



The Effect of Computerized Dynamic Assessment on Working Memory Span of EFL Learners while Listening

Masoomeh Estaji

Allameh Tabataba'i University, Tehran, Iran

Fatemeh Safari

Allameh Tabataba'i University, Tehran, Iran

Introduction

Assessment has a significant role in the academic success and failure of students in all disciplines. Considering the scope of this study, i.e., applied linguistics, it has been found that traditional and one-directional approaches to assessment in which the test taker is the object of the test rather than being a part of the process are challenging for the learners (Dochy, 2001). Various issues such as the negative washback effect, and low retention of content are the result of focusing on the tests as a product rather than a process (Butler & McMunn, 2006). As a result, alternative approaches to assessment which aim at reducing the tension between assessment and instruction are being investigated and implemented. One of these new trends in assessment is Dynamic Assessment (hereafter referred to as DA).

Azarian, Nourdad, and Nouri (2016) defined DA as a process-oriented approach to integrate assessment with instruction. DA takes a monistic view toward assessment and instruction. It is test-treatment-retest format based on modification of the test takers' needs. In simple words, being dynamic is the result of learners' response to amend and improve test performance. DA has been found to decrease the learners' anxiety and increase their positive attitudes (Lantolf & Poehner, 2004). It has also increased the language proficiency of the learners (Fahmy, 2013). The question left open is the extent to which DA can affect the psychological aspects of learning such as working memory span.

Poehner, Zhang, and Lu (2015) acknowledge that DA is an effective approach to the diagnosis of the learners' needs, as the approach is based on Vygotsky's (1986) socio-cultural theory which is a profound theory with practical implications. However, they also admit that Computerized DA (hereafter referred to as C-DA) can be even more effective than DA; due to it being more meticulous. They suggested the implementation of an online and computerized test to run C-DA.

Memory span, as a leading psychological factor in learning, has been defined as the number of items an individual can recall after exposure to the items (Lustig, May, & Hasher, 2001). It is an indicator of short term memory which is affected by a number of factors, such as the characteristics of materials used, addition of non-target elements, rhythm of presentation, rate of presentation, and modality of presentation (Lustig et al., 2001). Previous studies have revealed that many language learners have problems with the listening section of tests, which is as a result of low recall of information (Goh, 2000). Remembering the

items in listening tasks, after having listened to the audio files, is among the main causes affecting the students' performance in listening tests (Vandergrift, 2007).

This problem has been pinpointed in the English as a foreign language (EFL) context of Iran by Atasheneh and Izadi (2012) who suggest focusing on the process to increase the working memory span of the test takers. Although DA is a process-oriented approach to assessment, research dealing with the effect of DA and C-DA, in particular, on working memory span is sparse. Understanding the effect of C-DA on working memory span is crucial to educators in the field, as there is a constant quest for more practical approaches to memory span increase. Teachers can also take advantage of such studies to make decisions about using C-DA in their language classes. Likewise, the use of online tests is becoming more pervasive; thus, investigating the effects of such tests can help test developers and curriculum designers make more accurate decisions about the use of these tests.

To delve into the use and usefulness of C-DA, this study has pursued three main purposes. The first and main purpose of the study was to examine the extent to which the implementation of C-DA in the Iranian EFL context has affected the language learners' listening comprehension. The second purpose of this study was to investigate the extent to which C-DA has affected the working memory span of Iranian EFL learners. Finally, this study examined if the possible changes in the listening comprehension scores of the participants has been as a result of the interactional effect of C-DA and working memory. In particular, this study sought to address the following questions.

1. To what extent does the implementation of C-DA affect Iranian EFL learners' listening comprehension?
2. To what extent does C-DA affect the working memory span of Iranian EFL learners?
3. To what extent does change in the listening comprehension scores of Iranian EFL learners result from the interactional effect of C-DA and working memory?

Method

Participants

The participants of this study were selected non-randomly through convenience sampling. A total of 101 EFL learners who were already studying at intermediate level at a language institute in Tehran were given the pen-and-paper Oxford Placement Test (OPT) as a homogeneity test. Through considering the English usage scores of the test (Scoring between 30-47, B1 and B2 level of Common European Framework of Reference, according to the council of Europe), which were measured according to the acceptable and reliable conversion chart of the OPT, 51 EFL students were selected to participate in the study. Table 1 reveals the demographics of the participants.

TABLE 1
Demographics of the Participants

N	Age	Gender	Proficiency Level	Language learning Experience	Nationality	Educational Background
51	19-27	Male = 24 Female = 27	Intermediate	About two years	Iranian	31 undergraduates 20 graduates

These participants took a retired version of TOEFL iBT listening test as a pre-test. Based on the results of the listening section of the TOEFL iBT test, the participants were randomly divided into a control group (n = 25) and an experimental group (n = 26) with no statistical difference in their TOEFL iBT scores. To ensure there was no statistical differences between the groups prior to the study, an independent samples t-test was run.

TABLE 2

Descriptive Statistics; Independent Samples t-test for Learners Homogeneity Test

	Groups	N	Mean	Std. Deviation	Std. Error Mean
Homogeneity	Control	25	23.99	3.39	.77
	Experimental	26	24.35	3.66	.71

As can be seen, the difference between the control group (M = 23.99, SD = 3.39) and the experimental group (M = 24.35, SD = 3.66) is minimal.

Procedure

The main phase of the study began after getting the participants' consent, homogenizing, and grouping of the participants. The study lasted for 16 sessions over a period of 1.5 months. Table 3 reveals the research schedule.

