Abstract: This study was an initiative to investigate the extent to which synthesizing Transformative Education (TE) with Dynamic Assessment (DA) can contribute to developing EFL learners’ productive skills; speaking and writing. To do so, 105 Iranian university students majoring in English language were randomly assigned into four different classes, as control and experimental groups. The control groups were instructed based on pure TE-oriented approach, whereas the experimental groups were instructed based on the integration of TE with DA. Given the nature of the research questions, only the speaking and writing post-tests’ data meeting the required normality assumptions were analyzed based on parametric statistical approach (i.e. Independent Samples t-tests). The analyses revealed a significant difference between control and experimental groups in terms of both speaking and writing performances. So, TE integrated with DA is suggested as an innovative trend in foreign language education. The main implication might be for the practices in higher education to incorporate classroom activities such as critical thinking and class projects in their instructional delivery so that students would have the chance to participate in the teaching-learning process.

ABOUT THE AUTHOR

Azadeh Zarbafian, PhD candidate in TEFL at IAU, South Tehran Branch, Tehran, In Iran. Her research focuses on ELT methods and language assessment. She has published both in (inter) national journals & attended similar conferences.

Dr. Gholam-Reza Abbasian, is an assistant professor at Imam Ali University, Tehran, Iran. He offers Language Assessment, Research Methods & SLA courses at PhD program also as his research areas. He has published and presented more than 15 books, supervised many theses and dissertations, and published locally and internationally.

Dr. Ahmad Mohseni, associate professor at IAU, South Tehran Branch, English Dept., Iran. His research focuses on L2 teacher education. He has published and presented papers (inter)nationally. Abdollah Baradaran, associate professor at IAU, Central Tehran Branch, English Dept., Iran. His research interest is language teaching and teacher education. He has presented and published papers (inter)nationally.

PUBLIC INTEREST STATEMENT

The present research was conducted to improve the learning experience and the productive skills of the EFL learners through synthesizing Transformative Education with Dynamic Assessment, hoping that the results and findings would have practical implications for the EFL teachers. Two groups participated in the current research and the findings revealed that integration of TE with DA is an innovative trend in foreign language education which results in a significant difference between control and experimental groups at performing both speaking and writing skills.
1. Introduction
Transformative Education (TE) theory of Mezirow (1978) considers learning as a process within which the learners critically examine their world’s framework. It considers knowledge gain and fundamental changes in the learner’s beliefs, values, and perspectives within the process of human development as two critical aims of education. Based on TE principles, assessment is bifurcated with instruction as TE gets along with the notion of paradigm shift from testing to assessment. This is rooted in the notion that various forms of assessment, such as self-, peer-, authentic, dynamic, and in particular the formative one, facilitate the process of feedback exchange between the teacher and learning (Moran, 1997, & Hargreaves, 2005).

As the focus of this very study, Dynamic Assessment (DA) is rooted in Vygotsky’s Sociocultural Theory of mind (SCT); a theory that shares a lot with the notion of TE in principles. Bifurcating assessment and instruction, DA considers assessment “as the means to move toward an always emergent (i.e., dynamic) future” (Pohner, 2005, p. 20) through which the mediator tries to help the learners not only actualize their actual competence but also what they cannot perform independently, and move to the next stage within their Zone of Proximal Development (ZPD) in the light of mediation and interaction (Pohner, 2008).

Contrary to all theoretical claims and the resultant practical and pedagogical implementations, there is a missing link in the literature indicating specific study trying to integrate TE and DA in a single study focusing on a certain area of language education, let alone some sporadic studies focusing on developing language skills (Bohrer-Levi, 2012). Given this trend of developments, this study tries to shed [some innovative] lights on the development of the language productive skills in the light of integrating DA and TE in an EFL setting; better to say making academic peace between the principles of SCT and TE. To do so, the problem and purpose are realized in the form of the following comparative research questions and respective null hypotheses:

(1) Does integration of TE with DA affect speaking ability of EFL learners more significantly than TE does?

(2) Does integration of TE with DA affect writing ability of EFL learners more significantly than TE does?

The null hypotheses studied in the current research were as follows:

H01: Integration of TE with DA does not affect speaking ability of EFL learners more significantly than TE does.

H02: Integration of TE with DA does not affect writing ability of EFL learners more significantly than TE does.

2. Literature review
TE is rooted in the constructivist view holding that meaning is constructed and developed through personal experience and is validated through interaction and communication with others. According to Mezirow (2000), leaning happens through objective and subjective reframing (Kitchenham, 2008, p. 110). Objective reframing involves critical reflection on other’s assumption and subjective reframing involves self-reflection of one’s own assumption.
TE research continues to thrive since, according to Fergal Finnegan (2020), “key thinkers in the field such as Mezirow grasp at a very deep level that reality is highly complex and fluid and makes this central to the framing of the theory” (p.). Referring to all human daily problems and even traumas, he claims that “These sorts of social circumstances and life situations, our contributors suggest, also create opportunities for transformative learning”. According to him, TE, of course, is suggested as an object of further research, but it in turn ultimately changes to be a means of remedy of all these problems in claiming that “A great deal of transformative education research … … … can allow us to generatively respond to just such disjunctions and dilemmas”. Such a claim is rationalized on the grounds that TE develops rational, collaborative, democratic modes of thinking, bolsters the capacity for meaningful agency, and affords the opportunity to elaborate new narratives and to re-story ourselves.

Moreover, the core idea within transformative education is “substantive dialogue and a commitment to equality”. To this end, Buechner et al. (2020) argue that “we need to approach body–mind and the individual–within groups and shared-settings in a far more integrated and holistic way. This idea originates from Victor Turner (1969) notion of “liminality” (the state of being in-between, at a threshold). They argue that liminality presents opportunities for growth. The article brings into view something -the specific dynamics and qualities of intersubjective experience which lead to collective transformations.

In the same vein, Payne, H. et al. (2019) focus on the advantages of “persons in groups, the power of social institutions, and most centrally of all the intertwined nature of mind and body”. Referring to the current suffering from “medically unexplained symptoms” (MUS) in contemporary society, they argue transformative learning theory can help explain and support “self-management” and release from such symptoms …

In another feature article. Lee et al. (2020) look at transformative learning as a means to obtain “creative confidence.” Their study revealed that “enhancing creativity amongst students entailed the “demystification of creativity”. The findings suggest that the common understanding of creativity as a “gift”, and rare, possessed by individuals can block creativity and thus unpicks a dualistic way of thinking about individual and collective creativity.

Kang et al. (2020), discussing the effects of a yearlong professional development program on Chinese school leaders, say internationalization has certainly led to profound shifts in the assumptions of this group of school leaders. Since according to him personal meaning-making and distributed leadership models common in the Western educational system have had a transformative impact.

