

A Phenomenography of the Constraints to Dynamic Assessment: Data-driven Insights from the Iranian Academic Context

¹Asghar Ali Mohammadimanesh

²Abbas Bayat*

³Hamid Reza Khalaji

⁴Hossein Ahmadi

Research Paper

IJEAP- 2207-1887

Received: 2022-07-08

Accepted: 2022-09-01

Published: 2022-10-20

Abstract

Prompted by increased concerns regarding the shortcomings of traditional assessment, the current study used the unorthodox qualitative method of phenomenography to disclose significant constraints to the implementation of dynamic assessment (DA), which has robust theoretical, practical, and empirical foundations. To this aim, 19 university teachers were selected purposively from different language disciplines. A semi-structured interview was designed and participants' perceptions regarding the constraints were investigated. The interviews were audio-taped, transcribed, and analyzed based on principles of relevance, frequency, position, and pregnancy and then coded systematically under relevant categories. To make findings more concrete, the most relevant excerpts from the participants accompanied each category. Member checking strategy and respondents' validation were employed to secure the results' credibility and validity. Findings revealed that constraints to implementations of DA fall into three central categories technical constraints (scoring issues, validity issues, discriminating power of the test, fairness issues, consequential credibility, and concept load), practical constraints (demanding implementation, attitudinal issues, feasibility (practicality) issues, oral nature of feedback, and examiners' literacy), and contextual constraints (conservative nature of Iranian context, and resource constraints). The phenomenographic procedure used and findings have implications for language practitioners. This study had some limitations in sampling and enactment.

Keywords: Dynamic Assessment, Fairness, Scoring, Technical Constraints, Validity

1. Introduction

In classical views of testing and assessment, standardized testing was regarded as the hallmark of the discipline, claiming that it has remedied the sources of systematic and unsystematic measurement errors found in traditional approaches toward testing. However; according to Herman, Aschbacher, and Winters (1992), overemphasizing standardized assessment within classical approaches gave rise to a growing dissatisfaction with traditional, multiple-choice forms of testing, professing that traditional assessment cannot capture the actual competence of learners since test scores do not represent significant learning outcomes. They added that improvements in test scores do not represent substantial learning outcomes, and therefore, improvements in test scores are not necessarily the result of improved teaching and learning. These deficiencies laid the groundwork for the transition from classical approaches in testing to alternative forms of assessment (Mitchel, 1992). Accordingly; Herman et al., (1992), Huerta-Macias (1995), and Brown and Hudson (1998) called for an alternative form of assessment that encompasses problem-solving and higher-level thinking, embraces

¹PhD Candidate of TEFL; aamms2@yahoo.com, Department of English, Malayer Branch, Islamic Azad University, Malayer, Iran.

²Assistant Professor of TEFL (Corresponding Author), abbasbayat305@yahoo.com; Department of English, Malayer Branch, Islamic Azad University, Malayer, Iran.

³Assistant Professor of TEFL; hrkhalaj20@gmail.com, Department of English, Malayer Branch, Islamic Azad University, Malayer, Iran.

⁴Assistant Professor of TEFL; ahmadikh@gmail.com, Department of English, Malayer Branch, Islamic Azad University, Malayer, Iran.

meaningful tasks and instructional activities, employs real-world contexts, and lays emphasis on processes, etc.

Out of the approaches to alternative assessment in language teaching, the dynamic assessment (henceforth DA) stood sensitive to the principles outlined above and serves as a theoretically embedded and experimentally founded form of alternative assessment, which finds fault with the conventional views on teaching and assessment. On the one hand, the contribution of DA to the emerging alternative assessment movement is summed up by Sternberg and Grigorenko (2002), Lidz and Gindis (2003), and Poehner (2008), maintaining that dynamic intervention is embedded within the assessment procedure which interprets individuals' abilities, and leads them to higher levels of functioning. On the other hand, DA acknowledges the fluid nature of useful knowledge (Gergen & Dixon-Roman, 2014), criticizes the traditional linear views of the knowledge (Poehner, 2007), avoids bias in favor of systematic error (Garcia & Pearson, 1993), captures a dynamic picture of the pupils' learned materials (Elliot, 2003), shows consideration for principles of post-method pedagogy (Farokhipour, Khoshsima, Sarani & Ganji, 2020), and entails further instructional activity (Farrokh and Rahmani, 2017).

Although the transition from the traditional assessment to emerging types of alternative assessment such as DA is very important, it is not a practical and widespread approach in the Iranian language teaching context despite its distinctive educational and theoretical merits. Thus, both for practical and theoretical reasons, it is very crucial to uncover the obstacle to the implementation of DA in the country. In line with Brown et al., (2011) which hold that obtaining insights from teachers might alleviate and amend the shortcomings and the problems encountered, the current research attempted to address the following research question through a phenomenographic study of Iranian language teachers; what are the main constraints to the implementation of DA in the Iranian academic language teaching setting?

2. Literature Review

2.1. Theoretical and Practical Underpinnings

DA, a sub-type of interactive assessment, is an innovative approach in education that aims at assessing human abilities and learning potential by employing some basic cognitive concepts and meta-cognitive operations that identify and remove non-intellective barriers to the expression of examinees' intelligence (Haywood & Tzuriel, 2002). Many well-established theories and practices lend support to dynamic assessment. A) Vygotsky's sociocultural theory (1978): Gredler and Shields (2007) asserted that Vygotsky characterized cognitive development as constantly undergoing change, leading us to the assumption that the essence of higher forms of thinking is the individual's intervention in the processes of thinking. To put it in a nutshell, in sociocultural theory of learning, the supported scuffled talk from the side of more significant others can bring the emerging functions into light, meaning that what the learner can do at the present time with assistance can be done by the learner in future while unassisted. B) Structural cognitive modifiability theory: according to Feuerstein et al., (2002), the theory is an explanation of Man's unique predisposition to change or modify the structure of their cognitive functioning to become tuned to changing demands of life situations. C) Learning potential assessment (Budoff (1987), the earliest practical manifestation of these theories, is a test-train-retest procedure that offers children the opportunity to demonstrate that they can perform at a level average for their peers, on a nonverbal reasoning task, when offered a short period of training on a similar task. D) Graduated prompts (Campione & Brown, 1987) proposed a graduated prompting approach, which was mainly concerned with evaluating the academic progress of weak students and a rubric for counting the number of hints each learner requires to solve the problems presented to them in an approximating fashion. Testing the limits (Carlson & Wiedl, 1992) and the "Lerntest" (Guthke et al., 1992) are other practical approaches that are derived from the DA theory.

