Journal of Modern Research in English Language Studies
Print ISSN: 2676-5357
Online ISSN: 2676-5985

Computer-Mediated Diagnostic Assessment of Mixed-Ability EFL Learners' Performance on Tiered Tasks: Differentiating Mediation on Google Meet™

Fahimeh Rafi[1], Natasha Pourdana[2*], Farid Ghaemi[3]

[1]Department of Teaching English and Translation, Karaj Branch, Islamic Azad University, Karaj, Iran, *fahime.rafi@iauz.ac.ir*
[2*]Department of Teaching English and Translation, Karaj Branch, Islamic Azad University, Karaj, Iran, *natasha.pourdana@kiau.ac.ir*
[3]Department of Teaching English and Translation, Karaj Branch, Islamic Azad University, Karaj, Iran, *ghaemi@kiau.ac.ir*

Article Info

Article type:

Research Article

Received:

24/08/2021

Accepted:

02/05/2022

ABSTRACT

Grounded in Vygotsky's sociocultural theory of mind and the learner-centered approach to second/foreign language acquisition (SLA), this study investigated the extent to which the embedded differentiated instructions and diagnostic assessment, being mediated on Google Meet™ computer-mediated communication platform, would impact the improvement of mixed-ability English-as-a-Foreign-Language (EFL) learners' English words pronunciation and their degree of engagement in language learning. In a repeated-measures research design, an intact group of 66 EFL learners were partitioned into three tiers of higher, mid- and lower achievers to complete a virtual pretest of listening comprehension, followed by a series of parallel tiered performance tasks of English words pronunciation on a weekly basis. Their task outcomes were subsequently subjected to collective computer-mediated diagnostic assessment. After 10 sessions of intervention, the participants performed on an immediate virtual posttest of listening comprehension, and a post hoc interview. The results of mixed between-within subjects analysis of variance (ANOVA) indicated both the significant learning progress by the tiers, and the outperformance of the lower achievers on the tiered tasks. The statistical results of an analysis of covariance (ANCOVA) similarly reported significant improvement of the tiers' performance on the pretest-posttest summative assessment in this study. The inductive content analysis of the participants' responses to the structured interview elicited seven themes which were interpreted as the participants' strong approval of the usefulness of differentiated instructions, effectiveness of diagnostic assessment, and successful appeal of Google Meet platform.

Keywords: Computer-Mediated, Diagnostic Assessment, Differentiated Instructions, Google Meet, Mixed-ability

Cite this article: Rafi, F., Pourdana, N., & Ghaemi, F. (2022). Computer-mediated diagnostic assessment of mixed-ability EFL learners' performance on tiered tasks: Differentiating mediation on Google Meet™. *Journal of Modern Research in English Language Studies, 9*(2), 1-26. DOI: 10.30479/jmrels.2021.16118.1950

Publisher: Imam Khomeini International University

1. Introduction

The majority of second/foreign language (L2) teachers have to face the daily dilemma of serving diverse L2 learners in regular classrooms. Demographic heterogeneity such as giftedness, motivation, multiculturalism, or a mixed-ability classroom where an advanced L2 learner is sitting beside an underachiever requires the L2 teachers to manage learner diversity on their own (Tomlinson et al., 2003). Among the possible pedagogical solutions, differentiated instructions (DI) have recently received strong supports in second language acquision (SLA) pedagogy and research (Chen, 2007; Pourdana & Shahpouri Rad, 2017; Ur, 2005).

Mixed-ability language classrooms can be problematic unless the L2 teachers deliberately and constantly locate the learner differences and to provide them with suitable instructions which accommodate their various levels of readiness, personality, motivation and learning styles (Gomma, 2014). In this unmanageable situation, developing DI in order to address various levels of readiness and learning profiles in L2 learners is recommended as an ideal solution by several educators (de Graaf et al., 2018; Levy, 2008; Mahoney & Hall, 2017; Tomlinson et al., 2003). However, the L2 teachers' paucity of attention to learner diversity in real classroom contexts is largely evident in SLA literature (Nunley, 2006; Yatvin, 2004).

The Vygotskyan sociocultural approach to DI recommends scaffolding the L2 learners with working on a complex task, and mediating them *individually* to accomplish the task, until their learning needs are fulfilled (Vygotsky, 1987). As a result, DI is commonly operationalized by a detailed and procedural attention to the dynamic differences among L2 learners, and can take various forms, such as collaborative learning (Natsir & Asrawiah, 2013), tiered performance tasks (Tomlinson, 2014), and diagnostic assessment (Tomlinson et al., 2003). From the practical perspective, both DI and diagnostic assessment can suitably incorporate the teachers' diligent focus on the L2 learners' strengths and weaknesses (Ruiz-Primo, & Furtak, 2007; Yin et al., 2012).

Despite the fact that there are growing interests in *when* L2 teachers diagnose their students' learning difficulties, there is little solid evidence on the pedagogical aspects of diagnostic assessment in the SLA literature (Alderson, 2005). There is even less in-depth understanding of *how* L2 teachers can diagnose learner problems, specifically in the oral skills. In other words, the opportunity to embed the DI and diagnostic assessment in L2 context is still uncommon and under-documented, because it is argued to be implausible in practice (Martin & Miller, 2003). By the same token, despite the extensive SLA research on the potentials of computer-mediated communication (CMC), a few studies incorporated CMC as a tribune of the

DI and/or diagnostic assessment in the absence of face-to-face modality of interaction (Shekary & Tahririan, 2006). Therefore, it is a venue which calls for more in-depth research.

2. Literature Review

2.1. Differentiated Instructions in SLA

Inspired by the sociocultural theory (SCT) of mind, DI is known more as a pedagogical rather than a methodological approach to teaching language. *Differentiation* can be defined as a set of integrated strategies the L2 teachers can use (1) to adjust their teaching routines, curriculum contents, and tasks; (2) to respond to the student learning needs; and (3) to equally escalate learning opportunities for every individual student in the classroom (Bremner, 2008). Informed by the humanistic and learner-centered approaches to SLA, and as one of the dynamic DI strategies, *tiering* the tasks involves teaching similar language contents to the whole class, but assessing individual students with graded tasks well-suited to their cognitive capacities (Hogan, 2009). Theoretically, therefore, the "tiered tasks should engage students slightly beyond what they find easy or comfortable in order to provide genuine challenge and to promote their continued learning" (Buck, 2001, p. 53).

