

Review

Not peer-reviewed version

---

# Dynamic Assessment in L2 Settings: A Systematic Review and a Meta-Analysis

---

[Mohammed Aljohani](#)\*

Posted Date: 10 October 2025

doi: 10.20944/preprints202510.0770.v1

Keywords: Dynamic Assessment; Second/Foreign Language Learning; Assessment; Sociocultural Theory; Vygotsky



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Dynamic Assessment in L2 Settings: A Systematic Review and a Meta-Analysis

Mohammed Aljohani

Yanbu English Language and Preparatory Year Institute, Royal Commission for Jubail and Yanbu, Yanbu, Saudi Arabia; mohammad.aljohani@gmail.com

## Abstract

Dynamic assessment (DA) is an approach that intentionally brings instruction and learning together through adopting a social-constructivist framework in order to assess, diagnose, and improve L2 linguistic skills. DA in the L2 classroom has been applied but not surveyed for its effect and plausibility. To address that, I did a systematic review and a meta-analysis to report the effectiveness of DA in the L2 classroom in developing students' linguistic skills. Nine databases were searched to ensure a comprehensive coverage of the field including published dissertations. The characteristics of the studies were coded and analyzed systematically following a developed coding sheet. Means and standard deviations were coded, and the standardized mean difference was used to calculate effect sizes. A random effects model was used to compute the overall effect size of the included studies in the meta-analysis. Results indicated a positive effect of DA on the development of foreign language skills. The magnitude of the effect was significantly large ( $g = 1.07$ ; 95% CI = [0.82, 1.32]). The prediction intervals suggested that future research that meets inclusion criteria will report 95% of the time a positive effect (95% PI = [0.03, 2.10]). Sample size, age, and targeted linguistic skill did not significantly change the existing effect of DA in the L2 classroom. Based on the systematic review and existing literature, a framework was developed to help pedagogists include DA in their L2 classroom practices by explicating its purposes and functions, and the required conditions for a successful implementation.

**Keywords:** dynamic assessment; second/foreign language learning; assessment; sociocultural theory; Vygotsky

---

## Dynamic Assessment in L2 Settings: A Systematic Review and a Meta-Analysis

Assessment in the classroom context is a delicate exploration of growing minds influenced by interpersonal and intrapersonal factors with a set of priori assumptions regarding the subject-matter's nature and process of its acquisition. McMillan (2013, p.4) described classroom assessment as:

“... a process that teachers and students use in collecting, evaluating, and using evidence of student learning for a variety of purposes, including diagnosing student strengths and weaknesses, monitoring student progress toward meeting desired levels of proficiency, assigning grades, and providing feedback to parents.”

Shepard (2000) described the role of assessment in teaching and learning and predicted that an elaboration of a social-constructivist framework will be “... the middle-ground theory that will be accepted as common wisdom and carried into practice” (p. 6). She suggested that dynamic assessment is a very helpful “interactive assessment” that could move assessment into the ongoing teaching instead of being an “end-point of instruction” (p. 10).

One of the most celebrated types of assessment in the classroom is formative assessment. Yet, formative assessment is not clearly defined (Bennet, 2011), which needs to “... transform from a collection of abstract theories and research methodologies and become a creative and systematic classroom practice” (Clark, 2012, p. 24). On the other hand, a less widely implemented type of assessment, dynamic assessment, draws from the work of the psychologist Lev Vygotsky and his well-known zone of proximal development (ZPD) (Elliot, 2003). Dynamic assessment (DA) is an

intentional attempt to bring assessment and instruction together (van der Veen, Dobber, & van Oers, 2016) which aims to modify the cognitive situation through teaching (Tzurriel, 2001). Elliot (2003) encouraged the production of DA research that produces applicable practices for teachers, amongst others, but it seems that such call is not answered (Elliot, Resing, & Bechmann, 2018). In the world of second/foreign language teaching, the scene seems to be a bit different.

Kozulin and Garb's (2002) study was among the first to apply dynamic assessment (DA) in second language (L2) learning, influencing future research. Using a pretest-intervention-posttest design, they aimed to improve English text comprehension by mediating cognitive strategies (Feuerstein, 1979; 1990). The study involved 23 low-performing EFL students. After mediation, students showed significant improvement with an effect size of 1.2. They also developed a learning potential score (LPS) formula, now widely used to classify students based on future learning potential. Mediation was sandwiched between tests, illustrating DA's feasibility in L2 classrooms.

Sternberg and Grigorenko (2002, pp. 28-29) distinguished between static and dynamic testing. Importantly, *dynamic testing* explores "whether and how the participant will change if an opportunity is provided", while *dynamic assessment* is an intentional attempt to "intervene and change" (Sternberg & Grigorenko, 2002, p. 30). Poehner (2008) argues that such distinction is unnecessary as change can occur even in a single session of mediation (p. 18).

Another distinction adopted by many of the reviewed L2 DA papers is Lantolf and Poehner's (2004) distinction of two types of mediation: interventionist and interactionist. Interventionist DA follows a standardized format in which mediation is predetermined to help the learner(s) reach a predetermined reach point; on the other hand, interactionist DA offers mediation as it emerges between the learner(s) and the mediator in which an end point is not of any concern, nor the effort and time needed for development (p. 54).

The less frequent use of dynamic assessment in educational settings could be due to different reasons. One reason is of practical nature as most dynamic assessments are devised by psychologists which made them difficult to be applied in educational settings (van der Veen et. al., 2016). Another reason is that teachers need to apply forms of assessment to the whole group of students instead of an individual experience which is a property assumed for dynamic assessment (van der Veen et. al., 2016). Further, DA challenges the common practice in classrooms where testers view it as a teaching approach, and those interested in teaching disregard it and view it as an assessment tool not related to teaching (Poehner, 2008). Such gap between classroom practice and testing (or assessment) could require us to put dynamic assessment into a broader field of classroom assessment where it can be communicated very clearly to teachers that it can be part of their classroom practice and to create a research agenda that utilizes DA purposely to serve the classroom setting considering all possible obstacles. If a teacher reads a textbook in classroom assessment, the odds are high that they will not find any mention of DA.

Li and Li (2015) conducted the only literature review of L2 dynamic assessment (DA) over the past decade. They reviewed 25 papers published between 1994 and 2013, dividing them into interactionist and interventionist categories (Lantolf & Poehner, 2004). Most studies treated DA as a pedagogical tool, focusing on reading and grammar. The review noted a preference for group mediation and highlighted a lack of research using Kozulin and Garb's (2002) learning potential score (LPS) formula. They recommended further research on peer mediation and quantitative approaches for generalizability.

This systematic review and meta-analysis intend to quantify the effects of applying L2 DA practices in the classroom setting. It focuses on the use of DA to develop and assess L2 linguistic skills by detailing formats used, the types and nature of mediations offered, and the mediums used to deliver mediation. Further, the meta-analysis is used to quantify the effect of DA on the development of linguistic skills. The following research questions are attempted:

RQ1. How is dynamic assessment applied in the L2 classrooms and settings? What are the formats used in implementing DA, types and nature of mediation/intervention, and mediums used in delivering DA?

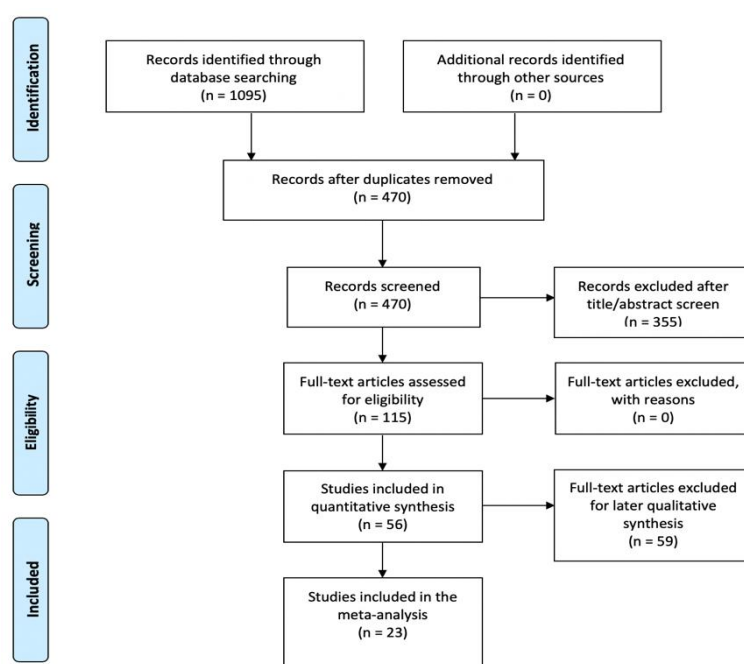
RQ2. What is the developmental effect of implementing DA on second/foreign language competencies for L2 learners?