Mezirow (1978) framed this theory based on his study on 83 women and he concluded that the reflection is one of the most crucial components of adult learning due to enabling people to reassess their existing assumptions that scaffold their beliefs, attitudes, and actions (Mezirow, 1990, 1991, 2000, 2009). Mezirovw’s theory is based on Habermas’ (1971) theory of knowledge constitutive interest. The perspective transformation has been a primary focus of adult education for over three decades. Perspective transformation is a term that was coined by Mezirow in 1975 to describe the central process of “adult development and the change toward thinking like an adult through a shift toward a more inclusive, differentiated, permeable, and integrated perspective” (1991a, p. 155). In this regard, Mezirow (1978) asserted that “by recognizing the social, economic, political, psychological, and religious assumptions that shape these [meaning] structures—presuppositions inherited but rarely examined critically—we can reconstruct our personal frame of reference, our self-concept, goals, and criteria for evaluating change” (p. 7). Besides, TE can be considered vital for adults’ education as they transfer the current knowledge and experiences of their native language to the target one (Kegan, 2000; Kuhn, 1996). However, it is significant for adult learners to intentionally change and reshape their existing frame of references. This process of reconstruction happens during the course of reexamining and reflecting the systems of the first language and adjusting to the new information received in the target language (Mezirow, 1990).
Mezirow’s TE theory is the approval of humanistic and constructivist perspectives in the philosophy of education (Baumgartner, 2012; Mezirow, 1994a, 2009). According to Mezirow (2000), there is no definite universal knowledge that can be applied to every individual. Hence, human beings need to constantly negotiate meaning individually (Mezirow, 2000). When there is a gap in observed knowledge or experience, adults will think and adopt themselves to move forward in human development. These gaps which are considered as collided situations are the opportunities for TE. As Mezirow (2003) has argued, “TE is a learning process that reconstructs, modifies, and transforms problematic or malfunctioning frame of reference—sets of fixed, unexamined, and unquestioned assumptions and expectations—to make them more inclusive, discriminating, open, reflective, and emotionally able to change” (p. 58).

Mezirow (2000) argued that “learning occurs in four ways: by elaborating existing frames of reference, by learning new frames of reference, by transforming points of view, or by transforming habits of mind” (p. 19). He explains that a frame of reference is a “meaning perspective” that has cognitive, affective and conative dimensions, the sum of assumptions and expectations through which we interpret our experiences and that offers the context for meaning-making. A frame of reference is composed of a habit of mind expressed through its resulting point of view. Finally, Mezirow (2009) defined TE as “learning that transforms problematic frames of reference to make them more inclusive, discriminating, reflective, open, and emotionally able to change” (Mezirow, 2009, p. 22).

King (2005) explained that international learners transform from inveterate silent members to class leaders through learning opportunities such as critical thinking, research paper presentations, and case studies. All these assist the learners to use their knowledge base in order to make informed and conscious decisions and be able to reflect on their experiences.

Harrison’s (2008) conducted research implementing various case studies, participating in adult literacy programs. This research aimed at studying the impacts of literacy programs on the lives of the learners. Using an analytic tool, he reached a new theory called “metamorphosis”. He concluded that there is a deep structural shift as participants reflected on their personal consciousness. This research was further used for other studies in TE, although it had the problem of making generalizations, which was due to the short sample size. Within this research, no quantitative data were offered to examine the correlation between TE and their educational background.

Fullerton (2010) conducted a research aiming at exploring how TE was incorporated into the experience of college students who were internationally exposed to TE strategies which were engaged in a leadership development program. Although this research was limited to only adult learners in the United States, he concluded that age was strong correlating a factor for TE to occur and TE can and does occur independently.

Schwartz (2013) conducted a study examining factors that promote TE experiences of college-level adult learners of foreign languages aiming at analyzing how college-level adult learners of foreign languages experience TE through educational and non-educational experiences. Due to the result of this research, 84.7% of participants did not have TE experiences while 15.3% of the participants had TE experiences. Although the participants attending this study reported experiencing non-educational perspective of transformation, those who who employed self-reflection during the mentioned process reported experiencing educationally-related perspective of transformation.

Stahl (2012) conducted a research on transformative professional development through the eyes of Mezirow and Guskey. This study investigated the process of professional development that valued teacher’s personal background, included their present teaching context and focused on real-time applications was considered effective by teachers and thus more likely to effect change in their classroom pedagogy. Additionally, it was beneficial when professional development included time for participants to both personally reflect on and to dialog with other colleagues about their learning.
Kaloupitsi (2016) studied that transformative effect of learning about culture through foreign language acquisition on Greek’s adult learning. Through that qualitative approach, she concluded a positive outcome revealing that transformative effect of learning is possible through language acquisition. Moreover, DA mode was investigated by Zhang (2008) for online EFL writing classes. The findings revealed that the DA greatly improved the learners’ writing abilities by mediation on encouraging remarks, writing strategies, and giving reference materials by the teacher. These factors not only bring out creativity for the learners, but also result in harmonious cooperation and collaboration between them.

Focusing on DA side of the coin, Safa, Donyaei and Mohammadi conducted a research on the effect of interactionist versus interventionist DA on EFL learner’s speaking skill. Hence, they studied 40 homogeneous Iranian EFL learners divided into DA group and Non-DA group. The analysis of the data revealed that interactionist model of DA had a statistically significant positive effect on Iranian EFL learners’ speaking ability; meanwhile, interventionist model of DA had a statistically significant positive effect on Iranian EFL learners’ speaking ability. The results also indicated that the three groups, namely, interactionist DA, interventionist DA, and non-DA had statistically significant effects on Iranian EFL learners’ speaking ability with the interactionist DA group outperforming. In the same vein, Fahmy (2013) conducted a research on the effect of DA on adult learners of Arabic at an institute level. The results of comparing the different evaluations conducted in both the pre- and post-DA phase showed that the structural control of Arabic improved for all participants. DA instruction was practical and successful in making a difference for the participants’ learning process. It reflected the success of the ILR-based rubrics in diagnosing accurately the students’ inabilitys whether in the interventionist-DA interviews or in the daily interactionist DA.

Mediation of meaning mentioned by Feuerstein (1979) as the main attributes of DA emphasize that the mediation and instruction would not be successful until it accompanies exploiting the relationships and making the connections clear. As Feuerstein says, the “episodic” grasp with which learners are faced, forces them to make a connection between the past and future projects.

In 2016, Ebadi researched the effect of DA on developing speaking abilities of EFL learners. Through Microgenetic and thematic analysis he explored a significant development in the participants’ cognition and their movement toward further self-regulation.

In 2015, Kordjazi and Derakhshan studied the Implications of DA in Second/Foreign Language Contexts. According to the data analysis, they concluded that DA has helped education through gathering data about students’ potential. It was also confirmed that it is able to adapt and accommodate students to the environment.

Son and Kim (2017) studied the potential of DA for English-speaking performance development. The results of the current study revealed that mediator’s mediational feedback and the learner’s reciprocating leads to a decrease in the mediator’s explicit feedback and an increase in the proactive moves of the learners. The data also showed that there is a positive relationship between the learner’s developing linguistic profiles, dynamic reciprocating moves, and the cognitive and collaborative strategies for meaning-making.