In a tiered classroom, L2 learners can collaborate or work in solidarity within the teacher-assigned tiers, while the tier membership is carefully determined according to the students' level of academic achievement. In designing an exemplar tiered performance task, the participants in Tier 1 are higher achievers whose level of language proficiency goes beyond the norm of the class; those in Tier 2 are mid-achievers whose level of readiness is usually defined as the norm of the class; and students in Tier 3 are lower achievers who can approach the norm only with the teacher assistance and mediation. Accordingly, while Tier 1 needs to engage in deeper and more complex contents, Tier 2 needs the teacher guidance and support occasionally, and Tier 3 demands constant teacher/peer scaffolding to complete the tasks.

2.2. Diagnostic Assessment in Perspective

The educational testing system has recently identified the vitality of the L2 learners' accountability for their own learning needs and weaknesses (Chen, 2007). Nonetheless, by evidence, it has fallen short to provide diagnostic assessment of the test takers to inform them of their accountable educational profiles (Brown & Hudson, 1998). As a result, the testing professionals call for more diagnosis in "guiding learning, improving instruction, and evaluating students' progress" (Mislevy et al., 2003, p. 18). As Gorin (2007) properly argued, diagnostic assessment can also provide rich

qualitative data about the individuals' "cognitive abilities, psychological pathologies, and personalities" (p. 7).

An early contribution to the discussion of L2 diagnostic assessment goes back to Spolsky (1992). He classified the language diagnostic tests into the educational and curricular types, but argued that "the traditional interest in curriculum did not lead to a concern for diagnosis, which was assumed to be a matter for the classroom teacher" (p. 30). Shohamy (1992) expanded upon the centrality of diagnostic testing to the curriculum to entail the diagnostic tests as the reliable reference to the observed language learning progress.

Promoting the computer-based diagnostic assessment, Hughes (2003) described what digital diagnostic tests should look like and offered a possible solution to the numerous problems of designing diagnostic tests. Following Hughes (2003), Alderson (2005) listed a number of hypothetical features of diagnostic assessment, such as (a) providing explicit focus on the remedies in future performance, (b) running detailed analysis of the problematic responses to certain items or tasks, (c) being more likely discrete-point than integrative, (d) focusing more on 'low-level' language skills such as sounds discrimination or letter-sound correspondence, and (e) casting immediate diagnostic feedback. Recently, in a seminal work, Jang and Wagner (2014) compared the L2 diagnostic assessment to the traditional test feedback, by their argument that traditional feedback is commonly product-oriented and relies upon the test scores or other summative information, whereas diagnostic feedback is more specific and learner-oriented which targets L2 learners' language processes, cognitive strengths and weaknesses. Therefore, the L2 learner can rigorously use this input to "confirm, add to, overwrite, tune, or restructure information in memory, whether it is domain knowledge, metacognitive knowledge, tasks, or cognitive strategies" (Butler & Winne, 1995, p. 275).

It has been argued that in SLA research and pedagogy, the diagnostic assessment of oral language skills is more widespread than reading and writing as literacy skills (Alderson, 2005; Pourdana et al., 2021). This is partly because the L2 speakers' inaccuracies are more immediately noticed, either in their oral use of the language or their reciprocity in listening. Yet, because SLA researchers have largely ignored the L2 learner's role in processing diagnostic feedback, so little is known about how they could integrate diagnostic feedback into learning development of language skills. So that, the available checklists, classroom observation grids, and approaches to provide diagnostic assessment seem to be intuitive, holistic, and ad hoc. On the other hand, those L2 learners who struggle to overcome weaknesses in their performance do not usually configure a uniform group which itself

makes the situation worse (Alderson, 2005). Therefore, the development of suitable diagnostic instruments is one of the aims of SLA research.

The notion that diagnosis should occur step-wise and procedural (Alderson, 2005) is relevant to the diagnostic assessment of oral skills at the level of letter-sound correspondence in speaking and listening in L2 classroom practice. By definition, sounds articulation consists of lower-level and higher-level processes, which both are relevant to diagnostic assessment. Lower-level processes enable the L2 learners to recognize the sounds of a language, lexical segments, and syntactic parsing, while the higher-level processes involve the (meta)cognitive strategies, inferences, and monitoring intelligibility. Metacognitive strategies play a critical role in articulating words as the good listeners are likely to plan, monitor, and evaluate their intelligibility (Atkinson, 2018).

SLA research on sounds production is within the psychomotor domain of the human learning. The psychomotor domain incorporates the progressive levels of behaviors from less complex (e.g., observation) to the more complex (e.g., mastery of articulation of certain sounds in a target language). Psychomotor skills are known as abilities which demand physical or tactile components. In other words, "rather than using the mind to think (cognition) and reflect (metacognition), or even the ability to develop social skills (affect), psychomotor behaviors are the things we do physically" (Atkinson, 2018). They need high degrees of flexibility, speed, precision, coordination and motor control.

Although the psychomotor domain taxonomy has been revisited several times by the scholars in education and psychotherapy (Anderson et al., 2001; Pohl, 2000), Atkinson's (2018) hierarchical levels of learning behaviors is by far one of the most referenced taxonomy of the psychomotor domain in SLA (Table 1). In his taxonomy, the psychomotor descriptors are used to indicate how the primary skills in an L2 learner progressively move towards the mastery, in order to safeguard the L2 teachers' diagnostic assessment of the learning progress. Accordingly, learning oral skills in the psychomotor domain proceeds from the primitive verbal behaviors (i.e., imitation) to the most complex (i.e., naturalization) level of speech production. In other words, in acquiring words pronunciation, mimicking/imitating another person's vocalization is the least demanding, whereas producing accurate and native-like sounds in a natural, consistent and fluent way is among the most demanding psychomotor tasks. Because one of the key characteristics of diagnostic assessment is the detailed analysis of L2 learner's performance, Atkinson's (2018) hierarchy of psychomotor domain might suitably serve as a theoretical framework to account for the learning processes of English words pronunciation.

Table 1

The Psychomotor Domain Taxonomy of Oral Skills (Adopted from Atkinson, 2018)

Stage	Proto-verb	Descriptor
Imitation	(to) imitate	The ability to observe, copy, replicate the speech sounds of others.
Manipulation	(to) manipulate	The ability to articulate sounds by memory or repeat/reproduce speech to prescribed standard instructions.
Precision	(to) perfect	The ability to articulate speech sounds with expertise and without interventions from others, with a high degree of accuracy with few errors.
Articulation	(to) articulate	The ability to adapt and integrate existing psychomotor skills in a non-standard way (e.g., a dialect or accent), in different or novel contexts, to assimilate.
Naturalization	(to) embody	The ability to articulate natural or near-natural speech in an automatic, intuitive or unconscious way appropriate to the context.