## Methods

The systematic review aimed at identifying all L2 DA published peer-reviewed articles and dissertations. Nine databases were used to search for eligible studies within the period from 2009 to 2019. The databases searched were: Education Research Complete; ERIC; PsychInfo; Web of Science: Social Sciences Citation Index; Web of Science: Arts & Humanities Citation Index; ProQuest: Linguistics & Language Behavior Abstracts; ProQuest: Education Database; Science Direct; and ProQuest: Theses & Dissertations. The search terms used included the following combinations: (“dynamic assessment” OR “dynamic testing”) AND (second language learning OR L2 OR EFL OR ESL OR ELL), Zone of proximal development AND (assessment OR testing OR evaluation) AND second language learning OR L2 OR EFL OR ESL OR ELL, (“mediated learning” AND “assessment”) AND (second language learning OR L2 OR EFL OR ESL OR ELL).

### Study Selection

The initial number of retrieved studies for all searched databases was 1095 studies; after removing duplicates, 470 studies were identified. Then, studies were screened by title and abstract which resulted 115 studies eligible for full-text screening. No studies were excluded after reading the full papers. Fifty-six studies were of a quantitative nature or included quantitative analyses as in mixed methods designs. Forty-five studies were qualitative and were saved for later analyses. Out of the 56 studies, 23 studies satisfied the inclusion criteria for the meta-analysis (see Figure 1).



**Figure 1.** Studies Selection Process.

### Eligibility Criteria

Eligible studies were empirical, published in English or Arabic, and focused on using dynamic assessment (DA) to develop or assess a second/foreign language (L2) linguistic skill in L2 contexts (e.g., schools, universities). Participants were mainstream L2 learners, excluding students with disabilities, gifted students, or balanced bilinguals. The target learners were those in formal classrooms in countries where the native language is spoken (Gass & Selinker, 2013).

Studies included in the meta-analysis followed a pretest-intervention-posttest design, with performance measured before and after the intervention using conventional, nondynamic tests. A control group was required to avoid inflated effect sizes (Lipsey & Wilson, 1993). If comparison groups were present, only data from the DA group and the control group were used to calculate the effect size. Studies using different DA modalities were combined if they targeted the same linguistic skill (Borenstein et al., 2011). In cases where both overall linguistic scores and sub-scores were reported, only the overall score was used for effect size calculation. All studies contributed one effect size.

### *Coding*

I extracted studies' descriptives which included author, year of publication, country where the study took place, the context (e.g., school, university, private/public language institution), sample's size and age, sampling technique, and targeted L2. Further the goal(s), research questions, description of the intervention and mediation offered, the approach followed (interventionist or interactionist), the mediator (computerized vs. not; teacher to individual vs. teacher to group vs. peer to peer), and the targeted L2 language skills. Methodologically, description of the design (e.g., pretest – intervention – posttest), and control and comparison groups use. The means and standard deviations were extracted to compute the effect sizes. Quality of papers were considered for discussion but not used to exclude papers to avoid imposing any idiosyncrasies.

### *Meta-Analytic Techniques*

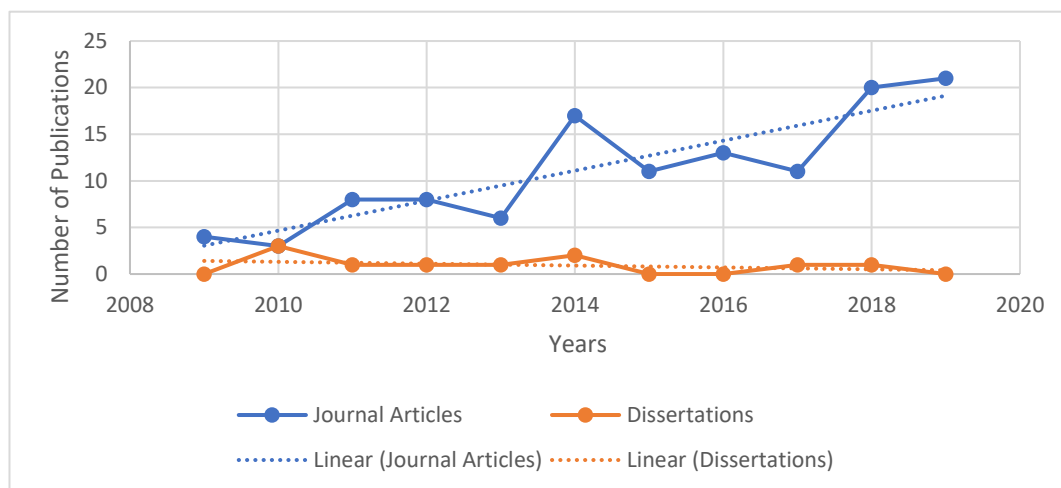
The random-effects model was used for the meta-analysis as in educational contexts it is not plausible to assume that all samples come from the same exact population (Lipsey & Wilson, 2001; Borenstein, Hedges, Higgins, & Rothstein, 2011). The estimator used was the Hartung-Knapp-Sidik-Jonkman method, and it was preferred because it is conservative and it “widens the confidence interval to reflect uncertainty in the estimation of between-study heterogeneity” (Cochrane Handbook for Systematic Reviews of Interventions version 6.0, 2019). Effect sizes were calculated based on the standardized mean difference. The calculations were computed using the ‘esc’ package in R (Lüdtke, 2018). The pooled deviation used the pre-test standard deviation for the treatment and control groups to control for possible threats to the validity of computed effect sizes due to sample size (Morris, 2008; Hoyt & Del Re, 2017). The meta-analysis and subgroup analyses were synthesized using the R package ‘meta’ (Schwarzer, Guido, Carpenter, & Rücker, 2015).

**Analysis Strategies.** After computing effect sizes for the included studies in the meta-analysis, a random-effects model was used to generate an overall mean effect size with the confidence and prediction intervals. I examined the within-study heterogeneity statistics of  $Q$ ,  $\tau^2$ , and  $I^2$  to reach more accurate interpretation of the overall effect size.  $I^2$  “describes the percentage of total variation across studies that is due to heterogeneity rather than chance” (Higgins, Thompson, Deeks, & Altman, 2003, p. 558) and is not “directly affected by the number of studies in the analysis” (Borenstein, et. al., 2009, p. 118). The larger the percentage of the  $I^2$ , the more is plausible to further investigate this variability through statistical techniques like subgroup analysis or meta-regression (Borenstein, et. al., 2009). As a rule of thumb, Higgins et. al. (2003) suggested that a 75% suggests high heterogeneity.

To assess heterogeneity, a graphical method, the Baujat plot, was used to pinpoint studies that were highly contributing to heterogeneity (Baujat, Mahé, Pignon, & Hill, 2002). Further, the computationally intensive graphical display of study heterogeneity (GOSH) was used as it performs “a meta-analysis on all possible subsets of the  $k$  studies in a meta-analysis” (Olkin, Dahabreh, & Trikalinos, 2012, p. 215). A subgroup analysis using the random-effect model was conducted to compare the intervention effects based on the variant outcomes of the meta-analysis included studies (Borenstein & Higgins, 2013).

### *Findings*

L2 Dynamic Assessment publications frequency has improved by year for journal articles, but it does not seem to be the case for dissertations (see Figure 2). This could suggest that the discussions of L2 DA in academia is relatively infrequent, and some serious attempts are needed to include dynamic assessment in university programs.



**Figure 2.** L2 DA number of publications from 2009 to 2019, per year.

Methodologically, L2 DA studies sample sizes increased compared to Li and Li (2015) whom they reported that only 32% of their included studies had sample sizes greater than 30. In this review, 77% of the reviewed papers had sample sizes above 30 in which 19% of them were above 100. Further, and similar to Li and Li's findings, the majority of the reviewed L2 DA studies had targeted reading and grammar in their interventions representing 55% of the total reviewed quantitative papers. The least L2 skill focused on was speaking. Ten studies (18%) used computers (cell phones in one study) to mediate, while the vast majority used teacher as mediator to the whole classroom (51%). Reviewed studies seem to use either an interventionist or an interactionist approach to DA roughly evenly, 47% and 43% respectively; while the rest of the studies tried to implement or contrast both approaches. Contextually, 44% of the studies were conducted in a university setting, 33% in a private/public language institution, 17% in a school, and 2% in a kindergarten; while two studies implemented DA in distance education situations: one used email correspondence (Ebadi, Vakilifard, & Bahramlou, 2018), and the other used mobile-based learning (Rashidi & Nejad, 2018).