Contrary to all these sporadic and separate studies; some on TE applications and some on DA (i.e. SCT), no single study could be traced to synthesizing the principles of these two theoretical paradigms in general and in a very single study on teaching language productive skills in particular in EFL educational context.

3. Methodology

3.1. Design and context of the study

As far as this study is concerned, since the first research question aimed at examining the probable effects of integration of TE and DA on a certain group of learners compared to the group receiving TE,
3.2. Participants
The participants were 105 Iranian mainly Persian university students of English major ranged from 20 to 25 years old who were selected based on convenient sampling but randomly organized in four classes. Given the fact that students initiate their studies at Iranian universities – for bachelor degree – at the age of 18 and there is no age restriction for the enrollment, the students entering the university might be from different age groups, which justifies the attendance of the 18–25 age range in this study. The candidates compete through a nation-wide high-stake Entrance Examination Test called “Konkour” which, to some extent, could be called a valid and reliable measure of language ability of the adult candidates and nullifies some selection concerns, though their language proficiency was additionally and specifically measured for the purpose of much more homogeneity in this study. For the purpose of the current study, the students were first conveniently selected from among those who had enrolled for the academic year of 2016–2017. The experimental group consisted of two classes and 55 learners, whereas the control group consisted of 50 learners who, following the initial convenient sampling, were randomly divided into the experimental and control groups.

3.3. Instruments
Multiple instruments including pre-test of speaking, pre-test of writing, posttest of writing and posttest of speaking, were used for the purpose of this study. The speaking section of IELTS Test was used as the pretest in order to test their speaking proficiency. In claiming primary use of the general proficiency model, the IELTS speaking sub-test asserts the notion that there is some varying technically analyzable, but fundamentally indivisible, body of language knowledge within each test-taker, and therefore, individuals can be ranked on the basis of this knowledge. Therefore, the test strives to discover proficiency through performance (Quaid, 2018). Additionally, the IELTS sub-tests were employed given the experience of close working with the participants and full knowledge of their language proficiency level. The speaking test was marked by using the IELTS speaking band descriptors and by assigning a mark to each descriptor (Appendix A). Meanwhile, the students’ writing abilities were assessed before the treatment. The writing assessment was used to assess the fundamental writing skills and measured the ability to write effectively. The topic of essay was “In cashless society, people use more credit cards. What are the advantages and disadvantages of this phenomenon? And it was the same for all students and it was scored based on how effectively the writing communicated the whole message. The score was based on ability to express, organize, and support opinions and ideas according to the IELTS band score description (Appendix B).

The posttest included speaking tests were recorded as well as writing tests. Posttest of writing included a topic as presented in IELTS writing test and the students were supposed to write an essay on the given topic individually within 40 min. The topic was “The current trend in education is to move away from traditional exams and instead have continuous assessment over the school year. What do you think of this trend?” The writing assessments of both groups were the same concerning the type of questions and topics. The same types of questions were also asked from experimental and control groups to collect data on their speaking performances. In order to estimate the reliability of the test, it was piloted again on another group of 30 university students. The piloting sample had similar features as the three groups. The reliability level was 0.70 as shown in a Cronbach’s alpha coefficient.

As the participants were at the level of self-expression in English, so the speaking section of IELTS Test was used as the pretest to test their speaking proficiency. The speaking test was marked by using the IELTS speaking band descriptors (Appendix B) and by assigning a mark to each descriptor. Meanwhile, the students’ writing ability was assessed before the treatment. They were required to write an essay on the given subject and their writing performance was measured through IELTS Writing band score (Appendix A) to assess their ability to express, organize, and
support opinions. The topic given to the participants was of general knowledge type in order to decrease the effect of topic influence on their writing.

In order to check the reliability of the writing and speaking pretest, they were piloted with a group of 30 university students besides the three target groups of the study. The piloting sample had similar features as those of the three groups. Also, the writings were assessed by two university professors to ensure their inter-rater reliability.

3.3.1. Inter-rater reliability of pretests and posttests of writing and speaking

Table 1 and Table 2 display the results of the Pearson correlations run to estimate the inter-rater reliability of the two raters who rated the participants’ performance on pretests and posttests of writing and speaking. Accordingly, there were significant agreements between the two raters on:

- Pretest of writing \( r(103) = .706 \), representing a large effect size, \( p = .000 \), and

- Posttest of writing \( r(103) = .578 \), representing a large effect size, \( p = .000 \).

Based on these results displayed in Table 2 it can be concluded that there were significant agreements between the two raters on:

- Pretest of speaking \( r(103) = .558 \), representing a large effect size, \( p = .000 \), and

- Posttest of speaking \( r(103) = .657 \), representing a large effect size, \( p = .000 \).

3.4. Data collection procedure

At first, the selected participants were interviewed separately on an assigned topic. This interview was based on IELTS exam speaking test which took about 15 min. They were also given the topic “In cashless society, people use more credit cards. What are the advantages and disadvantages of this phenomenon? They were asked to write an essay. The purpose was to elicit information regarding the current level of speaking and writing and to compare with the posttest data in order to evaluate the impact of the treatment on the participants.

In the second stage, the treatment was implemented into two different groups of students. One group (including two classes) received the treatment in writing and speaking skills based on TE. After the pretest and during the course, the instructor followed TE teaching strategies, namely, as (but not limited to):

- Associating the learner's goals to a predefined larger goals
- Providing challenging feedbacks for the learners
- Making students aware of what type of feedback they may received
- Fostering intellectual openness
- Cultivating reflective learners
- Creating opportunities for the learners to think in some more sophisticated ways
- Encouraging critical thinking and discourse
- Providing meaningful assignments

Moreover, the experimental group of students was given treatment based on the integration of TE and DA. This group got the benefit of TE as the first group. The difference in the treatment lied in the assessment procedure, which was DA. The treatment took 16 sessions.

The stages of TE included assigning the individuals in independent teams where they could communicate ideas and knowledge along with responsibilities exchange. The second stage was making
students involved in the content of the course being taught during certain sessions of the course. In the third stage, the students’ time for classroom discussions was expanded in order to develop their interactions. The final stage within the procedure of this research was developing assessments which themselves were learning opportunities; in other words, they were exposed to new challenges with which they could gain more knowledge on the subject area. The assessment in the experimental group was not only a test format but also ongoing DA of the learner’s progress on their transformational process. The point in the experimental group revolved around mediating the learners to overcome their problems dynamically and critically. The mediation offered by the instructor during the semester was per session and while the students were assigned into different groups. At each session of the instruction, at first, the teacher taught the topics of each session. The focus of the teaching materials regarding speaking was on parts of speech, fluency and task achievement. Meanwhile, the focus of the teaching material regarding writing was on parts of speech, cohesion and coherence, topic sentence, and task achievement. After the instruction, learners were assigned into groups. Based on the tasks students were given to complement within the group, either speaking or writing tasks, they were challenged as to receiving new viewpoints; These challenging moments were the times that instructor mediated the group members through the predefined scale of medication moving from implicitly to explicitly. In this study, instruction and assessment occurred simultaneously in that the teachers – as mediator – promote development by offering assistance to a student while concurrently assessing the student’s abilities. The instructor commenced by asking the learners to independently correct the errors and ended up with the provision of correct patterns and examples.