2.3. Communication Technologies in SLA Research and Pedagogy

The state-of-art information and communication technologies (ICT) are transforming teaching, learning, and using language on a daily basis. They are also changing the face of language assessment for diagnostic purposes. ICT has the potential to advance the efficiency of teacher diagnostic assessment and the subsequent decision making, both on-site the L2 classrooms and as a user-friendly self-assessment tool accessible to individual L2 learners (Alderson, 2005). Unlike the L2 teachers, computerized diagnosis rarely tires the students when it provides the automated synchronous feedback, or when it analyzes L2 learners' mistakes (Ritter, 2018).

Several researchers carried out numerous computer-mediated studies on teaching accuracy in pronunciation (Seferoglu, 2005), comparative studies on vowels and consonants (Wang & Munro, 2004), and case studies on native-like stress and intonation patterns (Levis & Pickering, 2004). The experimental findings supported the facilitating role of computer technologies in teaching pronunciation (Coman et al., 2020; Mahdi & Al Khateeb, 2019). Among popular computer-mediated communication (CMC) platforms, Google Meet™ (formerly known as Hangouts™) is a free-of-charge virtual video conferencing service, which has been developed by Google. Google Meet enables up to 100 participants to join a livestream meeting in a face-to-face modality of communication. They can speak, record

and share contents, photos, videos, and text messages to one another anywhere with the Internet access. From a pedagogical standpoint, Google Meet can properly generate an interactive learning environment with dynamic group sizes, such as whole class, small groups or pairs (Hismanoglu & Hismanoglu, 2011). Google Meet is also available as a digital application to download, install, and register into Microsoft Windows™, Android™, and IOS™ mobile operating systems.

2.4. The Study

This study set on a repeated measures design (Salkind, 2010) to bridge the gap in the SLA literature on DI and diagnostic assessment of English words pronunciation. In the virtual context of mixed-ability EFL classroom on Google Meet platform, the current researchers investigated the following research questions:

1. To what extent does the performance on tiered tasks embedded in diagnostic assessment have any differential impacts on the mixed-ability EFL learners' improvement of English words pronunciation?

2. To what extent does computer-mediated diagnostic assessment of the mixed-ability EFL learners' performance on tiered tasks of English words pronunciation impact their degree of engagement in language learning?

3. Method

3.1. Context and Participants

This study was carried out in the early COVID-19 pandemic outbreak in 2020, Iran. Sixty-six Iranian EFL learners (54 females, 81.81%) who were undergraduate students majoring in general psychology took part in the study. They were selected non-randomly by adopting the convenience method of sampling (Best & Kahn, 2006). Their ages ranged from 18 to 35 (M = 21.08, SD = .71) and their formal exposure to English was five years in average. Enrolled in the mandatory general English course at the university level, the participants received the language content mostly in English as the dominant medium of instruction (DMI) and randomly in Persian.

The participants' level of English language proficiency in terms of listening comprehension was determined by running the Preliminary English Test (PET): Listening sample Paper 1 (UCLES, 2004). The logic behind adopting this test was to narrow the scope of the summative assessment (i.e., the pre- and posttest) down to English words pronunciation as the dependent variable in this study. The 25 multiple-choice items of the test were converted into the Google Forms™, a free web-based survey administration software. The participants were required to take the online version of the test

in 35 minutes by listening to the downloadable audio file attached to the Google Forms (Cronbach's $\alpha = .872$).

Prior to the intervention, the participants were assigned into three equal-size ($N = 22$) tiers of higher, mid- and lower achievers, based on their obtained ranges of PET scores (Table 2). After the tiering assignment, the PET scores were compared across the tiers and no statistically significant between-group differences were found (F (2, 63) = .878, p = .52). In the following ten treatment sessions on Google Meet platform, the participants remained in their assigned tiers.

Table 2

Demographics of Participants

Proficiency level	PET score range	Tier	Gender (n)	Studying English (year)
Higher Achiever	21-25	1	Female, (20) 90.90%	> 6
Mid-Achiever	17-20	2	Female, (17) 77.27%	4-5
Lower Achiever	10-16	3	Female, (17) 77.27%	2-3

This study was conducted by three university professors whose Ph.D. was in Teaching English as a Foreign Language (TEFL), and had been teaching various EFL and EAP courses for 14 years. The researchers collaborated on data collection, diagnostic assessment and content analysis of the recorded interviews.

3.2. Materials and Instruments

Three sets of tiered performance tasks of English words pronunciation ($N = 3 \times 10$ sessions) were developed from the predetermined course content material (*Select Readings: Intermediate and upper-intermediate*, Lee & Bernard, 2011) and validated in a pilot study. Atkinson's (2018) hierarchical levels of learning behaviors was adopted as the theoretical framework to tier the developed tasks. Therefore, the tiered tasks were graded in terms of their modality (i.e., recognition, recognition-production, production) and the ascending psychomotor complexity demands they inhered. Accordingly, the tasks in Tier 1 were presumed having the highest psychomotor demand of naturalization in the production mode; the tasks in Tier 2 having the moderate psychomotor demand of precision and articulation in the integrated modes of recognition-production; and the tasks in Tier 3 having the lowest psychomotor demand of imitation and manipulation in the recognition mode (Table 3).

The tiered performance tasks were paralleled for their content which required the participants to recognize and/or produce a series of items testing

English words pronunciation on Google Meet platform. The content validation of the developed tiered performance tasks was carried out by five university professors majoring in TEFL to rate and review each item of the tasks based on the criteria of the appropriateness of the items, clarity of the rubric, length of tasks, and level of difficulty (Cronbach's α = .981). A revised version of tiered performance tasks was randomly selected and piloted with 42 undergraduate students similar to the main sample of participants (Cronbach's α = .821).

Table 3

The Levels of Modality and Complexity of the Tiered Performance Tasks

		Psychomotor Domain Taxonomy				
Tier	Modality	Imitation	Manipulation	Precision	Articulation	Naturalization
1	Production	-	-	-	-	+
2	Recognition-Production	-	-	+	+	-
3	Recognition	+	+	-	-	-

To summatively assess the participants' improvement in English words pronunciation, the researchers converted a virtual version of Preliminary English Test: Listening sample Paper 1 into the Google Forms and administered it as a 35-minute pretest. Similarly, the virtual version of Preliminary English Test: Listening sample Paper 2 was adopted as the immediate posttest after the 10-week intervention sessions in this study. The procedure of data collection was extended to a structured interview of the individual participants on Google Meet platform. The questions prompted their degree of engagement in (1) completing tiered performance tasks, (2) diagnostic assessment of their errors in the task outcomes, and (3) virtual learning experience on Google Meet. The responses were recorded and transcribed for the future coding and content analysis.