Table 1 describes the major extracted studies' descriptive characteristics.

**Table 1.** Included Quantitative Studies' Characteristics.

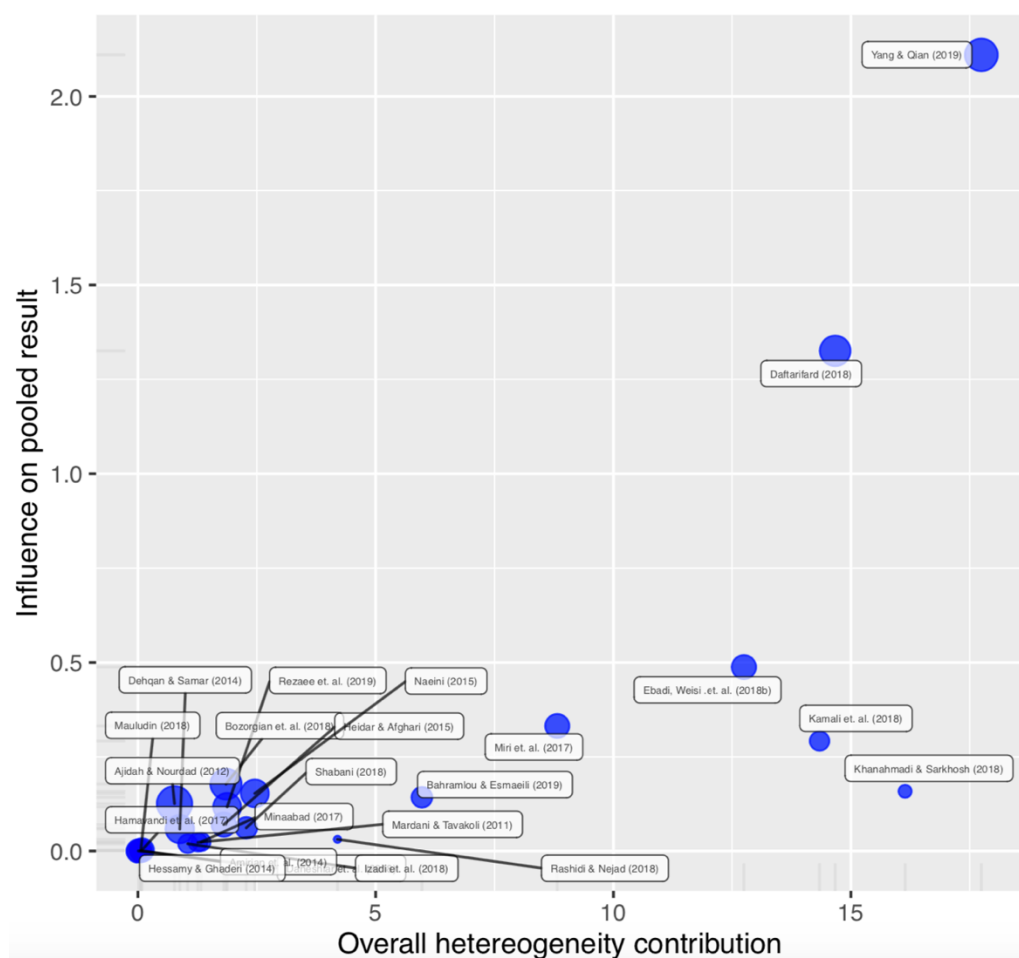
Author	Year	N	Age	Context	L2	L2 Targeted Skill(s)	Approach	Mediator
Ableeva & Lantolf	2011	7	18-20	University	French	L	interactionist	T-G
Aghaebrahimiana et. al.	2014	20		University	English	W	interventionist	T-I (G)
Ajidah & Nourdad	2012	197		University	English	R	interventionist	T-G
Alavi & Taghizadeh	2014	32 (m)	18-20	University	English	W	interventionist	T-G
Amirian et. al.	2014	28	17-20	Institute	English	R	interventionist	T-G
Bahramlou & Esmaili	2019	45	13-19	Institute	English	V	interventionist	T-G
Barabadi et. al.	2018	91	M=16	School	English	L	Both	C + T-I
Bozorgian & Alamdari	2018	180	16-24	Institute	English			
Daftarifrad	2018	185	N.M.	University	English	R	Interventionist	C

Daneshfar et. al.	2018	86		School	English	G	Interventionist	T-G
Ebadi et. al.	2018	75	18-28	Distance	Persian	V	Interventionist	C
Ebadi et. al	2018(b)	80	16-24	Institute	English	V	Interventionist	C
Ebrahimi	2015	44(F)	11-15	Institute	English	S	Both	
Erfani & Nikbin	2015	60(F)				W		T-G; P-P
Estaji & Farahanynia	2019	34	18-29	Institute	English	S	Both	T-G
Farangi & Saadi	2017	42	16-19	Institute	English	L	Interactionist	T-G
Ghonsooly & Hassanzadeh	2019	120	18-30	Institute	English	V	Interactionist	T-G
Hamavandi et. al.	2017	50 (F)	14 - 18	Institute	English	R	Interventionist	T-I
Hanifi et. al.	2016	25		University	English	V	Interventionist	T-G
Hashemi et. al.	2014	50	14-18	School	English	L	Interactionist	T-G
Heidar & Afghari	2015	60	25-32	Institute	English	L	Interactionist	T-G
Hessamy & Ghaderi	2014	50 (M)	13-16	Institute	English	V	Interactionist	T-G
Hidri	2019	132	18-20	University	English	W	Interactionist*	T-G
Izadi et. al.	2018	30	18-22	Institute	English	L	Interactionist	T-I
Jafary et. al.	2012	60 (M)	17-19		English	G	Interactionist	T-G
Kamali et. al.	2018	46(M)	15-20	Institute	English	G	Interventionist	T-G
Kamrood et. al.	2019	54	21-37	University	English	L	Interventionist	C
Khanahmadi & Sarkhosh	2018	45	14-19	Institute	English	G	Interactionist	T-G; P-P
Khonamari & Sana'ati	2016	61(M)			English	R	Interventionist	T-G
Kozulin & Levi	2018	80	17	School	English	S	Interactionist	T-I;T-G;P-P
Lin	2010							123
Lu & Hu	2019	50	10	School	English	L(phonology)	Interactionist	T-I
Mardani & Tavakoli	2011	30(M)		Institute	English	R	Interactionist	T-G
Mauludin	2018	44	16-18	School (technical institute)	English	W	Interactionist	T-G;T-I
Minaabad	2017	45		School	English	R	Both	T-I
Miri et. al.	2017	67(M)	15-17	School	English	G	Interactionist	T-I; T-G
Naeini	2014	46		University	English	R	Interactionist	T-I
Naeini	2015	102		University	English	R	Both	T-I
Nazari & Mansouri	2014	30	15-21	Institute	English	R	Interventionist	T-I
Pishghadam et. al.	2011	104	18-44(M=28)	University	English	R	Interventionist	C
Poehner & Lantolf	2013	163		University	Chinese; French; Russian	L & R	Interventionist	C
Poehner et. al.	2015	150		University	Chinese	L & R	Interventionist (based on interactionally developed prompts)	C
Rashidi & Nejad	2018	17	20-35		English	W	Interactionist	T-G
Rezaee et. al.	2019	120	18-30(m=20)	Mobile-based	English	S	Interventionist	T-I
Safa & Beheshti	2018	90(F)	15-26		English	L	Both	T-G
Shabani	2014	17	20-25	University	English	L	Interactionist	T-G
Shabani	2018	44			English	W	Interactionist	T-G
Teo	2012	68	18-19	University	English	R	Interventionist	C

Tocaimaza-Hatch	2016	9	20-26	University	English	V	Interactionist	P-P
Van Compernelle & Zhang	2014	1			English	G	Interventionist	T-I
Yang	2017	36		University	English	S	Interventionist	T-G
Yang & Qian	2017	46	M=20	University	English	R	Interventionist	C
Yang & Qian	2019	183		University	English	R	Interventionist	C
Ableeva	2010	7		University	French	L	Interactionist	T-G
Scotland	2017	52		University	English			
Lin	2010	63	3-4	Kindergarten	English	L&S	interventionist	T-I