There are some critical points in assisting learners in the experimental group in all stages. First, the assistance provided for the participants was gradual so that implicit help was offered initially and increasingly became more explicit as required by the learner. Second, help was contingent so that it was only offered when needed. Specifically, explicit forms of assistance were only provided when implicit forms were insufficient. Examples of such explicit forms included providing the correct form, or some explanations for using the correct form and even examples of correct pattern. Third, help was rendered through dialogue through which the teacher and learner co-constructed the intended meaning.

Multiple TE strategies were employed during the research. Among these strategies were creating opportunities for the learners to think in more sophisticated ways, encouraging critical thinking and giving meaningful assignments. These assignments included information, reasoning and opinion gap activities. For example, two students might have different schedules, but they want to find time to get together to work on their course project. The students were allowed to request information, and to negotiate the meaning whenever there were misunderstandings. The students were also challenged to solve group initiative problems. Many of these challenging activities were problem-solving which required critical reflection on the suggestions made by the group to solve the problems. During all the tasks given to the students to fulfill, the instructors took note of the reaction of the participants to the activities and topics presented to them.

Then, the participants received a posttest of writing and speaking. Posttest of writing included a topic as presented in IELTS writing test and the students were supposed to write an essay on the

| Table 1. Pearson correlations; inter-rater reliability of pretest and posttest of writing |
|-----------------------------------|------------------|------------------|
|                                  | Pre-Rater2       | Post-Rater2      |
|---------------------------------|------------------|------------------|
| Pre-Rater1                      | Pearson Correlation | .706**          |
|                                 | Sig. (2-tailed) | .000             |
|                                 | N               | 105              |
| Post-Rater1                     | Pearson Correlation | .578**          |
|                                 | Sig. (2-tailed) | .000             |
|                                 | N               | 105              |

**. Correlation is significant at the 0.01 level (2-tailed).
given topic individually within 40 min. The topic was “The current trend in education is to move away from traditional exams and instead have a continuous assessment over the school year. What do you think of this trend?” The writing assessments of both groups were the same regarding the type of questions and topics. The same types of questions were also asked from both groups to collect data on their speaking performances.

4. Results

4.1. Testing normality assumptions
Prior to deciding on the statistical approach, the collected data were first checked in terms of parametric test assumptions. Parametric tests assume four assumptions to be met including: “1) dependent variable, are interval scored or strongly continuous, 2) the data are normally distributed, 3) the distribution can be estimated in the population from which the respective samples have been drawn, and 4) the observations are independent” (Hatch & Lazarton, 1991, p. 237–238). Given the assumptions 1, 3, and 4 which are met, the data were checked in terms of normality. Table 1 displays the results of skewness and kurtosis and their ratios of their standard errors. The results indicated two different pictures for the pre-tests and post-tests’ scores of both skills as that the absolute values of these ratios for the pre-tests of both groups were higher than 1.96, while they were lower for the post-tests. That was why there was no choice other than resorting to the non-parametric Mann-Whitney U test for the pre-tests’ scores but independent samples t-test for the post-tests’ scores; pre-tests scores violating one of the normality assumptions but post-tests’ scores meeting them all! [A motive for more research in statistics as the subject’s performance may change the nature of the data]. Additionally, homogeneity measures of the groups on the pre-tests were also calculated. It should be noted that the assumptions of lack of univariate and multivariate outliers were also checked. Please refer to Appendix C.

4.2. Homogeneity measures based on speaking pretest
A non-parametric Mann-Whitney U test was run to compare the integrated and unintegrated groups’ means on pretest of speaking in order to prove that they were homogenous in terms of their speaking ability prior to the administration of the treatments. The Mann-Whitney U test was run because as displayed in Table 3 the distribution of scores violated the assumption of normality on pretest of speaking. Table 4 displays the median scores and mean ranks for the two groups on pretest of speaking. Based on these results it can be concluded that the integrated (Mdn = 5) and unintegrated (Mdn = 4.75) groups had fairly close median scores on pretest of speaking. Moreover, Table 5 displays the results of the Mann-Whitney U test. The results (Z = −.888, p = .374, Glass Rank Biserial Correlation Coefficient1 = .096 representing a weak effect size) indicated that there was not any significant difference between the integrated and unintegrated groups’ median scores on pretest of speaking (Figure 1). Thus, it can be concluded that the two groups were homogenous in terms of their speaking ability prior to the administration of the treatment.
Table 3. Testing normality of data

| Group            | N | Skewedness |       |       |       | Kurtosis |       |       |
|------------------|---|------------|-------|-------|-------|----------|-------|-------|
|                  |   | Statistic  | Statistic | Std. Error | Ratio | Statistic | Std. Error | Ratio |
| Integrated       |   | 55         | -1.00  | .322 | -0.31 | -1.324 | .634 | -2.09 |
| Pre-Speaking     |   | 55         | .436   | .322 | 1.35  | -2.224 | .634 | -0.35 |
| Post-Speaking    |   | 55         | .200   | .322 | 0.62  | -1.395 | .634 | -2.20 |
| Pre-Writing      |   | 55         | -.392  | .322 | -1.22 | .167   | .634 | 0.26  |
| Post-Writing     |   | 55         | -.392  | .322 | -1.22 | .167   | .634 | 0.26  |
| Unintegrated     |   | 50         | .069   | .337 | 0.20  | -1.436 | .662 | -2.17 |
| Pre-Speaking     |   | 50         | -.207  | .337 | -0.61 | .192   | .662 | 0.29  |
| Post-Speaking    |   | 50         | .080   | .337 | 0.24  | -1.458 | .662 | -2.20 |
| Pre-Writing      |   | 50         | -.177  | .337 | -0.53 | .107   | .662 | 0.16  |
| Post-Writing     |   | 50         | -.177  | .337 | -0.53 | .107   | .662 | 0.16  |
4.3. Homogeneity measures of groups on the pretest of writing

A non-parametric Mann–Whitney U test was run to compare the integrated and unintegrated groups’ means on pretest of writing in order to prove that they were homogenous in terms of their writing ability prior to the administration of the treatments. The Mann–Whitney U test was run because as displayed in Table 3 the distribution of scores violated the assumption of normality on pretest of writing. Table 6 displays the median scores and mean ranks for the two groups on pretest of writing. Based on these results it can be concluded that the integrated (Mdn = 4.5) and unintegrated (Mdn = 4.75) groups had fairly close median scores on pretest of writing.

Table 7 displays the results of the Mann–Whitney U test. The results (Z = −.201, p = .840, Glass Rank Biserial Correlation Coefficient = .021 representing a weak effect size) indicated that there was not any significant difference between the integrated and unintegrated groups' median scores on the pretest of writing.
Thus, it can be concluded that the two groups were homogenous in terms of their writing ability prior to the administration of the treatment.