3.3. Procedure

In a period of 10 weeks, the experimental procedure of this study was carried out in 90-minute regular sessions of the general English course for non-English major EFL learners at the university level. The participants were a mixed-ability intact group whose performance on the tiered tasks of English words pronunciation was decided as a partial fulfillment of the course requirements. A week before the study began, the Preliminary English Test: Listening sample Paper 1 was administered online as the pretest for the purpose of assigning the participants into three tiers, followed by a virtual tutorial session on the university online classroom platform

(vadana.iauec.ac.ir). Accordingly, in 90 minutes, one of the researchers introduced the participants to the notions of tiered performance tasks and diagnostic assessment through procedural examples.

In the following weekly sessions, the time of the class was split into 45 minutes of the regular instructions to cover the general English course materials, and 45 minutes of completing the tiered tasks and providing the diagnostic feedback. The task input was presented to the three tiers separately but simultaneously on Google Meet platform. The participants were required to respond to the items, save their task output, and share it with the researcher who hosted the session. For instance, when the higher achievers in Tier 1 were supposed to pronounce the word *plausible* by reading aloud (i.e., production), the mid-achievers in Tier 2 had to listen to the similar word pronunciation, and to locate the stress on the correct syllable in the phonetic representation */plɔ:zəbl/* (i.e., recognition and production); and the lower achievers in Tier 3 were required to listen to the pronunciation of the word */'plɔ:zəbl/* and decide whether it was pronounced accurately by monitoring the phonetic representation of the word (i.e., recognition) (see Appendix 1 for a sample of tiered tasks). Every participant's performance on the tiered tasks was subjected to the diagnostic assessment in the following session. To do so, the researchers provided collective diagnostic feedback on the errors committed by the individual participant's in every tier. The common errs were given priority to receive more time and focus.

The responses to the items in tiered performance tasks were scored by assigning 1 point to the correct and 0 point to the incorrect answers. The scoring procedure and tallying the task outcomes were carried out collaboratively by the researchers (Cronbach's α = .902). Immediately after the terminal treatment session, the participants took part in a virtual version of Preliminary English Test: Listening sample Paper 2, as the posttest (Cronbach's α = .872). Finally, the participants joined a virtual one-on-one interview with the researchers on Google Meet to express detailed perceptions of their learning experience in this study by answering the three interview prompts. The interviews were recorded and analyzed inductively for coding and content analysis by the researchers. Occasional disagreements were resolved case-wise to reach a full consensus.

Following Harding et al. (2015), diagnostic assessment in this study comprised four stages of (1) monitoring the participants' English sounds perception and articulation, (2) administering tiered performance tasks, (3) providing diagnostic assessment to resolve the individual participants' errors in the task outcomes (i.e., formative assessment), and (4) teacher post-intervention decision-making (i.e., summative assessment). Moreover, the *learner engagement* in this study was conceptualized as a metaconstruct or a holistic framework in which the students' self-assessment, critical thinking,

motivation, and enthusiasm were compounded in order to reach the intended language learning goals (Fredricks et al., 2004). The degree of learner engagement was assessed by the inductive content analysis of the responses to the post hoc structured interview.

4. Results and Discussion

The objective of the first research question was two-fold: investigating *how* diagnostic assessment embedded in the tiered performance tasks would impact the mixed-ability EFL learners' progress in learning English words pronunciation, and *what* difference this intervention would cause in participants' performance on the pretest-posttest summative assessment in this study.

4.1. Results

To address the first research question in this study, the tiered tasks outcomes and the total pre- and posttest scores were inserted into the Statistical Package for Social Sciences (SPSS) version 25, for running the test of normality and descriptive statistical analysis (Table 4).

Table 4

Testing Normality Assumption

		Skewness		Kurtosis		Levene		
		Statistic	Std. Error	Statistic	Std. Error	Statistic	df	Sig.
ask	1	-.336	.295	.376	.582	3.139	60	.051
	2	.502	.295	-.218	.582	1.960	60	.150
	3	.798	.295	.314	.582	.482	60	.201
	4	.852	.295	.964	.582	2.613	60	.082
	5	.852	.295	.964	.582	2.613	60	.063
	6	.331	.295	.150	.582	1.813	60	.172
	7	.331	.295	.150	.582	1.813	60	.172
	8	.029	.295	-.713	.582	3.303	60	.501
	9	.060	.295	-.345	.582	.521	60	.541
	10	.030	.302	-.787	.595	1.408	60	.253
Test	Pretest	.172	.295	-.010	.582	2.186	63	.342
	Posttest	-.629	.295	4.898	.582	10.696	63	.102

As the ratios of skewness and kurtosis were lower than ± 1, the normality of the data was retained (Bryne, 2010). The assumption of homogeneity of variances for the tiers' outcomes on Tasks 1 to 10, as well as the overall pre- and posttest scores was also met, referring to the indices of Levene's test of equality of error variances in Table 4.

As Table 5 reports, the tiers members showed progressive performance on Task 1 to Task 10. The observed progress, however; was more noticeable for Tier 3 where the lower achievers ended up with the

highest overall performance on tiered tasks (*M* = 6.608, 95% *CI* [5.519, 6.900]) relative to Tier 1 (*M* = 6.490, 95% CI [6.285, 6.692]) and Tier 2 (*M* = 5.724, 95% CI [5.862, 5.929]). Moreover, the participants in all tiers showed an overall improvement on their posttest performance (*M* = 20.17, *SD* = 2.33, 95% *CI* [19.59, 20.74]) relative to their pretest (*M* = 17.05, *SD* = .192, 95% *CI* [16.66, 17.43]) after receiving the intervention.

Table 5

Descriptive Statistics

				95% Confidence Interval	
Tiers	Task	Mean	Std. Error	Lower Bound	Upper Bound
1	1	5.667	.204	5.259	6.074
	2	5.667	.168	5.330	6.003
	3	5.952	.159	5.634	6.271
	4	5.952	.149	5.654	6.251
	5	5.952	.149	5.654	6.251
	6	6.476	.144	6.187	6.765
	7	6.476	.144	6.187	6.765
	8	7.381	.170	7.041	7.721
	9	7.381	.138	7.105	7.657
	10	8.000	.225	7.550	8.450
Total		6.490	.103	6.285	6.696
2	1	4.762	.204	4.354	5.169
	2	4.762	.168	4.425	5.099
	3	4.762	.159	4.443	5.081
	4	5.095	.149	4.797	5.393
	5	5.095	.149	4.797	5.393
	6	5.905	.144	5.616	6.194
	7	5.905	.144	5.616	6.194
	8	6.714	.170	6.374	7.055
	9	6.762	.138	6.486	7.038
	10	7.476	.225	7.026	7.926
Total		5.724	.103	5.519	5.929
3	1	4.238	.204	3.831	4.646
	2	4.476	.168	4.139	4.813
	3	4.476	.159	4.157	4.795
	4	5.524	.149	5.226	5.822
	5	5.524	.149	5.226	5.822
	6	6.476	.144	6.187	6.765
	7	6.476	.144	6.187	6.765
	8	6.810	.170	6.469	7.150
	9	7.810	.138	7.533	8.086
	10	8.857	.225	8.407	9.307
Total		6.608	.103	5.862	6.272
Pretest		17.05	.192	16.66	17.43
Posttest		20.17	.287	19.59	20.74