2.2. Related Research

Increasing interest in DA in education has led to a large body of research in different educational settings (Poehner, 2005; Haywood & Lidz, 2007, Alavi, Kaivanpanah & Shabani, 2012; Birjandi et al., 2013, Farokhipour et al., 2020, p.121). To lay the groundwork for this study only some of them are reviewed due to space limits. Showing adherence to curriculum-based and task-based DA procedures, Zarinkamar, Abdi, and Davaribina (2021) conducted a study on writing skill creativity. The study revealed that both sandwich and cake models of DA influenced the learners' writing creativity. Also, Ghahderijani et al., (2021) investigated the effect of group DA compared with computerized DA in an Iranian context and reported that the computer-assisted approach yielded better score gains. Furthermore, Farokhipour et al., (2020, p.121) proposed a model for implementing DA in the real educational setting, which eliminates some of the shortcoming associated with experimental and small-scale employment of DA.

Despite the large array of empirical studies submitting theoretical and empirical evidence in favor of DA in EFL and ESL contexts, the pathology of DA for rarity of its implementation is an intact research area. Therefore, the current research is an innovative attempt to investigate the obstacles to the successful implementation of DA in Iranian language education through a phenomenographic study.

3. Methodology

Educational pathology is construed as complex and multi-dimensional phenomenon that entails different aspects and might be interpreted differently in various settings. Thus, to capture the diversity and intricacy of the obstacles to the implementation of DA in the Iranian context a phenomenographic study was employed. Marton (1988) maintains that, unlike conventional qualitative research, which adopts an observational or "noumenal" approach to describe an event in itself, the phenomenal study adopts an experiential outlook to describe the event as it is happened. This idea was labeled a "phenomenographic study" by Marton (1981). Accordingly, phenomenography is a research method for mapping the qualitatively different ways people experience, conceptualize, perceive and understand various aspects of a phenomenon (Marton & Pong, 2005).

3.1. Participants

The sampling strategy employed in this study was purposive. Teaching and research experiences concerning alternative assessment guided the selection of 19 English language teacher at the university level specialized in applied linguistics (n=10, 52.6%), English literature (n=6, 31.5 %), translation (n=1, 5.2%) and pure linguistics (n=2, 10.5%). Miles and Hubberman (1994) maintain that purposive sampling permits identifying, selecting, and categorizing most data-rich cases in qualitative design. Another aim of the purposive sampling in this study was to include teachers of various academic ranks, gender, disciplinary background, and different lengths of teaching experience. Due to the researchers' adherence to the principle of data saturation which is a critical maxim in qualitative data collection, the number of participants was sensitive to the emerging findings. In addition, to fulfill ethical considerations, participants read and signed informed consent forms administered to them. The demographic information of participants is presented in Table 1.

Table 1: Demographic Information of Participants

Participants	Gender	Academic Rank	Discipline	Years of Experience
ELT 1	F	Assistant Pr.	Applied Ling	12
ELT 2	F	Assistant Pr.	Translation	7
ELT 3	F	Assistant Pr.	Applied Ling	5
ELT 4	F	Assistant Pr.	Applied Ling	3
ELT 5	F	Associate Pr.	Applied Ling	14
ELT 6	F	Assistant Pr.	Literature	6
ELT 7	M	Assistant Pr.	Applied Ling	2
ELT 8	M	Assistant Pr.	Applied Ling	6
ELT 9	M	Assistant Pr.	Applied Ling	10

ELT 10	M	Assistant Pr.	Applied Ling	9
ELT 11	M	Associate Pr.	Applied Ling	21
ELT 12	M	Assistant Pr.	Literature	6
ELT 13	M	Associate Pr.	Applied Ling	19
ELT 14	M	Assistant Pr.	Literature	9
ELT 15	M	Assistant Pr.	Literature	2
ELT 16	M	Assistant Pr.	Linguistics	8
ELT 17	M	Associate Pr.	Literature	23
ELT 18	M	Assistant Pr.	Linguistics	17
ELT 19	M	Assistant Pr.	Literature	7

ELT = English Language Teacher

3.2. Instruments

In line with Tight (2015), Akerlind (2005), and Marton and Pong (2005), the main instrument used in the current phenomenographic study was a semi-structured interview. Aimed at steering the direction of the research and arriving at a tentative framework that guides the interview content and questions, a thorough literature review was conducted. This stage had two purposes; a) to investigate the possible existing standardized scales, tools, and instruments that might measure our construct (Dornyei & Taguchi, 2010), b) to arrive at a tentative theoretical framework that guides the interview. After an extensive review of literature, using a small-scale grounded theory, the researchers collected a bulk of core concepts and proposals revolving around the subject of the study that formed our interview questions. The guidelines of Dornyei and Taguchi (2010, p.61) for scale development and validation were followed to secure the validity of the questionnaire. Also, two assistant professors in teaching English from Islamic Azad University, Qom branch, commented on the scale, according to which the required amendments and modifications were carried out. Having designed the framework for the semi-structured interviews, the meetings were held in different spatial and temporal settings. According to Patton (1990), the devised array of questions only guided the interview sessions. Therefore, due to the spur of time and occasion, extra detailed question were also asked to elicit the most possible comprehensive, elaborate and thorough information. Interviews were held in the Persian language to avoid meaning attrition. The interviewees were assured of the privacy of their data and the possibility of voluntary withdrawal from the course of the study. Each interview session lasted about one hour, wherein audio tape was used to record data.

3.3. Procedure of the Study

The procedure of the study was guided by Svensson (1997), Marton (1984, 1986), and Dornyei and Taguchi (2010, p 61). The schematic representation of the procedure of the study that entails three main stages is presented in Figure 1.