4.4. Testing the research hypotheses

As already stated and given the normality nature of the post-tests’ scores of both skills, independent-samples t-tests were run to compare the integrated and unintegrated groups’ means on first the posttest of speaking (Figure 3) in order to probe the first null-hypothesis and that of the writing skill to probe the second research question.

4.4.1. Testing the first null-hypothesis

The first null-hypothesis stated that the integration of TE with DA did not significantly affect learners speaking more significantly than TE did. Table 8 displays the two groups’ means on posttest of speaking. The results indicated that the integrated group (M = 6.45, SD = .515) had a slightly higher mean than the unintegrated group (M = 6.01, SD = .576) on posttest of speaking.
The results of the independent t-test (t (103) = 4.09, p = .000, r = .369 representing a moderate effect size) (Table 9) indicates that the integrated group significantly outperformed the unintegrated group on posttest of speaking. Thus, the first null-hypothesis as “integration of TE with DA did not significantly affect learners speaking more significantly than TE did” was rejected.

It should also be noted that the assumption of homogeneity of variances was met (Levene’s $F = .001$, $p = .988$). That was why the first row of Table 9, i.e. “Equal variances assumed” was reported.

4.4.2. Testing the second minor null-hypothesis

The second null-hypothesis stated that the integration of TE with DA did not significantly affect learners’ writing more significantly than TE did. An independent-samples t-test was run to compare the integrated and unintegrated groups’ mean on posttest of writing in order to probe the second null-hypothesis. Table 10 displays the two groups’ means on posttest of writing. The results indicated that the integrated group (M = 6.25, SD = .517) had a slightly higher mean than the unintegrated group (M = 6.01, SD = .490) on posttest of writing (Figure 4).

The results of the independent t-test (t (103) = 2.39, p = .019, r = .229 representing a weak effect size) (Table 11) indicates that the integrated group significantly outperformed the unintegrated group on posttest of writing. Thus, the second null-hypothesis as “integration of TE with DA did not significantly affect learners writing more significantly than TE did” was rejected, although the results should be interpreted cautiously due to the weak effect size value of .229.

It should also be noted that the assumption of homogeneity of variances was met (Levene’s $F = 1.94$, $p = .166$). That was why the first row of Table 11, i.e. “Equal variances assumed” was reported.

5. Discussion and conclusion

This research aimed to investigate whether the integration of TE and DA can have any significant effect on EFL learners’ productive skills; writing and speaking compared to TE. Based on the results of the t-test, there were significant differences between control and experimental groups at the learners’ speaking and writing scores. Hence, it was concluded that students receiving the treatment of the integration of DA and TE performed better in their speaking and writing compared to the group which benefited from TE. Based on the findings, the posttest scores in the experimental groups in both speaking and writing skills were significantly higher than their scores in the pretest.
| Levene's Test for Equality of Variances | t-test for Equality of Means |
|---|---|
| F   | Sig. | T   | Df   | Sig. (2-tailed) | Mean Difference | Std. Error Difference | =95% Confidence Interval of the Difference | Lower | Upper |
| Equal variances assumed | .001 | .988 | 4.090 | 103 | .000 | .435 | .106 | .224 | .647 |
| Equal variances not assumed | 4.068 | 98.799 | .000 | .435 | .107 | .223 | .648 |
The outcomes indicated that the learners’ paragraphs and speaking skill enhanced in terms of both clarity and content. The results revealed that almost all the learners who have participated in the study had a problem in the essay writing but improved their writings by experiencing the mediation. The mediation presented to the learners was different in each of them since the difficulty overcoming and struggling the tasks was different in different learners. During the mediation, some of the participants were able to enhance their writings since they could identify their own mistakes or errors. By comparing the pretest and posttest paragraphs written by the students, it was identified that mediation was successful most of all in the syntactic section and less than all in the spelling section. The reason behind this outcome is that in the case of syntax, the number of grammatical errors of the writings were limited to some common difficulties among the essays and as to the spelling the domain of the words to be learned was unlimited.

The findings obtained in this research are in line with the findings by Knodel (1997). He probed the effectiveness of the mediated learning experience and the graduated prompt (two kinds of DA approaches) on the assessment of composition writing of the poor writers with a learning disability and without a learning disability learners. The findings of the research showed that graduate prompt intervention was effectively beneficial. On the other hand, the learners who did not improve through graduated prompt received the mediated learning intervention. Moreover, DA mode was investigated by Zhang (2008) for online EFL writing classes. The findings revealed that the DA greatly improved the learners’ writing abilities through mediation provided by the teachers by the means of encouraging remarks, writing strategies, and giving reference materials. The mentioned factors brought out creativity along with cooperation and collaboration between learners.

The findings of the present study were also in line with the findings of Ahmadi Safa et al. (2015) who studied the effect of DA on speaking abilities of the Iranian EFL learners and concluded that DA had a statistically significant positive effect on Iranian EFL learners’ speaking ability.

As Duvall (2008) asserted, regardless of the mediation offered through DA, the important factor, which is very effective, is the ability of the teachers who engage the learners in speaking skills via cooperative activities. This enables learners to co-construct their ZPD.
| Levene's Test for Equality of Variances | t-test for Equality of Means |
|----------------------------------------|-----------------------------|
| F | Sig. | T | Df | Sig. (2-tailed) | Mean Difference | Std. Error | 95% Confidence Interval of the Difference |
|---|---|---|---|---|---|---|---|
| Equal variances assumed | 1.943 | .166 | 2.390 | 103 | .019 | .235 | .099 | .040 | .431 |
| Equal variances not assumed | 2.396 | 102.821 | .018 | .235 | .098 | .041 | .430 |
The findings of the present study, which are in line with those by Ebadi (2016), Hill and Sabet (2009), and Son and Kim (2017), revealed that DA over time had a significant positive impact on the participants’ learning process. The frequency of errors, mediations provided by the mediator, the participants’ responsiveness to the mediation and the types of their private speech were regarded as an index for assessing their progress in speaking skills.

Furthermore, it has been revealed by the current study that the integration of DA and TE is an effective means through which the performance of the students might be described. Hence, DA needs to be used for the learners failing to perform properly on the standardized instruments. This is rooted in the fact that mediation varies according to each individual.

Moreover, the findings were supported to the findings of the research done by Kalouptsi (2016). She studied the effect of TE and learning about a culture of learning a foreign language. Research delivered positive outcomes showed evidence that the transformative effect of learning about a culture through language acquisition is indeed possible.

Results of the present study indicated that the integration of TE and DA can influence both speaking and writing proficiency of the EFL learners at the university level. It can be concluded that not only this integration assists the writing ability of the learners, but also it is very influential in enhancing speaking proficiency.

As Vygotsky (1978) argued, the most effective instruction leads to human development. In the circumstances where the learners are exposed to a challenge beyond their current cognitive level, the expectation of the learners develops. In other words, when teachers convince the learners that their capabilities are much higher than their expectations, learners perform noticeably better to reach higher levels.