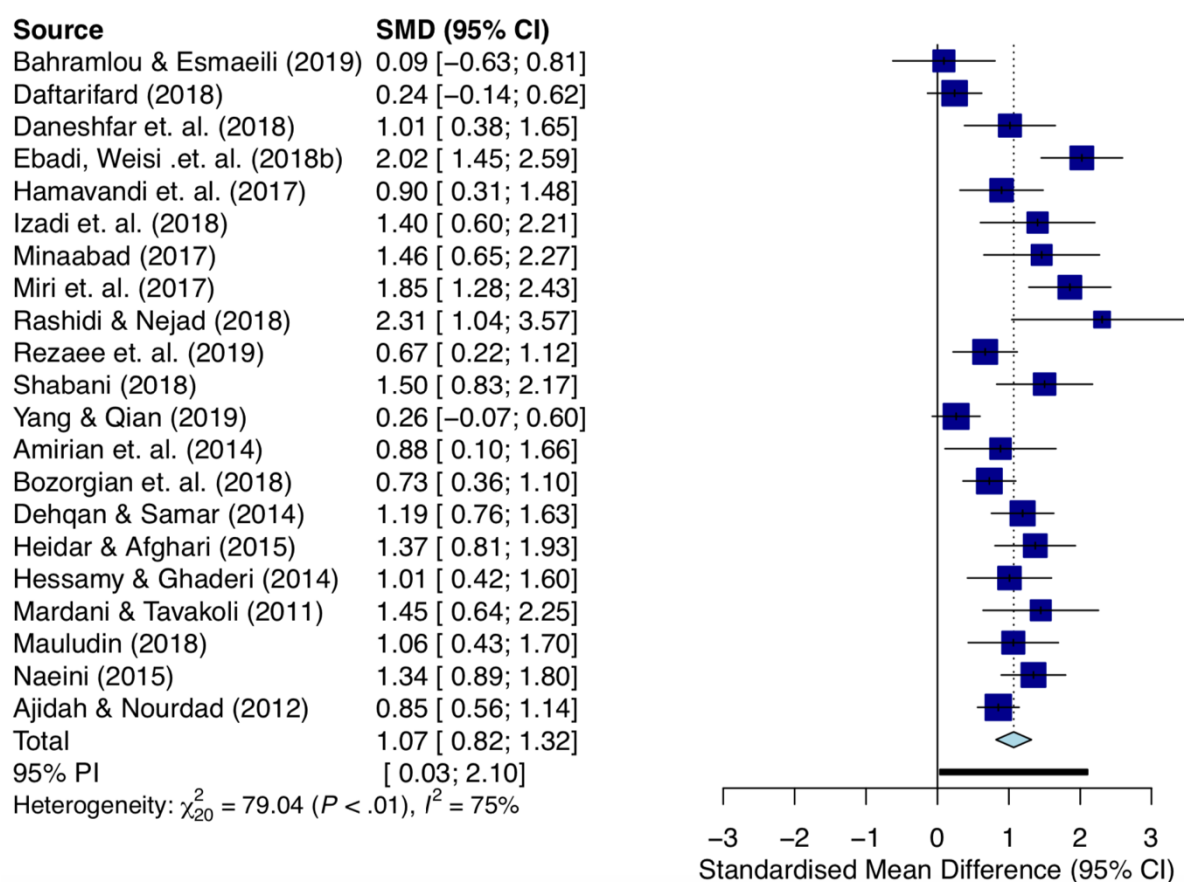
Note. (L: Listening, S: Speaking, R: Reading, W: Writing); T-G (teacher to group); T-I (teacher to individual); C (Computerized); P-P (peer to peer). (F) sample included females only; (M) sample included males only).

The initial meta-analysis run yielded high heterogeneity ( $Q = 110.42, p < 0.01$ ). The proportion of the true variation between studies that is not due to chance ( $I^2$ ) was 80%. The  $\tau^2 = 0.422$  ( $\tau = 0.65$ ) support the reported  $Q$  and  $I^2$  indicating high heterogeneity that needs to be explored. First, I used a number of techniques to identify outliers and studies contributing the most to heterogeneity. I identified extreme effect sizes based on the upper and lower confidence intervals of the pooled effect. Two studies were identified as reporting extreme effect sizes. I used the Baujat plot to graphically display studies that contributed the most to heterogeneity and the least to the pooled effect (Baujat, et. al., 2002) (see Figure 3). Furthermore, I used the GOSH plot (Olkin et. al., 2012) which helped identify possible clusters and the need for possible subgroup analysis. Further, it identifies studies that could influence the balance in these clusters. By triangulating the results from the three methods, two studies were identified to be reporting extreme effect sizes and contributing highly to heterogeneity.



**Figure 3.** Baujat Plot. Studies on the bottom right corner contribute the most to heterogeneity and the least to the pooled effect.

After removing two outlier studies, 21 studies remained included in the random-effects model. The overall mean effect size for using DA interventions in the L2 setting ( $g = 1.07$ ; 95% CI = [0.82, 1.32]) was significant (see Figure 4). The prediction interval (95% PI = [0.03, 2.10]) indicated that there is a 95% chance of future eligible studies will report a positively significant effect of DA on L2 learning. Despite that, the confidence and prediction intervals are wide, which is a feature of the Hartung-Knapp-Sidik-Jonkman method discussed earlier. Such broad intervals are accompanied with high heterogeneity in which true between-studies differences exist ( $Q = 79.04$ ,  $p < .01$ ;  $I^2 = 75\%$ ). The between-study variance ( $\tau^2 = 0.229$ ) and the estimated standard deviation for the effects ( $\tau = 0.48$ ) suggest that the overall effect size must be interpreted cautiously. Due to this heterogeneity, it is not plausible to generalize and assume a certain overall effect, but it is important to understand this variation. Thus, I did subgroup analyses to investigate the sources of variation.



**Figure 4.** Forest Plot for the random-effects model ( $k=21$ ).

The variables used in priori to explore potential heterogeneity were age, sample size, availability of a control group, medium (human vs. computer), and linguistic outcome (comprehension/perception or production). For age, there was no significant difference among school aged students (under 18), university aged students (18 and above), and mixed aged groups (including both under and above 18 years old students), ( $Q = 0.74$ ,  $p = 0.689$ ), where below 18 years old students ( $g = 1.209$  [95% CI: 0.812, 1.606],  $I^2 = 32\%$ ,  $k = 6$ ), 18 and above years old students ( $g = 1.013$  [95% CI: 0.601, 1.424],  $I^2 = 79\%$ ,  $k = 10$ ), and mixed ages classes ( $g = 1.035$  [95% CI: 0.121, 1.949],  $I^2 = 82\%$ ,  $k = 5$ ). Further, whether the mediator was a human or a computer, the effect of DA did not significantly differ ( $Q = 0.76$ ,  $p = 0.384$ ); the human mediator ( $k = 17$ ) overall effect was 1.1479 [95% CI: 0.923, 1.372], and  $I^2 = 48.1\%$ , and the computer's ( $k = 4$ ) overall effect was 0.776 [95% CI: -0.540, 2.093], and  $I^2 =$

90.4%. L2 DA did not differ across the different linguistic competencies as there were no significant difference ( $Q = 1.46$ ,  $p = 0.226$ ) between comprehension skills ( $g = 0.982$  [95% CI: 0.686, 1.277],  $I^2 = 76.4%$ ,  $k = 15$ ) and production skills ( $g = 1.304$  [95% CI: 0.717, 1.892],  $I^2 = 65.6%$ ,  $k = 6$ ).

I classified studies based on sample size in which a small sample had 30 participants or less, a medium sample had above 30 and less than 100, and a large sample had 100 or more participants. Studies almost showed a statistically significant difference in effect based on sample size ( $Q = 5.71$ ,  $p = 0.057$ ). The overall effect for large sample studies ( $k = 5$ ) was 0.676 [95% CI: 0.116, 1.234] with 80.7% true between-studies variance. Medium sample studies ( $k = 10$ ) effect was 1.248 [95% CI: 0.939, 1.558], and  $I^2 = 57.7%$ . Small sample studies ( $k = 6$ ) had an effect of 1.190 [95% CI: 0.445, 1.935], and  $I^2 = 62.8%$ .