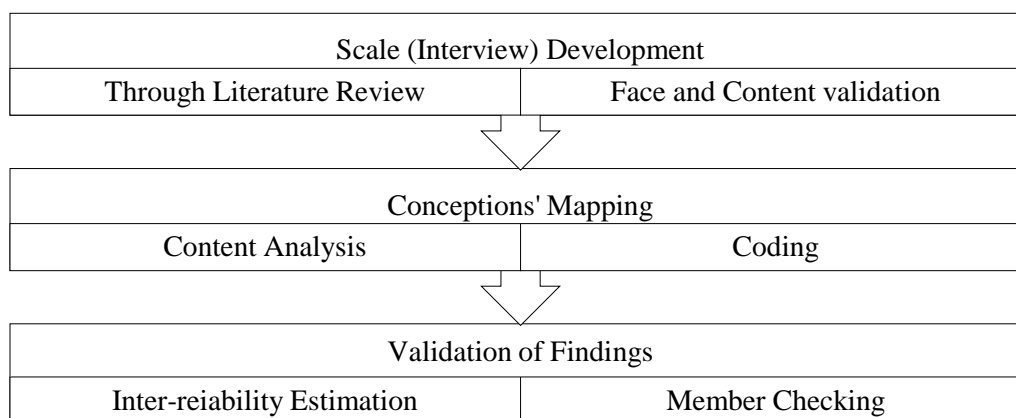


Figure 1: Procedure of the Study

Consistent with Marton (1986, p.42) and Tight (2015, p.40), a semi-structured interview was selected as the primary data collection tool. Aimed at steering research direction, determining the content and

the scope of the interview, and employment of the possible existing tools and instruments, a thorough literature review was run, and some central themes and concepts were extracted and used as the framework of the interview. The face and content validity of the questionnaire was examined by two experts. According to Marton (1986, p.41), using interview helps collect peoples' preconceived conceptions and thought in a controlled and structured condition, while the open-endedness of the questions secures the richness and multi-dimensionality of the information collected. In the next stage, the interviews were transcribed and emerging themes and concepts were collected. Rezaei et al., (2021, p.7) underline that identification of the conceptions should be carried out with absolute faithfulness to respondents' actual wording. The Selection of concepts was implemented based on some criterion. The first criterion was "relevance", meaning that a conception must be related to the question posed in the interview. Svensson (1997, p.163) maintains that the "relevance" of a conception is examined base on the meaning it carries and the context in which it occurred. Also, the tripartite criteria of "frequency", "position", and "pregnancy" suggested by Sjostrom and Dahlgren (2002) were followed in selecting concepts.

Accordingly, the density of the concepts in terms of the number of the occurrence (frequency), the association between the emerging concepts, established concepts and the subject of the study (position), and the weight given to each concept by the interviewees through verbal and non-verbal emphasis or re-iteration (pregnancy) were considered during concept mapping. When the "pool of concepts" was arrived at, the attention shifted from the interviewees to the meaning of concepts or "pool of meaning". In the final stage coding (open, axial, and selective coding) was run. Consistent with Marton (1986, p.43), verbal transcribed passages were sorted into piles, borderline cases were analyzed, and the criterion attributes for each category were identified to narrow down each passage into its respective category. Finally, an exemplar excerpt from the data that substantially represented the category was selected to accompany it. Guided by Rezaei et al., (2021, p.8), to extend the reliability of the finding, two researchers coded the collected data simultaneously. Disagreements in coding were reconciled through negotiation of views. Also, the credibility of the conclusions were elevated through a member-checking strategy and respondent validation, where the extracted factors (categories of concepts) were reviewed by a third analyst and turned back to five interviewees selected randomly out of the participants of the study to approve and assure their meaning. This strategy resulted in some subtle amendments in categories and concepts.

4. Results and Discussion

Concerns regarding shortcomings of traditional approaches to teaching and assessment gave rise to a critical shift in the tendency of language educational practitioners toward different modes of alternative assessment. However, research findings shows that the transition toward alternative assessment is impeded by many profound theoretical and practical problems (Rezaei et al., 2021, p.4). The same concerns are highlighted for implementing DA in a real educational context (Haywood & Lidz, 2007; Poehner, 2008; Fahmy, 2013, Farokhipour et al., 2020). To this aim, a phenomenographic study attempted to uncover the obstacles in the way of implementing DA in the real education context of Iran. The study revealed that technical constraints (including scoring issues, validity issues, discriminating power of the test, fairness issues, consequential credibility, and concept load), practical constraints (including demanding implementation, attitudinal issues, feasibility (practicality) issues, oral nature of feedback, and examiners' literacy), and contextual constraints (including conservative nature of Iranian context, and resource constraints) oppose implementation of DA in the Iranian academic setting.

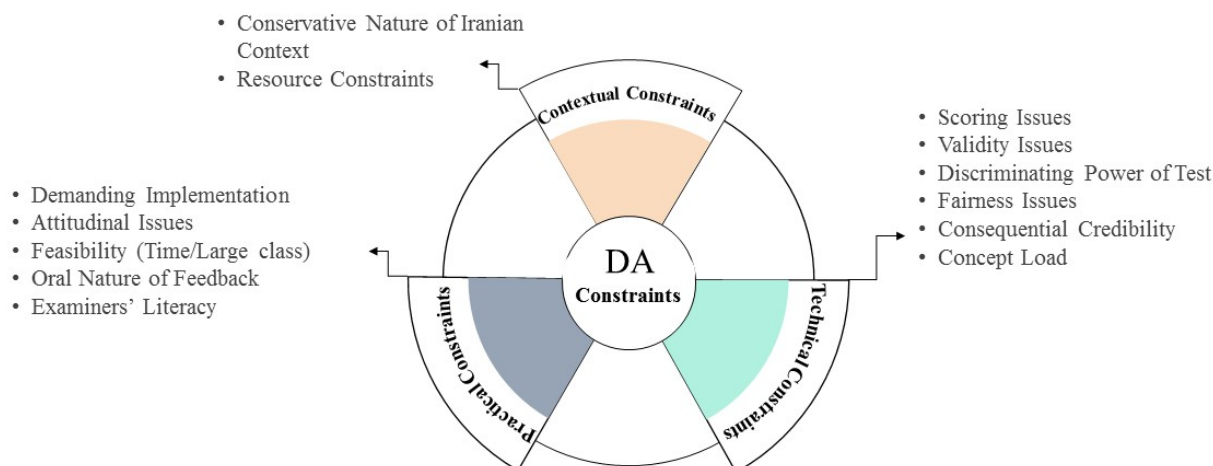


Figure 1. Fundamental Constraints to Implementation of DA in Iranian Academic Context

4.1. Technical Constraints

4.1.1. Scoring Issues

Collecting appropriate scores, which reveals the performances of a specific group of examinees on a specific set of items under a specific set of conditions is a crucial aspect of any assessment (Brown, 2013). Despite that, the findings of our study showed that Iranian language teachers believe that DA seems deficient in an identified procedure for exact and, or acceptable answer scoring based on the responsiveness moves of examinees. Therefore, language examiners fail to calculate accurate statistics for analyzing test's characteristics.