A mixed between-within subjects analysis of variance (ANOVA) (Tabachnick, & Fidell, 2013) – as an extension of the repeated measures designs - was run to cross-examine the progressive performance of the Tiers 1, 2, 3 on Tasks 1 to 10, as well as their performance on pre- and posttest of summative assessment. In other words, the observed differences in between-within subjects (tiers) could indicate the development in English words pronunciation in higher, mid-, and lower achievers across the time factor (i.e., 10 successive tasks).

Among the assumptions in the mixed between-within subjects ANOVA are (1) the equivalence of covariance matrices, and (2) the assumption of sphericity. The results of the Box's M statistics ($M = 14.197$, $p = .350 > .001$) indicated that the assumption of equivalence of covariance matrices was retained, and the Mauchly's test of sphericity, (χ^2 (2) = .946, $p = .178 > .05$) indicated that the assumption of sphericity in the data was also met.

Table 6

Multivariate Tests of Within-Group Effect: Performance on Tiered Tasks

Source		Value	F	Hypothesis df	Error df	Sig.	Partial η^2
Task	Pillai's Trace	.906	73.94	7.00	54.00	.000	.906
	Wilk's Lambda	.094	73.94	7.00	54.00	.000	.906
	Hotelling's Trace	9.585	73.94	7.00	54.00	.000	.906
	Roy's Largest Root	9.585	73.94	7.00	54.00	.000	.906
Task * Tiers	Pillai's Trace	.736	4.574	14.00	110.00	.000	.368
	Wilk's Lambda	.328	5.763	14.00	108.00	.000	.428
	Hotelling's Trace	1.858	7.035	14.00	106.00	.000	.482
	Roy's Largest Root	1.747	13.730	7.000	55.00	.000	.636

As Table 6 reports the results of the multivariate tests of within-group effect (task * tier), not only the participants in all tiers showed considerable improvement in their task outcomes (Wilks' Lambda = .094, F (7, 54) = 73.94, $p = .00$, Partial $\eta^2 = .906$, interpreting a large effect size) (Lenhard & Lenhard, 2016), but also they showed a significant interaction effect between the tier membership and the task outcomes (Wilks' Lambda = .328, F (14, 108) = 5.763, $p = .00$, Partial $\eta^2 = .428$, interpreting a large effect size). In order to further explore the between-group effect of the tier membership (higher, mid- and lower achiever) on the task outcomes, Table 7 and Figure 1 reported the significant difference (F (2, 60) = 14.033, $p = .000$, Partial $\eta^2 = .319$ representing a large effect size) among the three tiers.

Table 7

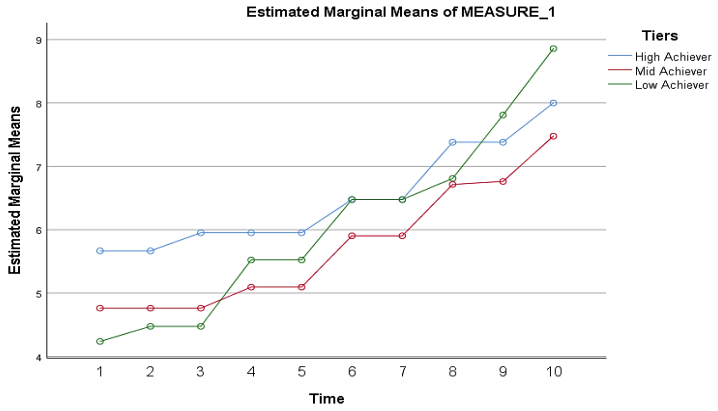
Multivariate Tests of Between-Group Effects: Performance on Tiered Tasks

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial η^2
Intercept	23393.525	1	23393.525	10599.008	.000	.994
Tiers	61.946	2	30.973	14.033	.000	.319
Error	132.429	60	2.207			

As it can be seen in Figure 1, all tiers showed considerable improvement in their task outcomes, while the lower achievers in Tier 3 were the most beneficiary. On the other hand, the higher and mid-achievers in Tiers 1 and 2 demonstrated a relatively expected pattern of improvement in their task outcomes.

Figure 1

Task Outcomes by Tiers



To further explore the performance of the three tiers on the pretest-posttest summative assessment in this study, a test of analysis of covariance (ANCOVA) was conducted, controlling for the non-random selection effect of the participants (Table 8).

Table 8

Multivariate Tests of Between-Group Effects: Performance on the Pretest and Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial η^2
Tier	106.295	8	13.287	3.068	.006	.301
Covariate (Pretest)	26.628	7	3.804	.878	.529	.097
Error	246.871	57	4.331			
Total	27195.000	66				
Corrected Total	353.167	65				

In Table 8, after controlling for the effect of the pretest (i.e., the covariate), the statistical results indicated that the participants in all tiers had a significant improvement on their posttest performance after the received intervention (F (8, 57) = 3.068, p = .006, partial η^2 = .301, representing a large effect size).

The second research question addressed the EFL learners' engagement in terms of their perceptions of the new experience they had in the embedded tiered performance tasks and diagnostic assessment on virtual Google Meet platform. The responses to the structured interview prompts were subjected to the interpretational analysis of the frequency counts (Tesch, 1990). The The second research question addressed the EFL learners' engagement in terms of their perceptions of the new experience they had in the embedded tiered performance tasks and diagnostic assessment on virtual Google Meet platform. The responses to the structured interview prompts were subjected to the interpretational analysis of the frequency counts (Tesch, 1990). The researchers improvised an open coding system to extract as many themes as possible out of the participants' responses (Appendix 2).