As a true effect size could be influenced by publication bias leading to a misinterpretation of the true effectiveness of an intervention, I conducted several publication bias tests. The first is a visual technique called the funnel plot (see Figure 5). To assess bias, we seek a symmetrical distribution of studies on the two sides of the plot and based on the plot there was an indication of asymmetry that signals potential bias. Other tests helped further examine this pattern. The first was a trim and fill analysis which added seven studies to reach a symmetrical distribution that does not lead to bias, and it produced a new overall effect ( $g = 0.801$ ); the new estimate could be compared with the original ( $g = 1.07$ ) and this showed that there is bias. Eggers test was significant ( $p = 0.0124$ ) adding more evidence to the existence of publication bias. Yet, Simonsohn, Nelson, and Simmons (2014b) argued that the trim and fill analysis might not be accurate in providing effect size estimates as it "...assumes that the publication bias process suppresses the publication of small effects (regardless of significance) rather than nonsignificant results (regardless of effect size)" (p. 667). They introduced  $p$ -curve as a solution to correct for the bias of publishing only significant results (p. 666) and offer an approach to estimating the true effect size (p.667). Ignoring nonsignificant results is one of the limitations of the  $p$ -curve although they stress that "it does not make  $p$ -curve biased" (p. 675). Eighteen studies with significant results out of the 21 articles were included in the  $p$ -curve analysis. The right skewness test indicated the  $p$ -curve is significantly skewed to the right which suggested the existence of a true effect. Further, the flatness test is nonsignificant suggesting an estimated power of 98% [96% , 99%]. Thus, as stated in  $p$ -curve's output, there was evidential value which was not inadequate indicating that there was a true effect that did not result from publication bias. (see Figure 6).

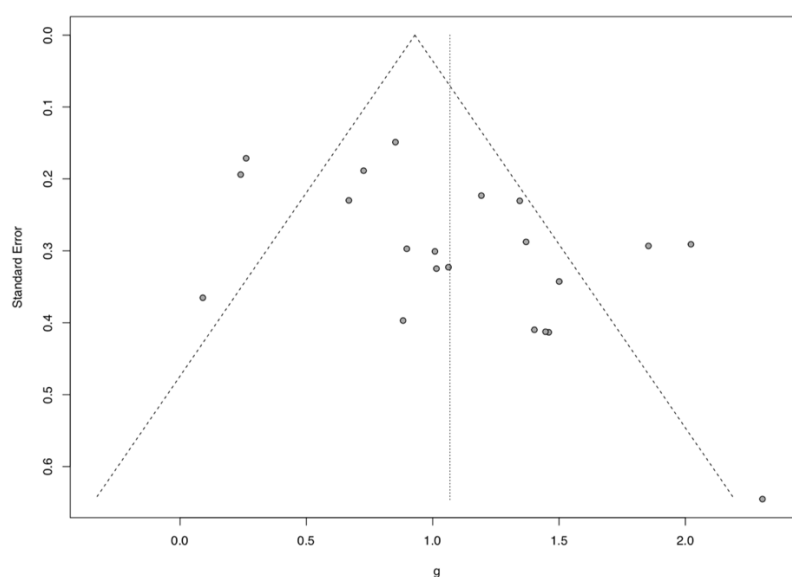


Figure 5. Funnel Plot.

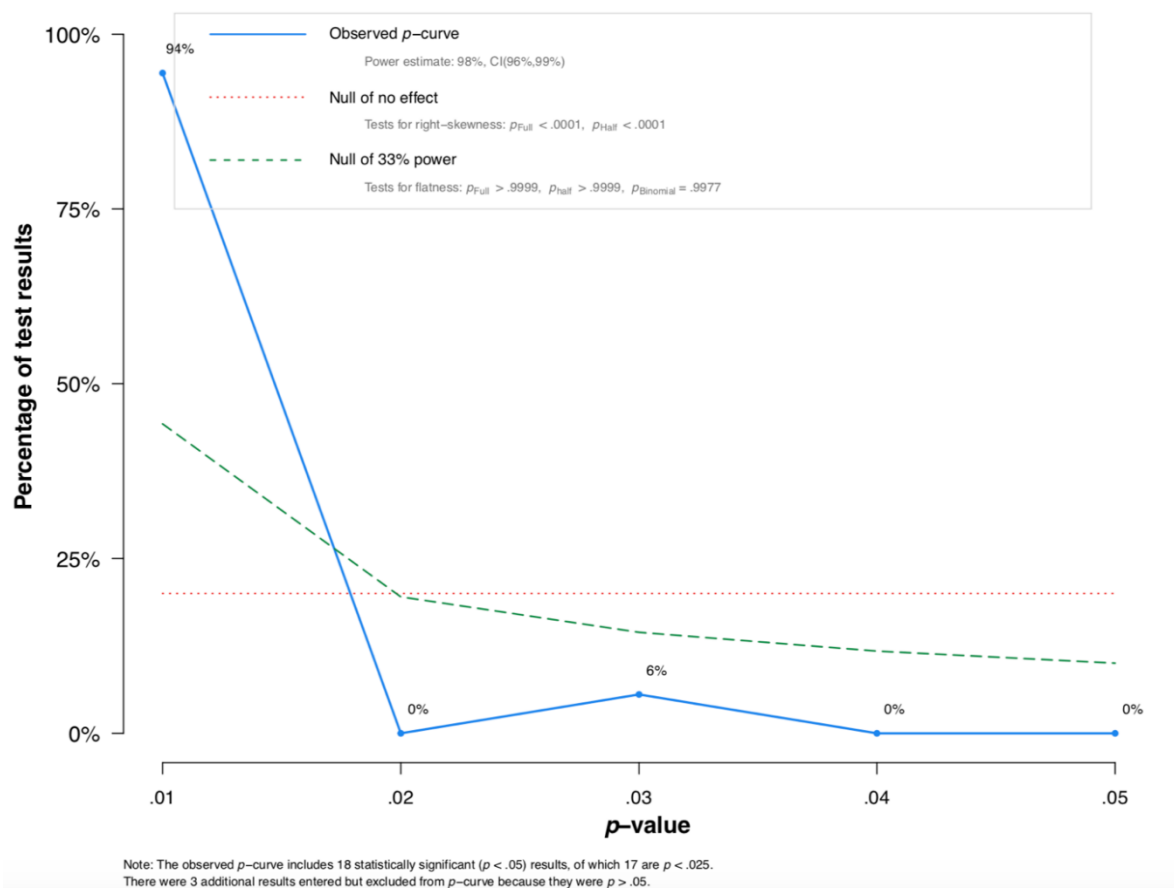


Figure 6. p-curve.

## Discussion

While 2. learning. Additionally, the p-curve analysis indicates a true effect in the meta-analyzed studies, suggesting further investigation is warranted.

Building on Li and Li's (2015) review, several findings are supported here. First, DA improves L2 learners' linguistic competencies, as most studies in this review reported significant positive effects (e.g., Ableeva & Lantolf, 2011; Ebadi et al., 2018). Second, both interventionist and interactionist approaches are commonly used in DA research, with 47% of studies utilizing the former and 42% the latter. Third, over half of the studies focused on reading and grammar, which aligns with Li and Li's (2015) results. Moreover, while computer-assisted DA is still in its infancy, 18% of studies implemented technology such as computers and cellphones (e.g., Rezaee et al., 2019). However, no dedicated DA software for L2 purposes exists at this time.

Collaborative efforts between educational technologists, applied linguists, and assessment experts could help establish a framework for DA technology development. However, validity and reliability issues must first be addressed, as DA currently lacks standardized approaches to these concerns (Kozulin & Garb, 2002). Poehner (2008) advocates for formative assessments that support learner development rather than strictly measuring abilities, aligning with Messick's unified model of validity. Still, Alonazi (2017) criticized computerized DA for its lack of complementary evidence regarding validity, highlighting ongoing challenges.

Li and Li (2015) also noted small sample sizes in their review, but this study shows an increase in L2 DA research with larger sample sizes—77% of studies had over 30 participants compared to 32% in the previous review. This shift indicates that L2 DA research may be reaching a critical point, requiring further investigation to identify future research directions.

The reviewed papers often failed to address the nature of the mediation provided, relying instead on L2-oriented mediation strategies or classical approaches that lacked precise definitions of 'language' and 'second language learning.' A clearer conceptualization of these constructs is essential in designing effective DA interventions. For instance, mediation should be tailored to interact with L2 acquisition factors like age, motivation, and attention, recognizing that mediation strategies effective for adults may not work as well for younger learners, given the implications of the critical period hypothesis (Lenneberg, 1967).

DA can serve various functions in L2 contexts, including diagnosing student weaknesses, informing instruction, and acting as a placement or predictive tool. Its application across different learning environments, student age groups, and skills demonstrates its versatility. However, further development of computerized DA, alongside guidelines for L2 practitioners, is necessary for its broader adoption.