Excerpt (ELT8): *In DA we cannot estimate examiners' performance owing to difficulties in scoring the correct or acceptable answer based on their responsive moves. Consequently, we cannot calculate some metrics such as item facility (difficulty). Thus, simple statistics such as normal distribution cannot be estimated with due care.*

These findings are in line with Plonsky and Gass (2011), Plonsky (2013), and Brown et al., (2016), which emphasized on setting identified rubrics and straightforward procedures for scoring performance of learners because deficient scoring procedures and rubrics not only undermines critical test statistics but also violates tests' assumptions and applications. Similarly, participants of the present research reported that scoring in DA seems vague and inaccurate to them.

4.1.2. Validity Issues

Though some prominent scholars, such as Messick (1989), maintain that validity is an inseparable standard for any assessment, Moss (2003) believed that examining the validity for small-scale assessments, wherein DA is an example, is subject to serious doubt. Despite that, DA researchers and practitioners (e.g. Poehner, 2008, p.22) hold that the evidence for establishing validity of DA is distinct from those in traditional high stakes assessment. The findings of the current research showed that Iranian language teachers are more familiar with traditional psychometric procedures of test validation, and therefore raise serious doubts on the validity of DA.

Excerpt (ELT11): *Assessment in DA is somehow formative, subjective and developmental. This is different from the traditional assessment that, guided by a positivist inclination for reaching psychometric evidence, generates objective results which can be manipulated statistically. In addition to that, evidence from an individual dynamic assessment or group dynamic assessment barely yields enough data for establishing content and construct validity. Do not forget that the way that DA is carried out by language teachers, breaches face validity because it seems that the examiner and the examinees are interacting rather than doing a formal assessment.*

Research findings in this section are consistent with Moss (2003, p.14) and Hughes (2003) who maintain that in classroom assessments where some standards are not met, inferential conclusions that secure validity cannot be obtained. Yet, in line with Vygotsky (1987) and Gipps (1994), Poehner (2007 and 2008, p.9) strongly avows that the validity of DA can be established through some other types of evidence such as indexing examinees' development to higher levels. However, the participants of the current study found it a formidable task to collect such evidence through non-psychometric measures.

4.1.3. Discriminating Power of the Test

Feuerstein, Falik, and Feuerstein (2003) reported that static assessment could not predict examinees' abilities and assess their cognition. Besides, Haywood (1992) challenged static assessment due to their detachment from classrooms' requirements. Lidz (1991) that revealed that information derived from static testing is deficient in estimating learners' potential and learning difficulties. Despite the aforesaid merits, DA, which contrasts sharply with traditional static assessment, was reported inadequate in its discriminating power. Findings also showed that in DA the relationship between dependent performance of the examinee and his/her achievement is blurred (Jespen & Lidz, 2000).

Excerpts (ELT1): *When I intervene in the flawed performance of learners with higher ability, for example, in a grammatical error, they respond correctly in the third move of the inventory. On many occasions in my experimental studies, when learners with much lower ability commits the same error and receives mediation, they also respond in the same move and even earlier. Thus, when DA promotes able/unable learner equally, it cannot be used for many educational purposes; for instance it is not a good indication of achievement.*

These findings endorse Budoff (1987), Lidz (1991, p.12), and Poehner (2005), which have postulated that assessing learning potential and cognitive modifiability is different from measuring static Knowledge. However, the discriminating power of DA is an intact area of research in the literature and needs further investigation.

4.1.4. Fairness Issues

Fairness is an extensive and multi-faceted concept in language assessment. While some scholars such as Messick (1989) subsume fairness under the umbrella term of validity, Moss et al., (2008) and Tierney (2013) examine fairness as a distinct concept related to testing ethics and equity. Accordingly, such an exposition of fairness as an equal and just treatment is associated with those theoretical perspectives that view tests as contests among individuals rather than opportunities to measure psychological constructs (Thissen, 2016). According to our respondents, this later view of fairness, contrasts sharply with DA implementation in several ways.

Excerpt (ELT15): *Due to the demands of our education context, we have to use tests as contesting and grading instruments among our students for some selection and progress check purposes. However, in dynamic assessment, we are faced with interaction fluidity, meaning that the level of examinees' abilities determines the way mediation is tuned by the examiner. Also, even the grading adaption is influenced by examinees' capabilities. In other words, uniform treatment of more able and less able students violates the equity and justness principles.*

These findings submit proof to Broer, Rizavi, and Powers (2005) and DeLuca, Lapointe, and Luhanga (2016), which maintained that prioritizing testing standards in favor of one group violates fairness and instills measurement bias into the test.

4.1.5. Consequential Credibility

Menken (2008) reports that test developers in recent decades stressed promoting testing standards, characteristics, and psychometric features while disregarding their implicit consequences. Young (2012) adds that a good test must be consistent with contextual and social expectations, and good testing standards do not guarantee the quality of test application and use. Research findings in this section showed that DA has little or no resemblance to real-word practice and requirements on many occasion.

Excerpt (ELT2): *Since DA practitioners implement it in an assisted manner, it does not engender enough power in language learners to fulfill their educational, social, and economic expectations. For example, my experience shows that a translation student that gets used to take a translation test in an assisted manner finds it a difficult task to perform an oral translation test for employment or real educational purposes; let us save the educational values of the DA.*

Many studies support such a conclusion drawn from DA. McNamara and Roever (2006), Shohamy (2008), and McNamara (2009) believe that a viable educational testing instrument or a robust psychometric one does not necessarily undertake consequential credibility. Therefore, social and consequential expectations are also concerned.