The three extracted themes for Prompt 1 included the participants' references to *the usefulness of tiered performance task* as being learner-centered versus teacher-dominating (N = 35), reducing task anxiety (N = 24), and generating more learning improvement than one-fits-all classroom tasks (N = 21). One of the major arguments of the participants was against the L2 teachers' regular misconception of the equality of the students' needs, interests, and potentials, which they perceived as one of the sources of frustration to the underachievers. They also indicated that working on tiered tasks which were often compatible to their level of readiness and language knowledge would benefit their language learning and boost their learning autonomy. The elicited key words were *fair*, *personal*, and *self*. The second extracted theme was the benefit of tiered performance tasks to lower debilitative task anxiety. The participants frequently referred to the engaging and relaxing environment which was created by completing the tasks tailored to their language potentials. The elicited key words for this theme were *warm*, *peaceful*, and *low-stress.* The final extracted theme was the effectiveness of tiered tasks to produce observable progress. The participants made frequent references to the gradual improvement in their English words pronunciation, their progressive ability to self-regulate, and monitor their own mispronunciations. The recorded key words were *effective, helpful* and *practical.*

The elicited themes for Prompt 2 addressed the *effectiveness of diagnostic assessment* with attributes of being detail-oriented and precise (N = 30), and encouraging student self-monitoring (N = 22). The major argument the participants expressed on the merits of the teacher diagnostic

assessment was the individualized and focused nature of the diagnosis which enabled them to precisely apply the feedback to resolve the target word mispronunciations. Moreover, they perceived the detail-oriented diagnostic feedback as facilitating which eventually mediated self-monitoring. The reference key words were *to-the-point*, *self-assessment*, and *learning.*

The extracted themes for Prompt 3 indicated the *appeal of Google Meet* platform by participants' references to its modernity ($N = 50$) and user-friendliness ($N = 27$). The majority of the participants came to the consensus that Google Meet created an engaging, smart and comfortable environment for their learning experience. By comparing Google Meet to the popular CMC platforms such as Zoom™ or Skype™, they approved its better accessibility and user-friendliness. The elicited key words were *user-friendly*, *easy*, *fast*, and *online*.

4.2. Discussion

To aggregate the findings in this study, the researchers found that when the mixed-ability EFL learners - who were partitioned into higher, mid- and lower achieving tiers - completed tiered performance tasks of parallel contents, followed by diagnostic assessment on Google Meet virtual classroom, they experienced noticeable progress in learning English words pronunciation on both formative and summative assessment. Yet, the lower achievers outperformed the higher and mid-achievers. Moreover, the participants' active engagement in language learning experience was perceived in their approval of the usefulness of tiered performance tasks, effectiveness of teacher diagnostic assessment, and the successful appeal of the Google Meet platform.

The discussion of the first research question is based on the statistical results which indicated that the lower, mid-, and higher achievers had steady and gradual progress in completing the tiered performance tasks. Moreover, the received intervention (i.e., the embedded tiered tasks and diagnostic assessment) caused differential impacts on the improvement of the tiers' learning words pronunciation. The research findings deemed the *cohort* impacts of the differentiated instructions and follow-up diagnostic assessment in a mixed-ability EFL context, where the L2 learners with various levels of readiness needed distinctive 'frames of reference' in their language learning (Henson, 2003). In this regard, when the academic diversity of the students was addressed through 'adaptation', in terms of DI and diagnostic assessment, it was inevitable that those who demanded more (i.e., the lower achievers) would benefit more from the newly transformed routines (Levy, 2008; Tomlinson et al., 2003). Moreover, 'the lower-level' nature of the diagnostic assessment (Alderson, 2005) which targeted the English words

pronunciation could properly meet the preliminary needs of the lower achievers than the self-sufficient mid- and higher achievers.

The results could be anchored to the Vygotskyan sociocultural theory of mind by reminding the concept of 'zone of proximal development' (ZPD). Inspired by the constructivist view of learning, Vygotsky (1987) insisted in matching the learning materials to the learners' capacity, by developing tasks at a suitable level "to stretch the learner's ability, but not to cause detrimental frustration" (Chen, 2007, p. 31), and paving the way for sustainable progress and enjoyment in the educational setting. In the same vein, the role of diagnostic assessment to produce observable language behaviors was circumstantial in the language learning progress of the tiers. As Gorin (2007) attributed the diagnostic assessment with the potential of 'immediate penetration', the obtained psychometric data from the participants' tiered task performance could scaffold them to bridge their weaknesses and to improve in learning words pronunciation.

The findings of this study on the effectiveness of tiered performance tasks were supported by Chen (2007) and Ritter (2018), while were partially contradicted by Pourdana and Shahpouri Rad (2017). To explore the college students' perspectives to the applicability of DI in EFL Taiwanese context, Chen (2007) collected data from 12 participants through a number of qualitative instruments, such as observation. Their responses were affirmative to the tiered performance tasks as an authentic, motivating and engaging summative assessment. The rate of approval was evident more by the lower achieving (82%), than higher achieving students (59%). The findings in Ritter (2018) who conducted separate case studies to explore high school teachers' perceptions of using digital contents or CMC platforms, reported the teachers' approval of the benefits of using educational technology to DI by creating dynamic assignments tailored to individual students' proficiency levels and interests. The findings in this study were in contrast to Pourdana and Shahpouri Rad (2017) whose findings in a case study with 46 mixed-ability EFL learners indicated the usefulness of DI but failed to show any significant association between the tiered performance tasks outcomes and the participants' mixed levels of language proficiency. Yet, their justification for unexpected results was the small sample size and the dynamic tier membership which likely caused data pollution in the study.

The substantial contribution of the diagnostic assessment to the participants' tired task outcomes was supported by Ardin (2018), and Nikmard and Tavassoli (2020). Ardin (2018) investigated the effects of diagnostic assessment of descriptive and narrative genres of writing on 40 EFL learners writing achievement, and reported the large impacts of diagnostic assessment on both writing genres. Similarly, to investigate the effect of diagnostic assessment on selective and productive reading tasks,

Nikmard and Tavassoli (2020) studied 60 EFL learners in a pretest-posttest research design and reported the significant improvement in both selective and productive tasks outcomes with pedagogical implications of regular diagnostic assessment to EFL reading comprehension practice.

The discussion of the second research question which queried the degree of engagement of the participants in completing tiered tasks, diagnostic assessment and their virtual learning experience is conclusive to the analytical results of the structured interview. The participants' general approval of the DI and diagnosis on Google Meet platform was largely supported by Doe (2015), Mahdi and Al Khateeb (2019), while was in contrast with findings by Colby-Kelly and Turner (2007). In a qualitative research, Doe (2015) examined how EFL learners would interpret diagnostic feedback and reported that while at the beginning of the course the students were skeptical about its benefits, they eventually interpreted the teacher diagnosis as appropriate and facilitating. Mahdi and Al Khateeb (2019) also promoted the computer-assisted pronunciation training (CAPT) in a meta-analysis of 20 experimental researches with 1014 L2 learners, and reported the strong impact (d = .68) of computer-assisted training and diagnosis of English pronunciation on L2 young and adult learners' mastery and sustained motivation, although it seemed more effective with lower-level than advanced L2 learners. Coming up with contradictory findings, in an assessment *for* learning (AFL) context, Colby-Kelly and Turner (2007) explored the evidence for an Assessment Use Argument (AUA) (Bachman & Palmer, 2010) of the diagnostic assessment from the L2 teachers' and students' perspectives. They reported mixed impressions towards diagnostic assessment, such as being "motivating and essential, unmerited, embarrassing, or untrustworthy" (p. 30).

