1), 46-57.
- Hatim, B., & Mason, I. (1997). *The translator as communicator*. Routledge.
- House, J. (2015). *Translation quality assessment: Past and present*. Routledge.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402.
- Kelly, D. (2008). Training the trainers: Towards a description of translator trainer competence and training needs analysis. *TTR*, 21(1), 99-125.
- Khorami, F., & Modarresi, G. (2019). A Rasch-based validation of the evaluation rubric for consecutive interpreting performance. *Sendebare*, 30, 221-244.
- Kim, M. (2009). Meaning-oriented translation assessment. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 123-157). John Benjamins.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197.
- Linacre, J. M. (2009). *Winsteps* (Version 3.68) [Computer Software]. Winsteps.com. <https://www.winsteps.com/winsteps.htm>
- Malone, M. (2008). Training in language assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education: Language testing and assessment* (pp. 225-239). Springer Media.

- Martínez-Mateo, R., Montero-Martínez, S., & Moya-Guijarro, A. J. (2016). The modular assessment pack: A new approach to translation quality assessment at the Directorate General for translation. *Perspectives*, 17, 1-31.
- Martínez Melis, N. & Hurtado Albir, M. (2001). Assessment in translation studies: Research needs. *Meta* 46, 272-287.
- Modarresi, G. (2019). Developing and validating involvement in translation scale and its relationship with translation ability. *Forum: International Journal of Interpretation and Translation*, 17(2), 225-248.
- Modarresi, G., & Alavi, S. M. (2014). Designing and validating a test battery of computerized dynamic assessment of grammar. *TELL*, 14(2), 1-29.
- Modarresi, G., & Ghoreyshi, S. V. (2018). Student-centered corrections of translations and translation accuracy: A case of BA translation students. *Translation Studies*, 15(60), 11-28.
- Muñoz, M. R. (2012). Cognitive and psycholinguistic approaches. In M. C. Millán & F. Bartrina (Eds.), *Handbook of translation studies* (pp. 241-256). Routledge.
- PACTE (2005). Investigating translation competence: Conceptual and methodological issues. *Meta*, 50(2), 609-619.
- Popham, W. J. (2014). *Classroom assessment: What teachers need to know* (7th ed.). Pearson Education.
- Scarino, A. (2010). Assessing intercultural capability in learning languages: A renewed understanding of language, culture, learning, and the nature of assessment. *The Modern Language Journal*, 94(2), 324-329.
- Stiggins, R. J. (2010). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 233-250). Taylor & Francis.
- Stoyonoff, S., & Chapelle, C. (2005). *ESOL tests and testing: A resource for teachers and program administrators*. TESOL.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Harper Collins.
- Tsagari, D. (2021). *Language assessment literacy: Concepts, challenges, and prospects*. In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 13-32). Routledge.
- Tseng, W. T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning*, 58, 357-400.
- Waddington, C. (2001). Should translations be assessed holistically or through error analysis? *Hermes Journal of Linguistics*, 26, 12-31.
- Williams, M. (2004). *Translation quality assessment: An argumentation centered approach*. University of Ottawa Press.
- Wilss, W. (1998). Decision making in translation. In M. Baker (Ed.), *Routledge encyclopedia of translation studies* (pp. 57-60). Routledge.
- Xu, Y., & Brown, J. T. L. (2016). Teacher assessment literacy in practice: A conceptualization. *Teaching and Teacher Education*, 58, 149-162.

(Received October 16, 2021; Revised November 20, 2021; Accepted December 18, 2021)