4.1.6. Concept Load

According to Farokhipour (2019), DA has many concepts that are theoretically loaded and their interpretations for educational and clinical use is not an easy task. A review of the literature shows that some DA concepts, such as the “zone of proximal development” (Valsiner, 2001; Poehner & Lantolf, 2005), “constructing the future” (Tharp & Gallimore, 1990; Lidz & Gindis, 2003; Poehner, 2005, p.16), “transfer of learning” or “transcendence” (Campione and Brown, 1987; Lantolf & Thorne; 2006; Anton, 2009), and graduated prompting (Tzuriel, 2000; Fraser, 2006) are theoretically complex, and their interpretation into practice is difficult. The findings of our study showed that Iranian language teachers, despite their familiarity with these theoretical conceptions, do not have a clear guideline to put them into effect during DA.

Excerpt (ELT17): *in DA procedures, the mediational moves are determined in advance because we adopt and, or adapt an existing mediation inventory. But, for example, when the border between mistake and error is not evident during the mediation, the current level of ability vs. potential level of ability cannot be determined as it is depicted by the ZPD theory. Also, we cannot become assured of internalization or transcendence of learning based on the sole responsiveness moves of examinees. Applied definitions of these concepts are not as straightforward as their theoretical exposition.*

Findings in this section support Selinker (1992) and Thouesny (2010), which stated that, though zone of proximal development and mediation are adequate theoretical expositions, their translation into applied educational or experimental concepts needs further endeavor and creativity. Accordingly, further research is required to shed more light on the accurate and fluent application of DA in real educational contexts.

4.2. Practical Constraints

4.2.1. Demanding Implementation

Malone (2013) asserts that an adequately implemented language test is a rich source of information about different dimensions of the instruction for various stakeholders. Davies (2013) believes that assessment implementation and adopting appropriate strategies for devising language tests is a critical Issue. While Poehner (2005) postulates that implementing DA is a rigorous activity due to the required consideration for cognitive, affective, and psychomotor activities, Tzuriel (2001, pp.55-56) posits that integrating different assessment modules into a coherent process is a difficult task for test developers.

Excerpt (ELT16): *When we want to develop a DA-based test, at the same time, many issues need careful consideration, such as selection of appropriate mediation strategy, mastery of mediational moves, and their presentation in due time and occasion, diagnosing the quality of cognitive function, and at the same time deciding on providing assistance or ignoring it, maintaining the relationship between instructional content and test task content, and at the same time evaluating and tracing change inside learners' mind. Also, many unforeseen issues might arise that make it more difficult for a language teacher to implement DA in the real classroom.*

Elliot (2000) states that a radical difference between DA and static assessment is the dynamism of the procedure of its implementation because DA is too fluid and unpredictable while static assessment has a prefabricated and predictable implementation and administration guideline. These

findings support Lidz (2003, p.115), which holds that during DA implementation, an examiner must consider the underlying process, pay attention to cognitive deficiencies, select the type and extent of the mediation required for treating malfunction, and stay an active agent and attentive to non-intellective factors.

4.2.2. Attitudinal Issues

Dornyei (1994), Ryan and Deci (2000), and Liu (2007) noted that a positive attitude toward language learning is a critical success variable. Babamoradi, Nasiri and Mohammadi (2008) reported that Iranian language learners hold a positive attitude toward DA, a finding reiterated by Taheri and Dastjerdi (2016). However, the results of the current phenomenography showed that Iranian language teachers find attitudinal issues an impediment to DA practice in the country.

Excerpt (ELT13): *my two decades of experience show that language teachers, parents and other influential stakeholders, such as language department deans and university presidents, adopt a negative attitude toward DA because they think that the possibility of cheating and violation of the rights of students during dynamic testing is increased. The way the assessment is held contrasts radically with traditional formal assessment and therefore induces a sense of informality and looseness, which in turn generates negative attitudes in different stakeholders.*

The research on teachers' and other stakeholders' attitudes toward DA is scarce. Further research is required to delve into this intact area.

4.2.3. Feasibility (Practicality)

The feasibility (practicality) of language tests is a fundamental principle (Alderson & Wall, 1993; Fulcher, 2010). Current research findings revealed that Iranian language teachers raise sever doubts about the practicality of DA in the real educational context of Iran.

Excerpt (ELT19): *DA needs continuous formative assessment, which is naturally time-consuming and slows the speed of instruction. In the Iranian academic context, we come up against a large class size, resource constraints due to economic sanctions, lack of equipment for covering the four language skills, instructional activity overload, and the impossibility of online dynamic assessment due to lack of personal computers and internet failure. These are some of the limitations which render DA into a time-consuming and unwieldy educational activity.*

These findings are in line with Farokhipour et al., (2020, p.120), which reported that, although DA can be implemented in the real educational settings through integrating perspectives from different areas of the discipline, it needs extra time and energy.

4.2.4. Oral Nature of Feedback

Feedback is a crucial element of language teaching and a significant source of input for internalization and modification of learning (Kim, 2004; Ellis, 2009). Since DA is built upon mediation, the weight attached to feedback is multiplied. Despite this weight, the oral nature of the mediation (feedback) brings about some serious problems for DA examiner.

Excerpt (ELT3): *When a DA examiner gives feedback to the examinee, due to the oral nature of the intervention, communication breakdown may happen. Also, qualities of an oral speech, such as rate and pronunciation, might influence the interaction. Furthermore, verbal interaction qualities might affect examinees' performance on a non-oral test task such as writing. Also, my experience shows that dyadic interaction between teacher and student, despite the provision of educational opportunities, frightens students and excites anxiety in them. These are validity and fairness issues.*

The findings in this section echoed Fulcher and Reiter (2003), Berry (2007) and Son (2016), which maintained that dyadic oral assessment entails some characteristics that threaten test standards (e.g., validity, equity, and fairness) and test objectives.