5. Conclusion and Implications

Dealing with the diversity of L2 learners is largely under-documented in SLA research and general education. A challenging dilemma is posed by the L2 teachers' hesitancy to tailor their teaching routines to the demanding oral skills, or to implement regular formative assessment of L2 learners. As Mehlinger (1995) properly argued, to "customize schooling for individual learners, rather than mass produce students is not a superficial change; but a deep cultural change" (p. 154). Yet, few teachers would like to make radical adjustments to their teaching practice in response to the learners' diversity. Therefore, this study which explored how DI and diagnostic assessment might successfully work in mixed-level and diverse educational contexts seems a promising venue with future pedagogical implications.

The current study suggests that the EFL learners can largely benefit from DI that targets their myriad of needs, goals and weaknesses. Therefore,

the L2 teachers are recommended to face the inconvenience of making their classroom a good fit for various L2 learners, by adopting a wider range of the differentiation strategies, and a more inclusive set of classroom management conventions. One of the conceptual frameworks which easily come to grips with DI is diagnostic assessment which addresses "a cognitive gap between a current level of performance and some desired level of performance or goal" in L2 learners (Westbroek et al., 2020, p. 109). Therefore, the research on language assessment needs a change of direction to become more diagnostic in practice. In other words, diagnostic assessment should become an integral part of the language curriculum, in-service professional development programs, or at least of L2 teaching practice. Filling this gap most likely encourages the L2 learners to invest their higher level of efforts in language learning.

The arguments in this research are still speculative due to some logistic and operational limitations. One of the major restrictions imposed on this study was the COVID-19 pandemic which caused countless readjustments to the researchers' contacts for data collection, content analysis, and regular discussions. Likewise, the researchers were aware that collecting a large body of data in the virtual classroom sessions, where the researchers' access to monitor the participants' task performance was minimal, could have failed to prevent data pollution. From the academic research design perspective, the researchers did not intend to isolate the effects of DI from diagnostic assessment by including a comparison group in this study. Nor did they plan to examine the sustained impacts of the conducted intervention by running a delayed posttest. As a result, the reported findings might be used with necessary precautions.

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* Continuum.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. Longman.

Ardin, M. (2018). *The effect of diagnostic assessment vs. dynamic assessment on EFL learners' descriptive and narrative writing.* Unpublished MA Thesis, Islamic Azad University, Karaj Branch, Iran.

Atkinson, S. P. (2018). Developing effective learning outcomes. Retrieved from *https://sijen.com/research-interests/8-stage-learning-design-framework/4-intended-learning-outcomes-ilos*

Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.

Best, J. W. & Kahn, J. V. (2006). *Research in education (3rd ed.).* Pearson Education Inc.

Bremner, S. (2008). Some thoughts on teaching a mixed ability class. *Scottish Languages Review*, *18,* 1-10.

Brown, J. D. & Hudson, T. D. (1998). The alternatives in language assessment: advantages and disadvantages. *TESOL Quarterly, 30*, 653-675.

Buck, G. (2001). *Assessing listening.* Cambridge University Press.

Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming (2nd ed.).* Routledge Taylor & Francis Group.

Chen, Y. H. (2007). *Exploring the assessment aspect of differentiated instruction: College EFL learners' perspectives on tiered performance tasks.* University of New Orleans Theses and Dissertations.

Colby-Kelly. C. & Turner, C. E. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *Canadian Modern Language Review, 64*(1), 9-37.

Coman, C., Tîru, L. G., Mesesan-Schmitz, L., Stanciu, C., & Bularca, M. C. (2020). Online teaching and learning in higher education during the Coronavirus pandemic: Students' perspective. *Sustainability,* 1-24.

de Graaf, A., Westbroek, H., & Janssen, F. (2018). A practical approach to differentiated instruction: How biology teachers redesigned their genetics and ecology lessons. *Journal of Science Teacher Education, 30*(1), 6–23.

Doe, C. (2015). Student interpretations of diagnostic feedback. *Language Assessment Quarterly, 12*(1), 110-135.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*, 59–109.

Gomma, O. M. K. (2014). The effect of differentiating instruction using multiple intelligences on achievement in and attitudes towards science in middle school students with learning disabilities. *International Journal of Psycho-educational Sciences, 3*(3), 109–117.

Gorin, J. S. (2007). Test Construction and diagnostic testing. In J. Leighton & M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*, (pp. 173 - 201). Cambridge University Press.

Harding, L., Alderson, C. J., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing, 32*(3), 1-20.

Henson, K. (2003). Foundations for Learner-Centered Education: A Knowledge Base. *Education, 124*(5), 3-13.

Hismanoglu, M., & Hismanoglu, S. (2011). Task-based language teaching: what every EFL teacher should do. *Procedia-Social and Behavioral Sciences*, *15*, 46-52.

Hughes, A. (2003). *Testing for language teachers (2nd ed.).* Cambridge University Press.

Hogan, R. E. (2009). Differentiated instruction and tiered assignments. *Mathematical and Computing Sciences Masters, 3*(5), 42-49.

Jang, E. E., & Wagner, M. (2014). Diagnostic feedback in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment: Approaches and development* (pp. 157-175). John Wiley & Sons, Inc.

Lenhard, W., & Lenhard, A. (2016). *Calculation of effect sizes.* Bibergau.

Lee, L., & Bernard, J. (2011). *Select Readings: Intermediate and Upper-intermediate*. Oxford University Press.

Levis, J., & Pickering, L. (2004). Teaching Intonation in Discourse Using Speech Visualization Technology. *System, 32*, 505-524.

Levy, H. M. (2008). Meeting the needs of all students through differentiated instruction: Helping every child reach and exceed standards. *Clearing the House: A Journal of Educational Strategies, Issues and Ideas, 81*(4), 161-164.

Martin, D., & Miller, C. (2003). *Speech and language difficulties in the classroom.* David Fulton.

Mahdi, H. S., & Al Khateeb, A. A. (2019). The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education, 7*(3), 733-753.

Mahoney, J. & Hall, C. (2017). Using technology to differentiate and accommodate students with disabilities. *E-Learning and Digital Media, 14*(5), 291 – 303.