### *Limitations*

One key limitation of this review is its focus on quantitative research, excluding qualitative studies. The initial focus on quantitative studies was driven by two factors: first, to conduct a meta-analysis that quantifies the effect of DA in the L2 setting, and second, to assess the practicality of large-scale DA implementation, as quantitative research often aims for generalizable findings.

A second limitation is the reliance on extracting a single effect size per study, as recommended by meta-analysis guidelines (e.g., Lipsey & Wilson, 2001). However, recent advancements in meta-analytic techniques, such as multilevel meta-analysis, allow for the inclusion of multiple effect sizes, which could address some issues associated with single-effect-size extraction, such as its impact on the overall mean effect size.

The third limitation is the lack of quality evaluations for the included papers. While rating paper quality may introduce subjectivity, it is important to assess the rigor of research, especially given the field's infancy. L2 DA research often replicates original studies, which helps establish foundational knowledge but may limit innovation. In a dynamic field like L2 teaching and learning, flexibility, practicality, and attention to learner needs, school contexts, and policy constraints are crucial.

### *Future Directions*

Future L2 DA research must incorporate language theories that explain language processing and build stronger validity arguments. Validity and reliability need greater attention, as dismissing the relevance of reliability could undermine DA's acceptance as an assessment method. Consistency in mediation should be considered, factoring in learners' levels and targeted linguistic constructs. Utilizing frameworks like item response theory can help assess the effectiveness of specific mediation prompts in promoting learner progress.

Additionally, research should examine DA's impact on learners' motivation. Motivation, self-esteem, self-efficacy, and agency all influence student performance, and these constructs should be studied to understand how they affect DA outcomes. For example, students may not benefit from DA due to low self-esteem, and their development might fluctuate based on agency levels. However, most reviewed studies have not explored these factors in L2 DA contexts.

The type of mediation offered also warrants investigation, as content-based hints may impact progress differently from strategy-based mediations. L2 DA research is still in its early stages, and establishing DA as a valid, reliable method that integrates instruction and assessment is crucial for advancing L2 learning. Researchers should address DA's practicality in the classroom and its various functions explicitly.

## **Conclusions & Recommendations**

DA in the L2 setting has proven to be a useful method for improving learners' performance and providing teachers with valuable insights for evaluating their practices. It offers information on

students' learning potential, which may correlate with unstudied linguistic abilities (e.g., Kozulin & Levi, 2018). Applied in various contexts and age groups, L2 DA consistently produces positive results. It can be delivered in-class or online, individually or in groups. However, more research is needed on constructs like self-efficacy, agency, self-esteem, socioeconomic status, and test anxiety for a comprehensive understanding of DA's impact.

Despite positive results, L2 DA is still in its early stages and requires further refinement. I recommend that L2 DA be integrated into classroom assessment, facilitating its inclusion in textbooks and teacher training, and promoting its use as both an instructional and assessment tool.

I propose a framework to guide future L2 DA research, which outlines three main purposes of DA and explores five key topics: 1) potential functions, 2) mediation types, 3) necessary requirements, 4) suitable mediums, and 5) practicality. Table 2 provides an elaborate discussion on each purpose. While still in development, this framework offers a practical approach to DA in L2 contexts, serving as a roadmap for future research to establish DA more firmly in the field.

**Table 2.** A framework to guide future research through the identification of purposes and functions of dynamic assessment in the L2 context.

Purpose	L2 DA as an instruction method	L2 DA as an assessment method	L2 DA as a service method
Functions	To help evaluate the effectiveness of a teaching method.	To check understanding while teaching. To determine if learning outcomes were met and plan near or far future learning experiences.	To help place students in their proper levels in second/foreign language learning courses/programs.
	To evaluate the appropriateness of learning outcomes to L2 learners.	To provide potentially less anxious assessment environments. To evaluate the appropriateness of assessments from students' responses before giving summative assessments that could be over or below the levels of the students.	To help provide curriculum developers and textbook writers with the actual skills needed for each level based on the identified skills from DA practices.
	To target students' actual needs by identifying their zones of proximal development and plan instruction accordingly.	To assess skills that are of an interactional nature such as speaking by creating more authentic experiences (authenticity).	To pilot summative and large-scale test items to learners to assess its suitability for the targeted level and identify problems with items which cannot be measured but could be investigated interactionally as with cultural bounded issues.
	To equip students with learning strategies and techniques to develop self-regulation (and related constructs like agency).	To evaluate the progress of students within their zones of proximal development and provide accurate estimations of the odds of success in a course.	To maximize students learning experiences in writing centers and one-to-one conferencing by providing a rich reciprocal experience.
		To provide detailed reports about students' progress that contain information about their actual performance, mediated performance, and learning potential to parents, policy makers, and curriculum developers.	To help identify talented learners through unassisted scores and learning potential scores.
Mediation	Could be applied using an interactionist or an interventionist approach. Different types of prompts can be used mapped with the adopted approach.	Can be applied using an interactionists or interventionist approach. Mediation targets content; the formation of prompts may not focus on metacognition as it assesses change towards achieving the learning outcomes of a course.	Can be of various forms according to the targeted function. Mediation is of an exploratory nature hence gradual prompts might be suitable for situations where the goal is to predict instead of modify learning.
	Prompts may not only focus on helping acquire knowledge but also leaning strategies that contribute to ones metacognition.		

Requirements	<p>A conceptualization of the nature of language and language learning by adopting theories from fields such as second language acquisition and psycholinguistics.</p> <p>Understanding the effect of constructs such as attention, motivation, and age on second language learning.</p>	<p>Strong validity arguments.</p> <p>Clear purposes of assessment aligned with the learning outcomes.</p> <p>Short, weekly or unit based, to allow for rich reflections and continuous development of the course and swift interreference to address students' learning difficulties.</p> <p>Careful design of prompts which focuses on helping learners achieve the outcomes content wise.</p> <p>Reliability of mediation offered (prescribed mediation) must be investigated and compared among learners who share similar characteristics.</p>	<p>A clearly articulated purpose and goal of the DA intervention.</p> <p>A well-designed intervention that aligns with the intended function (e.g., an interactionist approach with no prescribed prompts for writing conferences vs. an interventionist approach with a prescribed mediation offered to place a student in the accurate level within a program).</p> <p>Reliability must be investigated and addressed thoroughly for functions that require producing accurate results such as placing students in accurate learning levels and identifying students of talent or special need.</p> <p>Validity evidence must be appropriate for the intended function.</p>
Medium	<p>teacher to students</p> <p>computerized</p>	<p>teacher to student</p> <p>teacher to students</p> <p>computerized</p>	<p>teacher to student</p> <p>Computerized</p>
Practicality	<p>Allows for the modification/adoption of teaching approaches by providing necessary evidence.</p> <p>Targets the whole class and emphasizes an equal opportunity learning experience considerate of whole class development.</p> <p>Allows for the identification of underperforming or overperforming students and creating meaningful groupings based on actual levels which will maximize learning from classes and learning from peers.</p>	<p>Can be applied in class or outside class: hence saving time for more instruction time.</p> <p>Can provide further assessments in response to previous assessments to evaluate change between assessments.</p> <p>Helps design less anxious assessment experiences.</p> <p>Rich assessments of few or single learning outcomes yielding richer information.</p> <p>Meaningful assessment reports comprehended by parents and students even; number scores is meaningful as it reflects performance unassisted and assisted.</p>	

## References

1. Ableeva, R. (2010). *Dynamic assessment of listening comprehension in second language learning* (Ph.D., The Pennsylvania State University). Retrieved from <http://search.proquest.com/pqdtglobal/docview/816855305/abstract/EFD10DAB16AD46B5PQ/2>
2. Ableeva, R., & Lantolf, J. (2011). Mediated dialogue and the microgenesis of second language listening comprehension. *Assessment in Education: Principles, Policy & Practice*, 18(2), 133–149. <https://doi.org/10.1080/0969594X.2011.555330>
3. Aghaebrahimian, A., Rahimirad, M., Ahmadi, A., & Alamdari, J. K. (2014). Dynamic Assessment of Writing Skill in Advanced EFL Iranian Learners. *Procedia - Social and Behavioral Sciences*, 98, 60–67. <https://doi.org/10.1016/j.sbspro.2014.03.389>
4. Ajideh, P., & Nourdad, N. (2012). The Immediate and Delayed Effect of Dynamic Assessment on EFL Reading Ability. *English Language Teaching*, 5(12), 141–151.