Stremmel and Fu (1993) argued that during establishing DA, teachers playing the role of the mediator, have critical encouraging effects on the provision of support leading to learning and development. They approved that when the teachers do not demand high or employ less demonstration, the most effective teaching through the assessment happens including the equal partnership; which is in line with the findings of this research within which the role of the mediator is facilitated through integration with TE.

The other major finding of the current study is the role of the teacher as the cause of challenges for the students to experience TE which is mostly facilitated while being integrated with DA. As King (2005) discussed, while instructors can provide strategies and opportunities for transformative learning to occur, it is up to the learner to change or not. As Mezirow pointed out, “One cannot become emancipated through indoctrination” (1991a, p. 88). An individual alone chooses how he/she perceives and makes meaning of the world and his/her experiences within it. Instructors should not force this process, but rather should provide nurturing challenges and support toward this vision. Only in this way can the promise of transformative learning be realized. As King (2005) observed, there are no guarantees of transformative learning outcomes, only opportunities. The life experiences of students must lead them to intellectual and emotional readiness, meet conditional thresholds, and they must be psychologically willing. So, in addition to the nature of the approach implemented in this study, the learners’ willingness can also be among the factors determining the results achieved.

6. Implications and further recommendations
An important educational implication from this study could be for educational administrators who should take the probable impacts of language learning and the context of the learning into considerations because learners might transfer them to their academic life. According to the results, there could be counseling sessions at each academic level through which students are allowed to discuss their challenges and critical situations that have faced during their academic life.
Another educational implication might be that faculties in higher education could practice theories in the classroom as documented in the quantitative findings of the current study. Faculties might integrate activities such as classroom discussions or projects, critical thinking and role-play in their teaching plans within which the learners will find the opportunity to participate in the learning process. Meanwhile, the learners could be familiar with the concepts of different techniques of teaching as findings from the study show that classroom activities were highly associated with experiencing TE of learners.

The present study attempted to integrate DA and TE to examine the probable effect on the productive skills of the learners. This study was limited to DA that is the mediated learning experience. Thus, future studies can consider the various domains of DA concerning the classroom. Addressing the process of learning and teaching in a DA approach is needed for future research in order to shift the focus from the performance to the type and amount of intervention required. This study considered only the intermediate proficiency level learners. Future research can investigate the impact of the integration of DA and TE on other levels of proficiency. Within the current research, all the participants were university students of English. It would be fine to investigate whether there are any similarities and differences between the perceptions of the teachers and learners at other educational levels.

Funding
The authors received no direct funding for this research.

Author details
Azadeh Zarbafian
E-mail: azadehzarbafian@yahoo.com

Gholam-Reza Abbassian
E-mail: gabbasian@gmail.com

Ahmad Mohseni
E-mail: omohseny1328@gmail.com

Abdollah Baradaran
E-mail: baradaranabdollah@yahoo.com

1 Foreign Language Department, Islamic Azad University, South Tehran Branch, Borujerd, Iran.
2 Applied Linguistics, Imam Ali University, Iran.
3 Islamic Azad University, South Tehran Branch, Borujerd, Iran.
4 Islamic Azad University, Central Tehran Branch, Borujerd, Iran.

Correction
This article has been republished with minor changes. These changes do not impact the academic content of the article.

Citation information
Cite this article as: Synthesizing transformative education with dynamic assessment in developing EFL learners' productive skills, Azadeh Zarbafian, Gholam-Reza Abbassian, Ahmad Mohseni & Abdollah Baradaran, Cogent Arts & Humanities (2020), 7: 1750840.

Note
1. Glass Rank Biserial Correlation Coefficient is computed as $r_g= \sqrt{\frac{2(\text{mean rank1}-\text{mean rank2})}{\text{rank1}^2+\text{rank2}^2}}$. (Gray and Kinnear 2012, p. 207).

References
Ahmadi, S., Donyaei, S., Mohammadi, M., & Reza, A. (2015). An investigation into the effect of interactionist versus interventionist models of dynamic assessment on Iranian EFL learners’ speaking skill proficiency. Teaching English Language Journal, 9(2), 146–166.
Anton, M., & DiCamilla, F. (1998). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. Baumgartner, L. M. (2012). Mezirow's theory of transformative learning from 1975 to present. In E. W. Taylor & P. Cranton (Eds.), The handbook of transformative learning: Theory, research, and practice (pp. 37–55). Jossey-Bass.
Boher-Levi, T. (2013). The effect of dynamic assessment on the performance of students in oral proficiency tests in English as a foreign language. Tel Aviv University.
Buechner, B., Dirks, J., Konvisser, Z. A., Myers, D., & Peleg-Baker, T. (2020). From liminality to communitas: the collective dimensions of transformative learning. Journal of Transformative Education, 18(2), 87–113. https://doi.org/10.1177/1541344619900881
Duval, E. (2008). No secrets to conceal: Dynamic assessment and state mandated, high stakes reading assessments for children with learning disabilities [Unpublished doctoral dissertation, The Pennsylvania State University]. (Google Scholar).
Ebadi, S. (2016). Mediation and reciprocity in online L2 dynamic assessment. CALL-EJ, 17(2), 16–40.
Fahmy, M. (2013). The Effect of Dynamic Assessment on Adult Learners of Arabic: A Mixed-Methods Study at the Defense Language Institute Foreign Languages. Unpublished doctoral dissertation, University of San Francisco, San Francisco.
Feuerstein, R. (1975). The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques. University Park Press.
Finnegan, F. (2020). Editor’s notes: Exploring the collective and embodied dimensions of transformative learning. Journal of Transformative Education, 18(2), 83–86. https://doi.org/10.1177/15413446209006632
Fullerton, J. (2010). Transformative learning in college students: A mixed methods study [Doctoral dissertation]. Available from ProQuest Dissertations and Theses database. (AAT 3398454). Retrieved from Dissertations & Theses @ The University of Nebraska-Lincoln. (No. 250713987).
Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. Cambridge Journal of Education, 35(2), 213–224. https://doi.org/10.1080/03057640500543880
Harrison, A. E. (2008). The far reaching impact of transformative learning: An ethnographic case study [Doctoral dissertation]. Available from ProQuest Dissertations and Theses database. (AAT 33473316). Tampa: University of South Florida.
Hatch, E. & Lazarton, A. (1999). The research manual: design and statistics for applied linguistics. Rowley, MA: Newbury House.
Hebermans, J. (1971). Knowledge and human interests. Boston Beacon Google Scholar.