Mehlinger, H. D. (1995). *School reform in the information age.* Indiana University Press.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of education assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3–62.

Natsir, R. Y., & Asrawiah, A. (2013). Improving the students' reading comprehension using tiered tasks strategy. *Exposure, 2*(1), 76-94.

Nikmard, F., & Tavassoli, K. (2020). The effect of diagnostic assessment on EFL learners' performance on selective and productive reading tasks**.** *Journal of Modern Research in English Language Studies, 7*(1), 79-104.

Nunley, K. F. (2006). *Differentiating the high school classroom: Solution strategies for 18 common obstacles*. Thousand Oaks.

Pohl, M. (2000). *Learning to think and thinking to learning*. Hawker Brownlow Education.

Pourdana, N., & Shahpouri Rad, M. (2017). Differentiated instructions: Implementing tiered listening tasks in mixed-ability EFL context. *Journal of Modern Research in English Language Studies, 4* (4), 45-63.

Pourdana, N., Nour, P., & Yousefi, F. (2021). Investigating metalinguistic written corrective feedback focused on EFL learners' discourse markers accuracy in mobile-mediated context. *Asian-Pacific Journal of Second and Foreign Language Education*, *6*(7), *doi.org/10.1186/s40862-021-00111-8*

Ritter, O. N. (2018). *Integration of educational technology for the purposes of differentiated instruction in secondary STEM education*. Ph.D. Dissertation. University of Tennessee.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84.

Salkind, N. J. (2010). *Encyclopedia of research design.* Sage.

Seferoglu, G. (2005) Improving students' pronunciation through accent reduction software. *British Journal of Educational Technology, 36*(2), 303-316.

Shekary, M., & Tahririan, M. H. (2006). Negotiation of meaning and noticing in text-based online chat. *The Modern Language Journal, 90*(4), 557-573.

Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal, 76*(4), 513-521.

Spolsky, B. (1992). The gentle art of diagnostic testing revisited. In E. Shohamy & A. R. Walton (Eds.), *Language assessment for feedback: Testing and other strategies* (pp. 29-41). Kendall/Hunt.

Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics (6th ed.)*. Pearson Education.

Tesch, R. (1990). *Qualitative research: Analysis types and software tools*. Palmer.

Tomlinson, C.A. (2014). *The differentiated classroom: Responding to the needs of all learners (2nd ed.).* Alexandria.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahn, C. M., Brimijoin, K., Conover, L. A., & Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted, 27*(2/3), 119-145.

Ur, P. (2005). *A course in language teaching*. Cambridge University Press.

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber, & A. S. Carton (eds.), *The collected works of L. S. Vygotsky: Problems of general psychology* (pp. 39-285). Plenum.

Wang, X. & Munro, M. J. (2004). Computer-Based Training for Learning English Vowel Contrasts. *System: An International Journal of Educational Technology and Applied Linguistics, 32*(4), 539-552.

Westbroek, H. B, van Rens, L., van den Berga, E., & Janssen, F. (2020). A practical approach to assessment for learning and differentiated instruction. *International Journal of Science Education, (42)*6, 955–976.

Yatvin, J. (2004). *A room with a differentiated view: How to serve ALL children as individual learners*. Heinemann.

Yin, M., Sims, J., & Cothran, D. (2012). Scratching where they itch: Evaluation of feedback on a diagnostic English grammar test for Taiwanese university students. *Language Assessment Quarterly, 9*(1), 78–104.

Appendix 1: A Sample of Tiered Tasks

Tier 1 (High Achievers)

1. Pronounce and record the following words.
A. dem. on. strate B. prot. ag. on. ist
C. sub. sti. tute D. ob. so. lete
E. re. ac. tion. ary F. im. mi. grant
G. a. nal. o. gous H. ex. po. nen. tial
I. crim. in. olo. gist J. plau. si. ble

Tier 2 (Mid-achievers)

1. Pronounce and record the following words orally and decide whether their stress patterns are correct (C) or incorrect (I).
A. dem. on. ′strate ----------
B. ′prot. ag. on. ist ----------

2. Pronounce and record the following words and underline the syllable that gets primary stress.
A. sub. sti. tute B. ob. so. lete

3. Pronounce and record the following words, then decide whether the given pronunciations are correct (C) or incorrect (I).
A. re. ac. tion. ary / ri ′ækʃəneri / ----------
B. im. mi. grant / ′ɪmegrənt / ----------

4. Look at the following words, then underline the correct pronunciation.
A. / æn′æləgəs / / ə′næləgəs / / ɛ′næləgəs/
B. /ekspə′nenʃl / /ekspəʊ′nenʃl/ /ɪkspə′nenʃl/

5. Pronounce and record the following words and choose the syllable(s) with /a:/ sound.
A. crim. in. olo. gist B. plau. si. Ble

Tier 3 (Low achievers)

1. Listen to the recorded words then decide whether the following stress patterns are correct (C) or incorrect (I).
A. dem. on. ′strate ----------
B. ′prot. ag. on. ist ----------

2. Listen to the recorded words and underline the syllable that gets primary stress.
A. sub. sti. tute
B. ob. so. lete

3. Listen to the recorded words, then decide whether the given pronunciations are correct (C) or incorrect (I).
A. re. ac. tion. ary / ri ′ækʃəneri /----------
B. im. mi. grant / ′ɪmegrənt / ----------

4. Listen to the recorded words, then underline the correct pronunciation.

A. / æn′æləgəs / /ə′næləgəs / /eˈnæləgəs/
B. /ekspə′nenʃl//ekspəʊ′nenʃl/ /ɪkspə′nenʃl/

5- Listen to the recorded words and choose the syllable(s) with /a: / sound.
A. crim. in. olo. gist B. plau. si. ble

Appendix 2: Extracted themes and their distribution in responses to the interview prompts

Example	Theme	F
1. Are tiered tasks working for you to improve your pronunciation?		
Yes. In other classes, all of the students are [seen] like equal. But we are not equal. Teachers think we are the same. But tiered tasks can solve this problem. The teacher has different tasks for different students. I really enjoy it.	- Learner-centered	35
	- Reducing stress	24
	- Effective	21
2. What do you think about individualized feedback the teacher provides on your errors?		
It helps me to pay more attention to [the] details. When I pay attention, I do the tasks better. I like to revise my errors and learn better.	- Detail-oriented	30
	- Self-monitoring	22
3. How do you like learning English on Google Meet?		
Google Meet is very good. It is [the] first time I use it. It is great I can see others these days. I can share files easily by my laptop or mobile everywhere.	- Modern	50
	- User-friendly	27