5. Alavi, S. M., & Taghizadeh, M. (2014). Dynamic Assessment of Writing: The Impact of Implicit/Explicit Mediations on L2 Learners' Internalization of Writing Skills and Strategies. *Educational Assessment*, 19(1), 1–16. <https://doi.org/10.1080/10627197.2014.869446>
6. Amirian, M. R., Davoudi, M., & Ramazanian, M. (2014). The Effect of Dynamic Assessment on Iranian EFL Learners' Reading Comprehension. *Advances in Language and Literary Studies; Footscray*, 5(3), 191–194.
7. Bahramlou, K., & Esmaeili, A. (2019). The Effects of Vocabulary Enhancement Exercises and Group Dynamic Assessment on Word Learning Through Lexical Inferencing. *Journal of Psycholinguistic Research*, 48(4), 889–901. <https://doi.org/10.1007/s10936-019-09638-x>
8. Bakhoda, I., & Shabani, K. (2019). Enhancing L2 learners' ZPD modification through computerized-group dynamic assessment of reading comprehension. *Innovation in Language Learning and Teaching*, 13(1), 31–44. <https://doi.org/10.1080/17501229.2017.1286350>
9. Barabadi, E., Khajavy, G. H., & Kamrood, A. M. (2018). Applying Interventionist and Interactionist Approaches to Dynamic Assessment for L2 Listening Comprehension. *International Journal of Instruction*, 11(3), 681–700.
10. Bozorgian, H., & Fakhri Alamdari, E. (2018). Multimedia listening comprehension: Metacognitive instruction or metacognitive instruction through dialogic interaction. *ReCALL : The Journal of EUROCALL; Cambridge*, 30(1), 131–152. <http://dx.doi.org.mutex.gmu.edu/10.1017/S0958344016000240>
11. Chen, Y. (2019). Developing Students' Critical Thinking and Discourse Level Writing Skill Through Teachers' Questions: A Sociocultural Approach. *Chinese Journal of Applied Linguistics*, 42(2), 141–162. <https://doi.org/10.1515/CJAL-2019-0009>
12. Daftarifard, P. (2018). The effect of mediational artifacts on EFL learners' reading comprehension performance. *International Journal of Language Studies*, 12(3), 73–90.
13. Daneshfar, S., Aliasin, S. H., & Hashemi, A. (2018). The Effect of Dynamic Assessment on Grammar Achievement of Iranian Third Grade Secondary School EFL Learners. *Theory and Practice in Language Studies*, 8(3), 295–305. <http://dx.doi.org.mutex.gmu.edu/10.17507/tpis.0803.04>
14. Dehqan, M., & Samar, R. G. (2014). Reading Comprehension in a Sociocultural Context: Effect on Learners of Two Proficiency Levels. *Procedia - Social and Behavioral Sciences*, 98, 404–410. <https://doi.org/10.1016/j.sbspro.2014.03.433>
15. Dynamic Assessment of Incidental Vocabularies: A Case of Iranian ESP Learners. (2016). *Advances in Language and Literary Studies*, 7(2). <https://doi.org/10.7575/aiac.all.v.7n.2p.163>
16. Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning*, 32(5–6), 527–555. <https://doi.org/10.1080/09588221.2018.1527362>
17. Ebadi, S., Asakereh, A., & Link to external site, this link will open in a new window. (2017). Developing EFL Learners' Speaking Skills through Dynamic Assessment: A Case of A Beginner and an Advanced Learner. *Cogent Education*, 4(1), 18. <http://dx.doi.org.mutex.gmu.edu/10.1080/2331186X.2017.1419796>
18. Ebadi, S., Link to external site, this link will open in a new window, Vakilifard, A., Bahramlou, K., & Link to external site, this link will open in a new window. (2018). Learning Persian vocabulary through reading: The effects of noticing and computerized dynamic assessment. *Cogent Education; Abingdon*, 5(1). <http://dx.doi.org.mutex.gmu.edu/10.1080/2331186X.2018.1507176>
19. Ebadi, S., Weisi, H., Monkaresi, H., & Bahramlou, K. (2018). Exploring lexical inferencing as a vocabulary acquisition strategy through computerized dynamic assessment and static assessment. *Computer Assisted Language Learning*, 31(7), 790–817. <https://doi.org/10.1080/09588221.2018.1451344>
20. Ebrahimi, E. (2015). The effect of dynamic assessment on complexity, accuracy, and fluency in EFL learners' oral production. *International Journal of Research Studies in Language Learning*, 4(3). <https://doi.org/10.5861/ijrsl.2015.982>
21. Elola, I., & Oskoz, A. (2016). Supporting Second Language Writing Using Multimodal Feedback. *Foreign Language Annals; Alexandria*, 49(1), 58–74. <http://dx.doi.org.mutex.gmu.edu/10.1111/flan.12183>
22. Erfani, S. S., & Nikbin, S. (2015). The Effect of Peer-Assisted Mediation vs. Tutor-Intervention within Dynamic Assessment Framework on Writing Development of Iranian Intermediate EFL Learners. *English Language Teaching*, 8(4), 128–141.

23. Estaji, M., & Farahanynia, M. (2019). The Immediate and Delayed Effect of Dynamic Assessment Approaches on EFL Learners' Oral Narrative Performance and Anxiety. *Educational Assessment*, 24(2), 135–154. <https://doi.org/10.1080/10627197.2019.1578169>
24. Farangi, M. R., & Saadi, Z. K. (2017). Dynamic assessment or schema theory: The case of listening comprehension. *Cogent Education*, 4(1). <http://dx.doi.org/mutex.gmu.edu/10.1080/2331186X.2017.1312078>
25. Ghonsooly, B., & Hassanzadeh, T. (2019). Effect of interactionist dynamic assessment on English vocabulary learning: Cultural perspectives in focus. *Issues in Educational Research*, 29(1), 70–88.
26. Hamavandi, M., Rezai, M. J., & Mazdayasna, G. (2017). Dynamic assessment of morphological awareness in the EFL context. *Cogent Education*, 4(1). <http://dx.doi.org/mutex.gmu.edu/10.1080/2331186X.2017.1324254>
27. Hanifi, S., Nasiri, M., & Aliasin, H. (2016). Dynamic Assessment of Incidental Vocabularies: A Case of Iranian ESP Learners. *Advances in Language and Literary Studies; Footscray*, 7(2), 163–170.
28. Hanjani, A. M., & Li, L. (2014). Exploring L2 writers' collaborative revision interactions and their writing performance. *System*, 44, 101–114. <https://doi.org/10.1016/j.system.2014.03.004>
29. Hashemi Shahraki, S., Ketabi, S., & Barati, H. (2014). Group dynamic assessment of EFL listening comprehension: Conversational implicatures in focus. *International Journal of Research Studies in Language Learning*, 4(3). <https://doi.org/10.5861/ijrsl.2014.955>
30. Heidar, D. M., & Afghari, A. (2015). The Effect of Dynamic Assessment in Synchronous Computer-Mediated Communication on Iranian EFL Learners' Listening Comprehension Ability at Upper-Intermediate Level. *English Language Teaching*, 8(4), 14–23.
31. Hessamy, G., & Ghaderi, E. (2014). The Role of Dynamic Assessment in the Vocabulary Learning of Iranian EFL Learners. *Procedia - Social and Behavioral Sciences*, 98, 645–652. <https://doi.org/10.1016/j.sbspro.2014.03.463>
32. Hidri, S. (2019). Static vs. dynamic assessment of students' writing exams: A comparison of two assessment modes. *International Multilingual Research Journal*, 13(4), 239–256. <https://doi.org/10.1080/19313152.2019.1606875>
33. Izadi, M., Khoshshima, H., Nourmohammadi, E., & Yarahmadzehi, N. (2018). Mediation Strategies in a Dynamic Assessment Approach to L2 Listening Comprehension: Different Ability Levels in Focus 1. *Italian Sociological Review; Verona*, 8(3), 445–466.
34. Jafary, M. R., Nordin, N., & Mohajeri, R. (2012). The Effect of Dynamic versus Static Assessment on Syntactic Development of Iranian College Preparatory EFL Learners. *English Language Teaching*, 5(7), 149–157.
35. Kamali, M., Abbasi, M., & Sadighi, F. (2018). The Effect of Dynamic Assessment on L2 Grammar Acquisition by Iranian EFL Learners. *International Journal of Education and Literacy Studies*, 6(1), 72–78.
36. Mehri Kamrood, A., Davoudi, M., Ghaniabadi, S., & Amirian, S. M. R. (2019). Diagnosing L2 learners' development through online computerized dynamic assessment. *Computer Assisted Language Learning*, 1–30.
37. Kao, Y.-T. (2014). *Vygotsky's theory of instruction and assessment: The implications on foreign language education* (Ph.D., The Pennsylvania State University). Retrieved from <http://search.proquest.com/pqdtglobal/docview/1639700013/abstract/B97860CF0EB34B8EPQ/2>
38. Khanahmadi, F., & Sarkhosh, M. (2018). Teacher-vs. Peer-Mediated Learning of Grammar through Dynamic Assessment: A Sociocultural Perspective. *International Journal of Instruction*, 11(4), 207–222.
39. Khonamri, Fatameh, & Sana'ati, M. K. (2016). The Effect of Dynamic Assessment on Iranian EFL Students' Critical Reading Performance. *Kesan Penilaian Dinamik Terhadap Penguasaan Pembacaan Kritis Pelajar EFL Iran.*, 41(2), 115–123.
40. Khonamri, Fatemeh, & Sana'ati, M. K. (2014). The Impacts of Dynamic Assessment and CALL on Critical Reading: An Interventionist Approach. *Procedia - Social and Behavioral Sciences*, 98, 982–989. <https://doi.org/10.1016/j.sbspro.2014.03.508>
41. Kozulin, A., & Levi, T. (2018). EFL Learning Potential: General or Modular? *Journal of Cognitive Education & Psychology*, 17(1), 16–27. <https://doi.org/10.1891/1945-8959.17.1.16>