4.2.5. Examiners' (assessment) Literacy

The development and administration of a standard assessment require sophisticated competencies (Boyles, 2005), which are collectively called language assessment literacy (Inbar-Lourie, 2008). Many researches are conducted to shed light on different dimensions of assessment literacy (e.g., Combee et al., 2020, Malone, 2013; Fulcher, 2012; Davies, 2008). Gilardo (2018) reports that knowledge about tenets of applied linguistics, knowledge about assessment theories and procedures, and knowledge about assessment context are three significant dimensions of assessment literacy. Current research findings showed that the standard implementation of DA requires extra expertise, knowledge and skill because it is rooted in the psychological and clinical foundations that are usually at odd with conventional educational practices.

Excerpt (ELT5): *Many Iranian language teachers at the academic level are familiar with the theories of DA, its formats, and its modes. But it seems that, due to the lack of pre-service and in-service instructional courses, many of these teachers avoid not only DA but also other types of alternative assessment. They still prefer traditional summative tests. The development, validation, rating procedure, teacher talk strategies for mediation, scoring procedure, standardization, and many other aspects of DA need sophisticated knowledge and expertise, which require time, energy, and institutional support to master. Thus, assessment literacy and awareness of Iranian language teachers about new modes of alternative assessment, including DA, even at the academic level, need to be promoted to a great extent. Then, we can expect detachment from traditional approaches.*

The findings of our study affirm Shahzamani and Tahririan (2021), and Homayounzadeh and Razmjo (2021), which investigated different dimensions of Iranian language teachers' assessment literacy and concluded that mastery of some knowledge and skill are necessary for implementing a standard assessment that informs instruction and educational decision making. These findings uncovered some serious demerits associated with teachers' assessment literacy in the country which needs particular attention at the scholarly and policy levels.

4.3. Contextual Constraints

4.3.1. Conservative Nature of Iranian Context

The impact of context, including cultural, ethnic, and religious variables, on language learning has been the subject of considerable research in different countries. For instance, Berhane (2019) and Eshete (2013) reported that the nature of language context might impose sever limitations on language education owing to cultural, religious, and social reasons. In Iran, too, Behtash, Farokhipour, and Hashemi (2017) reported that religious identity, for example, negatively impacts motivation and inclination for learning English in the country. Our findings revealed that managerial conservatism, cultural conservatism and a tendency toward nativization of English in the country has impacted DA implementation, although indirectly.

Excerpt (ELT1): *many Iranian educational policymakers and managers are unaware of the shortcoming of traditional views of education and assessment that are an inaccurate translation of outdated western views, and resist any radical change. Also, language teachers, despite their awareness of pitfalls inherent in traditional assessment, are not motivated for alternative assessment because of contextual pressures, institutional conservatism, stakeholders' extreme care for aggravation of negative washback, and raised concern for nativization of language education, which has impeded almost all innovational attempts at the national level.*

Findings in this section are consistent with Farokhipour et al., (2021) and Salimi and Safarzadeh (2018), which concluded that context is a determinant factor in language learning because teaching and assessment methods, research methods, and management methods are sensitive to context. The theorizers of DA (e.g., Lantolf, 1994; Lantolf & Appel, 1994) also believe that socio-cultural atmosphere affects various dimensions of language learning.

5. Conclusion

The current research was an unorthodox attempt to employ ethnography and probe primary constraints to the implementation of DA in the Iranian context, despite the assessment's rich theoretical, practical, and empirical bases. Results of holistic interviews, their analysis, and their validation revealed three categories of constraints, including technical, practical, and contextual constraints. Also, findings showed that the technical problems have rendered DA into an intricate practice that requires sophisticated knowledge, skill, and experience. Furthermore, findings showed that the elimination of practical and contextual constraints, in addition to assessment literacy, need managerial and institutional attention at the policy level. These findings have implications for language researchers, teachers, and practitioners. It is recommended that language researchers emulate the phenomenographic procedure used in this study because this qualitative method captures a holistic picture of similar educational phenomena. Also, the innovative validation procedure used can be emulated and extended by future researchers. Findings also shed light on different dimensions of limitations and constraints to not only DA but also other alternative assessment approaches. This study had some limitations. Participants were sampled purposively and conveniently from three universities. Future researchers are recommended to include more participants from more universities through more random sampling procedures, secures the generalization of findings and promote their validity and credibility. Another limitation resided in the feasibility of the research. Embracing many participants, running lengthy interviews and analyzing them was time and money consuming.

6. Acknowledgement

We would like to express our deepest thanks to the language professors in department of English at Malayer branch, Islamic Azad University, and all those language teachers and professors that took part in our interviews.

7. Declaration of Conflicting Interests

The authors, hereby, declare that they have no competing interests.

8. Funding Details

No funding was received from any institution or academic center.

References

- Akerlind, G. (2005). Variation and commonality in phenomenographic research methods. *Higher Education Research and Development*, 24, 321–334.
- Alavi, S. M., Kaivanpanah, Sh., & Shabani, K. (2012). Group dynamic assessment: An inventory of mediational strategies for teaching listening. *The Journal of Teaching Language Skills*, 3(4), 27-58.
- Alderson, J.R., & D. Wall. (1993). Does washback exist? *Applied Linguistics* 14, 115-29.
- Anton, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42, 576-598.
- Babamoradi, P., Nasiri, M., Mohammadi, E. (2018). Learners' attitudes toward using dynamic assessment in teaching and assessing IELTS writing task one. *International Journal of Language Testing*, 8(1), 1-11.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Behtash, I. Z., Hashemi, S. H., Farokhipour. (2017). Influence of religious identity on a foreign language learning: A case of Iranian EFL learners. *International Journal of English Language & Translation Studies*, 5(2), 16–20.
- Berhane G. (2019). Foreign Languages in Ethiopia: History and status: *International Journal of Research and Analytical Reviews*, 6(1) 1431-1439.