Hill, K., & Sabet, M. (2009). Dynamic speaking assessment. TESOL Quarterly, 43(3), 537–545. https://doi.org/10.1002/tesq.2009.tb00251

Kaloupints, M. (2016). The transformative effect of learning about a culture through foreign language acquisition: A case study of Greek adults learning about Hispanic culture in Spanish language classes in Greece. ISRN nr: LIU-IBI/IMPALG–A–16/009–SE

Kang, H., Sun, Q. & Lyu, L. (2020). Learning to transform through interplay between the confucian and western cultural heritages: A case study of school leadership development in Beijing, China. Journal of Transformative Education, 18(2), 163–182. https://doi.org/10.1177/1541344619877166

Kegan, R. (2000). What form transforms: A constructive-developmental approach to transformative learning. In J. Mezirow (Ed.), Learning as transformation (pp. 56–72). San Francisco: Jossey-Bass.

King, K. P. (2003). Bringing transformative learning to life. Krieger.

Kitchenham, A. (2009). The evolution of Mezirow’s Transformative Learning Theory. Journal of Transformative Education. Retrieved from Research Gate.

Knodel, M. K. (1997). Dynamic assessment of written language (Doctoral dissertation, University of Calgary). MQ20833.

Kuhn, T. S. (1996). The structure of scientific revolutions (3rd ed.). The University of Chicago Press.

Lee, J. H., Portillo, M., & Meneely, J. (2020). Insights into three frames of creative minds: igniting perspective transformation among first-year university students. Journal of Transformative Education, 18 (2), 138–162. https://doi.org/10.1177/1541344619893314

Mezirow, J. (1978). Education for perspective transformation: Women’s re-entry programs in community colleges. Teacher’s College, Columbia University Press.

Mezirow, J. (1993b). How critical reflection triggers transformative learning. In J. Mezirow & Associates (Eds.), Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning (pp. 1–20). Jossey-Bass.211.

Mezirow, J. (1991a). Transformative dimensions of adult learning. San Francisco: Jossey-Bass.

Mezirow, J. (1994a). Understanding transformative theory. Adult Education Quarterly, 44(4), 222–232. https://doi.org/10.1177/074171369404400403

Mezirow, J. (2000). Learning to think like an adult: Core concepts of transformation theory. In J. Mezirow (Ed.), Learning as transformation (pp. 3–33). Jossey-Bass.

Mezirow, J. (2003). Transformative learning as discourse. Journal of Transformative Education, 1(1), 58–63. https://doi.org/10.1177/1541344603252172

Mezirow, J. (2009). An overview on transformative learning. In K. Illeris (Ed.), Contemporary theories of learning: Learning theorists ... in their own words (pp. 90–105). Routledge.

Moran, J. J. (1977). Assessing adult learning: A guide for practitioners. Malabar: Krieger. Snyder, C. (2008). Grabbing hold of a moving target: Identifying and measuring the transformative learning process. Journal of Transformative Education, 6(3), 159–181.

Payn, H., Robert, A., & Jarvis, J. (2019). The body mind approach as transformative learning to promote self-management for patients with medically unexplained symptoms. Journal of Transformative Education, 18(2), 114–137. https://doi.org/10.1177/1541344619883892

Pohner, M. E. (2000). Dynamic assessment of oral proficiency among advanced L2 learners of French [Unpublished doctoral dissertation, Pennsylvania State University].

Pohner, M. E. (2008). Dynamic assessment: A Vygotskian approach to understanding and promoting second language development. Springer Publishing.

Quaid, E. D. (2019). Reviewing the IELTS speaking test in East Asia: Theoretical and practice-based insights. TESOL Quarterly, 4(2). https://doi.org/10.1186/s40668-018-0056-5

Schwartz, T. (2013). Examination of factors that promote transformative learning experience of college level adult learners of foreign languages. (Doctoral dissertation). Retrieved from http://etd.auburn.edu/etd/ Son, G., & Kim, S. (2017). The potentials of dynamic assessment for the development of English speaking performance: A microgenetic analysis. http://scholar.dkyobook.co.kr/searchDetail.laf?barcode=4010025100380 [Google Scholar].

Stohl, L. D. (2012). Transformative Professional Development through the Eyes of Jack Mezirow and Thomas GuskeyRetrieved from http://www.proquest.com/en-US/products/dissertations/individuals.ahtml

Stremmel, A. J., & Fu, V. R. (1993). Teaching in the zone of proximal development: Implications for responsive teaching practice. Child and Youth Care Forum, 22(5), 337–350. https://doi.org/10.1007/BF00760943

Tebachnick, B. G & Fidell, L. S. (2014). Using Multivariate Statistics. Pearson. Boston.

Turner, V. (1965). The ritual process: structure and anti-structure. Chicago: Aldine [Google Scholar].

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. (M. Cole, V. John–Steiner, S. Scribner, & E. Souberman, ed. by). Harvard University Press.

Zhang, Y. H. (2008). The theoretical construction of a dynamic assessment mode in Chinese tertiary EFL writing class with online teaching and scoring systems. CALL-EJ, 14(2), 38–50.
### Appendix A IELTS TASK 2 Writing band descriptors

| Band | Task Achievement                                                                                                                                                                                                 | Coherence and Cohesion                                                                 | Lexical Resource                                                                                                                                                                                                                                                                                                                                 | Grammatical Range and Accuracy                                                                                                                                                                                                 |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 9    | • Fully addresses all parts of the task  
• Presents a fully developed position in answer to the question with relevant, fully extended and well supported ideas                                                                                     | • Uses cohesion in such a way that it attracts no attention  
• Skilfully manages paragraphing                                                                                      | • Uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as "slips"                                                                                                                                          | • Uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as "slips"                                                                                                                                           |
| 8    | • Sufficiently addresses all parts of the task  
• Presents a well-developed response to the question with relevant, extended and supported ideas                                                                                                  | • Sequences information and ideas  
• logically  
• Manages all aspects of cohesion well  
• Uses paragraphing sufficiently and appropriately                                                                        | • Uses a wide range of vocabulary  
• Fluently and flexibly to convey precise meanings  
• Skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation  
• Produces rare errors in spelling and/or word formation                                                                                              | • Uses a wide range of structures  
• The majority of sentences are error-free  
• Makes only very occasional errors or inappropriacies                                                                                                      |
| 7    | • Addresses all parts of the task  
• Presents a clear position throughout the response  
• Presents, extends and supports main ideas, but there may be a tendency to overgeneralise and/or supporting ideas may lack focus                                                                 | • Logically organises information and ideas; there is clear progression throughout  
• Uses a range of cohesive devices appropriately although there may be some under-/over-use  
• Presents a clear central topic within each paragraph                                                                 | • Uses a sufficient range of vocabulary to allow some flexibility and precision  
• Uses less common lexical items with some awareness of style and collocation  
• May produce occasional errors in word choice, spelling and/or word formation                                                                                                                 | • Uses a variety of complex structures  
• Produces frequent error-free sentences  
• Has good control of grammar and punctuation but may make a few errors                                                                                                                             |

(Continued)
### Task Achievement
- Addresses all parts of the task although some parts may be more fully covered than others
- Presents a relevant position although the conclusions may become unclear or repetitive
- Presents relevant main ideas but some may be inadequately developed/unclear

### Coherence and Cohesion
- Arranges information and ideas coherently and there is a clear overall progression
- Uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical
- May not always use referencing clearly or appropriately
- Uses paragraphing, but not always logically