42. Lin, Z. (2010). Interactive Dynamic Assessment with Children Learning EFL in Kindergarten. *Early Childhood Education Journal; New York*, 37(4), 279–287. <http://dx.doi.org.mutex.gmu.edu/10.1007/s10643-009-0356-6>
43. Lu, Y.-Y., & Hu, C.-F. (2019). Dynamic assessment of phonological awareness in young foreign language learners: Predictability and modifiability. *Reading and Writing*, 32(4), 891–908. <https://doi.org/10.1007/s11145-018-9894-x>
44. Mardani, M., & Tavakoli, M. (2011). Beyond Reading Comprehension: The Effect of Adding a Dynamic Assessment Component on EFL Reading Comprehension. *Journal of Language Teaching & Research*, 2(3), 688–696. <https://doi.org/10.4304/jltr.2.3.688-696>
45. Mauludin, L. A. (2018). Dynamic assessment to improve students' summary writing skill in an ESP class. *Southern African Linguistics and Applied Language Studies*, 36(4), 355–364. <https://doi.org/10.2989/16073614.2018.1548296>
46. Mawlawi Diab, N. (2016). A comparison of peer, teacher and self-feedback on the reduction of language errors in student essays. *System*, 57, 55–65. <https://doi.org/10.1016/j.system.2015.12.014>
47. Minaabad, M. S. (2017). Study of the Effect of Dynamic Assessment and Graphic Organizers on EFL Learners' Reading Comprehension. *Journal of Language Teaching & Research*, 8(3), 548–555. <https://doi.org/10.17507/jltr.0803.13>
48. Miri, M., Alibakhshi, G., Kushki, A., & Bavarsad, P. S. (2017). Going beyond One-to-One Mediation in Zone of Proximal Development (ZPD): Concurrent and Cumulative Group Dynamic Assessment. *Eurasian Journal of Applied Linguistics; Ankara*, 3(1), 1–24.
49. Naeini, J. (2014). On the Study of DA and SLA: Feuerstein's MLE and EFL Learners' Reading Comprehension. *Procedia - Social and Behavioral Sciences*, 98, 1297–1306. <https://doi.org/10.1016/j.sbspro.2014.03.546>
50. Naeini, J. (2015). A Comparative Study of the Effects of Two Approaches of Dynamic Assessment on the Reading Comprehension of Iranian EFL Learners. *International Journal of Applied Linguistics & English Literature; Footscray*, 4(2), 54–67. <http://dx.doi.org.mutex.gmu.edu/10.7575/aiac.ijalel.v.4n.2p.54>
51. Nazari, B., & Mansouri, S. (2014). Dynamic assessment versus static assessment: A study of reading comprehension ability in Iranian EFL learners. *Dil ve Dillilimi Çalışmaları Dergisi*, 10(2).
52. Park, K. (2010). *Enhancing lexicogrammatical performance through corpus-based mediation in L2 academic writing instruction* (Ph.D., The Pennsylvania State University). Retrieved from <http://search.proquest.com/pqdtglobal/docview/1732856900/abstract/EFD10DAB16AD46B5PQ/6>
53. Pishghadam, R., Barabadi, E., & Kamrood, A. M. (2011). The Differing Effect of Computerized Dynamic Assessment of L2 Reading Comprehension on High and Low Achievers. *Journal of Language Teaching & Research*, 2(6), 1353–1358. <https://doi.org/10.4304/jltr.2.6.1353-1358>
54. Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA). *Language Teaching Research*, 17(3), 323–342. <https://doi.org/10.1177/1362168813482935>
55. Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337–357. <https://doi.org/10.1177/0265532214560390>
56. Rashidi, N., & Bahadori Nejad, Z. (2018). An Investigation Into the Effect of Dynamic Assessment on the EFL Learners' Process Writing Development. *SAGE Open*, 8(2), 215824401878464. <https://doi.org/10.1177/2158244018784643>
57. Rezaee, A. A., Alavi, S. M., & Razzaghifard, P. (2019). The impact of mobile-based dynamic assessment on improving EFL oral accuracy. *Education and Information Technologies*, 24(5), 3091–3105. <https://doi.org/10.1007/s10639-019-09915-1>
58. Safa, M. A., & Beheshti, S. (2018). Interactionist and Interventionist Group Dynamic Assessment (GDA) and EFL Learners' Listening Comprehension Development. *Iranian Journal of Language Teaching Research*, 6(3), 37–56.

59. Scotland, J. (2017). *Participating in a shared cognitive space: An exploration of working collaboratively and longer-term performance of a complex grammatical structure*. Retrieved from <https://ore.exeter.ac.uk/repository/handle/10871/32739>
60. Shabani, K. (2014). Dynamic Assessment of L2 Listening Comprehension in Transcendence Tasks. *Procedia - Social and Behavioral Sciences*, 98, 1729–1737. <https://doi.org/10.1016/j.sbspro.2014.03.600>
61. Shabani, K. (2018). Group Dynamic Assessment of L2 Learners' Writing Abilities. *Iranian Journal of Language Teaching Research*, 6(1), 129–149.
62. Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, 143(2), 534.
63. Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666–681.
64. Tavakoli, M., & Nezakat-Alhossaini, M. (2014). Implementation of corrective feedback in an English as a foreign language classroom through dynamic assessment. *Dil ve Dilbilimi Çalışmaları Dergisi*, 10(1), 211–232.
65. Teo, A., & Jen, F. (2012). Promoting EFL students' inferential reading skills through computerized dynamic assessment. *Language Learning & Technology*, 16(3), 10–20.
66. Tocaimaza-Hatch, C. C. (2016). Mediated Vocabulary in Native Speaker-Learner Interactions During an Oral Portfolio Activity. *Foreign Language Annals; Alexandria*, 49(2), 336–354. <http://dx.doi.org.mutex.gmu.edu/10.1111/flan.12190>
67. van Compernelle, R. A., & Zhang, H. (Stanley). (2014). Dynamic assessment of elicited imitation: A case analysis of an advanced L2 English speaker. *Language Testing*, 31(4), 395–412. <https://doi.org/10.1177/0265532213520303>
68. Yang, X. (2017). Dynamic Assessment in English Pronunciation Teaching: From the Perspective of Intellectual Factors. *Theory and Practice in Language Studies; London*, 7(9), 780–785. <http://dx.doi.org.mutex.gmu.edu/10.17507/tpls.0709.10>
69. Yang, Y., & Qian, D. D. (2017). Assessing English reading comprehension by Chinese EFL learners in computerized dynamic assessment. *Language Testing in Asia; Heidelberg*, 7(1), 1–15. <http://dx.doi.org.mutex.gmu.edu/10.1186/s40468-017-0042-3>
70. Yang, Y., & Qian, D. D. (2019). Promoting L2 English learners' reading proficiency through computerized dynamic assessment. *Computer Assisted Language Learning*, 1–25. <https://doi.org/10.1080/09588221.2019.1585882>
71. Zhao, Y. (2018). *Both Sides of the Interaction: Native Speaker and Nonnative Speaker Tutors Serving Nonnative Speaker Students in a University Writing Center* (Ph.D., Oklahoma State University). Retrieved from <http://search.proquest.com/pqdtglobal/docview/2271929299/abstract/EFD10DAB16AD46B5PQ/7>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.