- Birjandi, P., Estaji, M., & Deyhim, T. (2013). The impact of dynamic assessment on reading comprehension and metacognitive awareness of reading strategy use in Iranian high school learners. *Iranian Journal of Language Testing*, 3(2), 61-77.
- Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *National assessment summit papers* (pp. 11-15). Iowa State University.
- Broer, M., Lee, Y. W., Rizavi, S., & Powers, D. (2005). *Ensuring the fairness of GRE writing prompts: Assessing differential difficulty*. ETS Research Report, RR 05-11. Princeton, NJ: Educational Testing Service.
- Brown, J. D. (2013). My twenty-five years of cloze testing research: So what? *International Journal of Language Studies*, 7(1), 1-32.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653. <https://doi.org/10.2307/3587999>.
- Brown, J. D., Trace, J., Janssen, G., & Kozhevnikova, L. (2016). How well do cloze items work and why? In C. Gitsaki & C. Coombe (Eds.), *Current trends in language evaluation, assessment and testing: Research perspectives* (pp. 2-39). Newcastle Upon Tyne, UK: Cambridge Scholars.
- Brown, K. W., West, A. M., Loverich, T. M., & Biegel, G. M. (2011). Assessing adolescent mindfulness: Validation of an adapted mindful attention awareness scale in adolescent normative and psychiatric populations. *Psychological Assessment*, 23(4), 1023-1033. <https://doi.org/10.1037/a0021338>.
- Budoff, M. (1987). The validity of learning potential assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 52-81). New York: Guilford Press.
- Campione, J. C., & Brown, A. L. (1987). Linking dynamic testing with school achievement. In C. S. Lidz (Ed.), *Dynamic testing* (pp. 82-115). New York: Guilford Press.
- Carlson, J. S., & Wiedl, K. H. (1992). The dynamic assessment of intelligence. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 167-186). New York: Springer.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(3), 1-16.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347.
- Davies, A. (2013). Fifty years of language assessment, *The Companion to Language Assessment*, 11(1) 1-18.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. In *Educational Assessment, Evaluation and Accountability*, (pp. 1-22). <https://doi.org/10.1007/s11092-015-9233-6>.
- Dornyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal*, 78(3), 273-284.
- Dornyei, Z., & Taguchi, T. (2010). *Questionnaires in Second Language Research Construction, Administration and Processing* (2nd Ed.). New York, Routledge.
- Elliott, J. (1993). Assisted Assessment: If it is dynamic why it is so rarely employed? *Educational and Child Psychology*, 10, 48-58.
- Elliot, J.G. (2000). Dynamic Assessment in educational contexts: Purpose and Promise. In C.S. Lidz, & J.G. Elliot (Eds.), *Dynamic assessment: Prevailing models and applications*. (pp.713-740). New York, USA: Elsevier Science Inc.
- Elliot, J.G. (2003). Dynamic assessment in educational settings: Realizing potential. *Educational Review*, 55(1), 15-32. <https://doi.org/10.1080/00131910303253>.

- Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, 1(1), 3-18.
- Eshetie, B. A. (2013). *Language Policies and the Role of English in Ethiopia*. A presentation paper at the 23rd Annual Conference of IATEFL BESIG, Bielefeld, Germany.
- Fahmy, M.M (2013). *The effect of dynamic assessment on adult learners of Arabic: A mixed-method study at the defense language institute foreign language center*. Unpublished PhD dissertation, University of San Francisco.
- Farrokh, P., & Rahmani, A. (2017). Dynamic assessment of writing ability in transcendence tasks based on Vygotskian perspective. *Asian-Pacific Journal of Second and Foreign Language Education*, 2(1), 1-23. <https://doi.org/10.1186/s40862-017-0033-z>.
- Farokhipour, S., Khoshshima, H., Sarani, A., Ganji, M. (2020). Presenting and investigating the effect of a local model of dynamic assessment in diagnosing and removing learning difficulties of high school students in productive skills. *Foreign Language Research Journal*, 10(1), 120-134.
- Farokhipour, S., Mostafapoor, E., Shibak, M., Khoshnoodi, H. (2021). Departure from the West in foreign language teaching strategies for consolidation of the ideological system in the post-method. *Scientific Journal of Islamic Education*, 28(49), 121-139.
- Feuerstein, R., Feuerstein, R. S., Falik, L. H., & Rand, Y. (2002). *The dynamic assessment of cognitive modifiability: The learning propensity assessment device, theory, instruments, and techniques*. Jerusalem: ICELP Press.
- Fulcher, G. (2010). *Practical language testing*. Hodder Education.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132.
- Fulcher, G., & Reiter, R. M. (2003). Task difficulty in speaking tests. *Language Testing*, 20(3), 321-344.
- García, G. E., & Pearson, P. D. (1994). Assessment and diversity. In L. Darling- Hammond (Ed.), *Review of research in education* (pp. 337-392). Washington, DC: American Educational Research Association.
- Gergen, K. J., & Dixon-Román, E. J. (2014). Social epistemology and the pragmatics of assessment: Teachers college record: *The Voice of Scholarship in Education*, 116(11), 1–22. <https://doi.org/10.1177/016146811411601111>.
- Ghahderijani, B. H., Namaziandost, E., Tavakoli, M., Kumar, T., & Magizov, R. (2021). The comparative effect of group dynamic assessment (GDA) and computerized dynamic assessment (C-DA) on Iranian upper-intermediate EFL learners' speaking complexity, accuracy, and fluency (CAF). *Language Testing in Asia*, 11(1), 1-20. <https://doi.org/10.1186/s40468-021-00144-3>
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195.
- Haywood, H. C. (1992). Interactive assessment: A special issue. *Journal of Special Education*, 26(3), 233-234.
- Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational application*. Cambridge University Press.
- Haywood, H. C., & Tzurriel, D. (2002). Applications and challenges in dynamic assessment. *Peabody Journal of Education*, 77 (2), 38–61.

- Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Association for supervision and curriculum development (ASCD).
- Homayounzadeh, Z., & Razmjoo, S. A. (2021). Examining 'assessment literacy in practice' in an Iranian context: Does it differ for instructors and learners? *Journal of Teaching Language Skills*, 40(2), 1–45.
- Huerta-Macías, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5(1), 8–11.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 328–402
- Jepsen, R. H., & Lidz, C. S. (2000). Group dynamic assessment: Reliability and validity of a cognitive assessment procedure with adolescents with developmental disabilities. *Journal of Cognitive Education and Psychology*, 1(1), 26–39.
- Lantolf, J. P. (1994). Sociocultural theory and second language learning: introduction to the special issue. *Modern Language Journal*, 78, 418–420.
- Lantolf, J. P., & Appel, G. (1994). *Vygotskian approaches to second language research*. Norwood, NJ: Ablex.
- Lantolf, J.P. & S.L. Thorne. (2006). *The socio-genesis of second language development*. Oxford: Oxford University Press.
- Lidz, C.S. (1991). *Practitioner's guide to dynamic assessment*. New York: Guilford Press.
- Lidz, C.S. (2002). Mediated learning experience (MLE) as a basis for an alternative approach to assessment. *School Psychology International*, 23(1), 68 – 84.
- Lidz, C.S. (2003). *Early childhood assessment*. New Jersey: John Wiley & Sons Inc.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33–48). Lawrence Erlbaum Associates, Inc.
- Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller (Eds.), *Vygotsky educational theory in cultural context* (pp. 99-119). Cambridge University Press.
- Liu, M. (2007). Chinese students' motivation to learn English at the tertiary level. *Asian EFL Journal*, 9(1), 126-146.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344.
- Marton, F. (1981). Phenomenography? Describing conceptions of the world around us. *Instructional Science*, 10, 177–200.
- Marton, F. (1986). Phenomenography: A research approach to investigating different understandings of reality. *Journal of Thought*, 21, 28–49.
- Marton, F. (1988). Phenomenography: Exploring different conceptions of reality. In D. Fetterman (Ed.), *Qualitative approaches to evaluation in education: The silent scientific revolution* (pp. 176–205). New York, NY: Praeger.
- Marton, F., & Pong, W. Y. (2005). On the unit of description in phenomenography. *Higher Education Research & Development*, 24, 335-348.
- McNamara, T. F. (2009). Australia: The dictation test Redux? *Language Assessment Quarterly*, 6, 106–111.

- McNamara, T. F., & Roever, C. (2006). *Language testing: The social dimension*. Oxford, UK: Blackwell.
- Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Bristol: Multilingual Matters.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). New York, NY: American Council on Education/Macmillan
- Mitchell, R. (1992). *Testing for learning. How new approaches to learning can improve American schools*. The Free Press.
- Moss, P.A. (2003). Re-conceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13-25.
- Moss, P. A., Pullin, D. C, Gee, J. P., Haertel, E. H., & Young, L. J. (Eds.). (2008). *Assessment, equity, and opportunity to learn* (pp.333-352). Cambridge University Press.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Sage.
- Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, 35(4), 655–687.
- Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, 61(2), 325–366.
- Poehner, M. E. (2005). *Dynamic assessment of oral proficiency among advanced L2 learners of French*. The Pennsylvania State University, United States: Pennsylvania.
- Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal*, 91(3), 323–340. doi.org/10.1111/j.1540-4781.2007.00583.x.
- Poehner, M.E. (2008). *Dynamic assessment: a Vygotskian approach to understanding and promoting second language development*. Springer Publishing.
- Poehner, M. E. (2011). Validity and interaction in the ZPD: interpreting learner development through L2 Dynamic Assessment. *International Journal of Applied Linguistics*, 21(2), 244–263. <https://doi.org/10.1111/j.1473-4192.2010.00277.x>.
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3), 233-265.
- Rezai, A., Alibakhshi, G., Farokhipour, S., & Miri, M. (2021). A phenomenographic study on language assessment literacy: hearing from Iranian university teachers. *Language Testing in Asia*, 11(1). <https://doi.org/10.1186/s40468-021-00142-5>.
- Ryan, R. M. & Deci, E. L. (2000). *Intrinsic and extrinsic motivations: Classic definitions and new directions*. Retrieved June, 21, 2020 from: <http://www.idealibrary.com>
- Salimi E, Safarzadeh M M. (2018). From globalization to glocalization: An investigation of EFL teachers' perceptions of glocalized language education. *Language related research*, 9 (4):157-180
- Shahzamani, M., & Tahririan, M. H. (2021). Iranian medical ESP practitioners' reading comprehension assessment literacy: Perceptions and practices. *Iranian Journal of English for Academic Purposes*, 10(1), 1–15.
- Shohamy, E. (2008). Language policy and language assessment: The relationship. *Current Issues in Language Planning*, 9(3), 363–373.
- Sjostrom, B., & Dahlgren, L. O. (2002). Nursing theory and concept development or analysis: Applying phenomenography in nursing research. *Journal of Advanced Nursing*, 40(3), 339–345

- Sternberg, R.J. and E.L. Grigorenko. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge University Press.
- Svensson, L. (1997). Theoretical foundations of phenomenography. *Higher Education Research & Development*, 16(2), 159–171. <https://doi.org/10.1080/0729436970160204>.
- Taheri, R. & Dastjerdi, H. (2016). Impact of dynamic assessment on Iranian EFL learners' picture-cued writing. *International Journal of Foreign Language Teaching and Research*, 4(13), 129-144.
- Tharp, R.G. & Gallimore, R. (1990). Teaching mind in society. In L.C. Moll. (Ed). *Vygotsky and education: Instructional implications and applications of socio-historical psychology* (pp.175-205). Cambridge: Cambridge University Press.
- Thissen, D. (2016). Commentary on the assessment of the fairness of comparisons under divergent measurement conditions. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp.203-214). Routledge.
- Thouësny, S. (2010). *Assessing second language learners' written texts: An interventionist and interactionist approach to dynamic assessment*. Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA), Toronto, Canada.
- Tight, M. (2015). Phenomenography: the development and application of an innovative research design in higher education research. *International Journal of Social Research Methodology*, 19(3), 319–338.
- Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (ed.), *The Sage handbook of research on classroom assessment* (pp. 125-144). Sage Publications.
- Tzuriel, D. (2000). Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology Review*, 12 (4), 385-420.
- Tzuriel, D. (2002). Cognitive education: The menace and hope. In W. Resing, W. Ruijssenaars, & D. Aalsvoort (Eds.), *Learning potential assessment and cognitive training: Actual research perspectives in theory building and methodology* (pp. 355–363). New York. Elsevier.
- Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Human Development*, 44, 84–97.
- Vygotsky, L.S. (1978). *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Young, R. F. (2012). Social dimensions of language testing. In G. Fulcher & F. Davidson. (Eds.), *The Routledge handbook of language testing* (pp. 178-193). Routledge, New York.
- Zarinkamar, N., Abdi, R., Davaribina, M. (2021). Developing EFL Learners' Creativity in Writing Skill Through Dynamic Assessment-based Teaching. *Research in English Language Pedagogy*. 9(2), 309-338.