### Lexical Resource
- Uses an adequate range of vocabulary for the task
- Attempts to use less common vocabulary but with some inaccuracy
- Makes some errors in spelling and/or word formation, but they do not impede communication

### Grammatical Range and Accuracy
- Uses a mix of simple and complex sentence forms
- Makes some errors in grammar and punctuation but they rarely reduce communication
## Appendix B IELTS Speaking band descriptors

| Band | Fluency and coherence | Lexical Resource | Grammatical range and accuracy | Pronunciation |
|------|-----------------------|-----------------|-------------------------------|---------------|
| 9    | * Speaks fluently with only rare repetition or self-correction;  
      * Any hesitation is content-related rather than to find words or grammar  
      * Speaks coherently with fully appropriate cohesive features  
      * Develops topics fully and appropriately  
   | * Uses vocabulary with full flexibility and precision in all topics  
      * Uses idiomatic language naturally and accurately  
   | * Uses a full range of structures naturally and appropriately  
      * Produces consistently accurate structures apart from “slips” characteristic of native speaker speech  
   | * Uses a full range of pronunciation features with precision and subtlety  
      * Sustains flexible use of features throughout  
      * Is effortless to understand |
| 8    | * Speaks fluently with only occasional repetition or self-correction; hesitation is usually content-related and only rarely to search for language  
      * Develops topics coherently and appropriately  
   | * Uses a wide vocabulary resource readily and flexibly to convey precise meaning  
      * Uses less common and idiomatic vocabulary skilfully, with occasional inaccuracies  
      * Uses paraphrase effectively as required  
   | * Uses a wide range of structures flexibly  
      * Produces a majority of error-free sentences with only very occasional inappropriacies or basic/ non-systematic errors  
   | * Uses a wide range of pronunciation features  
      * Sustains flexible use of features, with only occasional lapses  
      * Is easy to understand throughout; L1 accent has minimal effect on intelligibility |
| 7    | * Speaks at length without noticeable effort or loss of coherence  
      * May demonstrate language-related hesitation at times, or some repetition and/or self-correction  
      * Uses a range of connectives and discourse markers with some flexibility  
   | * Use vocabulary resource flexibly to discuss a variety of topics  
      * Uses some less common and idiomatic vocabulary and shows some awareness of style and collocation, with some inappropriate choices  
      * Uses paraphrase effectively  
   | * Uses a range of complex structures with some flexibility  
      * Frequently produces error-free sentences, though some grammatical mistakes persist  
   | * all the Shows positive features of Band 6 and some, but not all, of the positive features of Band 8 |

(Continued)
| Band | Fluency and coherence | Lexical Resource | Grammatical range and accuracy | Pronunciation |
|------|-----------------------|-----------------|-------------------------------|---------------|
| 6    | • Is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation  
• Uses a range of connectives and discourse markers but not always appropriately | • Has a wide enough vocabulary to discuss topics at length and make meaning clear in spite of inappropriacies  
• Generally paraphrases successfully | • Uses a mix of simple and complex structures, but with limited flexibility  
• May make frequent mistakes with complex structures though these rarely cause comprehension problems | • Uses a range of pronunciation features with mixed control  
• Shows some effective use of features but this is not sustained  
• Can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times |
| 5    | • Usually maintains flow of speech but uses repetition, self-correction and/or slow speech to keep going  
• May over-use certain connectives and discourse markers  
• Produces simple speech fluently, but more complex communication causes fluency problems | • Manages to talk about familiar and unfamiliar topics but uses vocabulary with limited flexibility  
• Attempts to use paraphrase but with mixed success | • Produces basic sentence forms with reasonable accuracy  
• Uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems | • Shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6 |
| 4    | • Cannot respond without noticeable pauses and may speak slowly, with frequent repetition and self-correction  
• Links basic sentences but with repetitious use of simple connectives and some breakdowns in coherence | • Is able to talk about familiar topics but can only convey basic meaning on unfamiliar topics and makes frequent errors in word choice  
• Rarely attempts paraphrase | • Produces basic sentence forms and some correct simple sentences but subordinate structures are rare  
• Errors are frequent and may lead to misunderstanding | • Uses a limited range of pronunciation features  
• Attempts to control features but lapses are frequent  
• Mispronunciations are frequent and cause some difficulty for the listener |
| 3    | • Speaks with long pauses  
• Has limited ability to link simple sentences  
• Gives only simple responses and is frequently unable to convey the basic message | • Uses simple vocabulary to convey personal information  
• Has insufficient vocabulary for less familiar topics | • Attempts basic sentence forms but with limited success, or relies on apparently memorised utterances  
• Makes numerous errors except in memorised expressions | • Shows some of the features of Band 2 and some, but not all, of the positive features of Band 4 |
Continued

| Band | Fluency and coherence | Lexical Resource | Grammatical range and accuracy | Pronunciation |
|------|-----------------------|-----------------|-------------------------------|---------------|
| 2    | • Pauses lengthily before most words  
      • Little communication possible | • Only produces isolated words or memorised utterances | • Cannot produce basic sentence forms | • Speech is often unintelligible |
| 1    | • No communication possible  
      • No rateable language |                      |                               |               |
| 0    | • Does not attend |                      |                               |               |
Appendix C Testing Lack of Univariate Outlier

The standardized scores (Z-scores) were computed for pretests and posttests of speaking and writing. Table C1 displays the descriptive statistics for the Z-scores for the pretests and posttests of speaking and writing. None of the variables had a Z-score higher than ± 3. That is to say, the data did not have any univariate outliers.

To check multivariate outliers, the Mahalanobis Distances were computed. The Mahalanobis Distances were compared with the chi-square critical value of 20.51 at .001 levels of significance for five dependent variables (Tabachnick and Fidell 2014). Table C2 displays the descriptive statistics for the Mahalanobis Distances. The maximum observed value of 14.27 was lower than 20.51 indicating that the present data did not have any multivariate outliers.

| Table C1. Descriptive statistics and Z-scores for posttests of speaking and writing |
|---------------------------------|---|---|---|---|---|
|                                | N  | Minimum | Maximum | Mean | Std. Deviation |
| Z-score: Pretest of Speaking   | 105| -1.250   | 1.466   | .000 | 1.000          |
| Z-score: Posttest of Speaking  | 105| -2.117   | 2.158   | .000 | 1.000          |
| Z-score: Pretest of Writing    | 105| -1.137   | 1.535   | .000 | 1.000          |
| Z-score: Posttest of Writing   | 105| -2.198   | 1.681   | .000 | 1.000          |
| Z-score (LAS)                  | 105| -2.913   | 2.082   | .000 | 1.000          |

| Table C2. Descriptive statistics of mahalanobis distances |
|----------------------------------------------------------|
|                                | N  | Minimum | Maximum | Mean | Std. Deviation |
| Mahalanobis Distance           | 105| .493    | 14.278  | 4.952| 2.775          |