TABLE 3

Research Schedule

Session	Control Group	Experimental Group	Duration
1	Homogeneity Test+ Pre-test	Homogeneity Test+ Pre-test	75 minutes
2-15	Conventional education	Intervention	45 minutes
16	Post-test (listening+ memory)	Post-test	60 minutes

In the non-computerized dynamic group (Non-C-DA, the control group), the participants were asked to listen to a TOFEL iBT listening test every session and answer the test questions. The test takers were given around 35 listening questions in each session. The time allocated to answer the questions was around 45 to 50 minutes. The listening tasks were in the form of lectures, discussions, and conversations. This process was similar to the conventional approach to answering listening questions and did not entail any particular treatment.

As for the computerized dynamic assessment (C-DA, the experimental group), the researcher opted for the interventionist approach (sandwich format, as suggested by Grigorenko, Sternberg, & Ehrman, 2000), to be used in the study. Based on this model, the participants took the pre-test in the form of a traditional static test (the iBT) in which no mediation was provided. In the intervention stage, computerized mediation was conducted for the students through the computerized dynamic assessment (C-DA) approach.

In the intervention phase, the course instructor (not the researcher) had mediated conversations with the participants and asked about their difficulties with answering the TOEFL iBT listening questions. Suggestions were given to the participants based on their problems. After determining the problems and giving solutions, the teacher played an audio file for the participants and asked them to listen to the file. The listening questions were not presented to the participants at this point; however, they were asked to bear in mind the main points stated by the speakers in the conversation. The participants were guided on what information was deemed necessary by the teacher before listening to the audio file. After listening to the audio files, the participants were given the questionnaires and were asked to answer the listening questions. This method of increasing and measuring working memory span is adapted from Conway et al. (2005). Working memory span of the learners was also gauged using the Conway, Cowan, Bunting, Theriault, and Minkoff's (2002) procedure. The intervention phase lasted for 14 sessions in the experimental group.

The main differences between the control group and experimental group procedure were as follows:

- a. The control group participants were given the TOEFL iBT test containing the listening questions as they listened to the audio files, whereas in the experimental group the participants were given the questions after they listened to the audio files.

- b. Implementation of dynamic assessment in the experimental group was based on Conway et al.'s (2005) model. The current listening skill of the participants was checked first, the listening problems were identified and worked on in groups through question-and-answer discussions, and the performance was tested again through a posttest.

After the intervention phase of the study was completed, the post-test was administered to all participants in both groups. Therefore, the researcher could compare the learners' performance before and after the mediation phase. The post-test was an iBT listening test. The participants' memory span was also gauged based on the administered listening test. To measure the memory span, the procedure suggested by Karpicke and Pisoni (2000) was used. They believe working memory can be measured according to the number of words a person can remember. To measure the participants' memory span, at the posttest, they were asked to write down the number of facts they remembered about the listening audio file. The same test was given to the participants in the control group.

Results

To carry out the study, first descriptive statistics of the homogeneity test and the groups' post-test scores and tests of normality assumptions were computed.

TABLE 4
Descriptive Statistics; Normality of Distribution

	Homogeneity Test	Post-test Control Listening	Post-test Experimental Listening	Post-test Control Memory	Post-test Experimental Memory
N	101	25	26	25	26
Std. Deviation	3.54	2.27	4.23	2.74	4.82
Skewness	.07	.42	.24	.52	.40
S.E. of Skewness	.28	.54	.54	.25	.25
Kurtosis	-.64	.89	-.29	.78	.32
S.E. of Kurtosis	.41	.80	.80	.71	.71

Normality of distribution of scores was investigated through ratios of skewness and kurtosis at the initial phase of the study. Based on George and Mallery (2003), when distribution of data is ± 1.96 , normal distribution of scores can be assumed. In the next phase of data analysis and prior to the main analysis, the reliability of the homogeneity test was assessed.

TABLE 5
Descriptive Statistics; OPT Test

	N	Mean	Std. Deviation	Variance
Nelson	101	24.17	3.54	9.11
KR-21	.78			

The reliability index for the homogeneity test was .78; thus, the consistency of scores was approved. Table 6 shows the reliability of the post-test of listening.

TABLE 6
Reliability of Post-test of listening by KR-21

	N	Mean	Std. Deviation	Variance
Pretest	51	18.15	2.116	4.321
KR-21	.78			
Posttest	51	25.11	2.764	6.978
KR-21	.80			

To determine the homogeneity of the groups at the outset of the study, an independent samples t-test was run (Table 7).

TABLE 7
Independent Samples t-test: Pre-test of Experimental vs Control Groups

	Levene's Test for Equality of Variances				t-test for Equality of Means				
	F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.87	.43	-.41	49	.68	-.36	1.01	-2.51	1.75

The results of independent samples ($t_{(49)} = .41, p = .68 \geq \alpha = .005$) [-2.51, 1.75] show that the two groups are homogenous.

Results to Respond to the First Research Question

The answer to the first question was investigated by comparing the participants' performance in the experimental group (C-DA group) with that of the control group (Non-C-DA group) in the listening test. Table 8 shows the descriptive statistics of the two groups' listening scores in TOEFL iBT.

TABLE 8
Descriptive Statistics for iBT Listening Test: Experimental Groups vs. Control Groups

Groups	N	Mean	Std. Deviation	Std. Error Mean
Experimental	26	24.83	4.71	.68
Control	25	20.63	4.01	.67

As observed in Table 8, the experimental group has a higher mean ($M = 24.83, SD = 4.71$) than the control group ($M = 20.63, SD = 4.01$). Table 9 presents the main results, comparing the groups in terms of their post-test listening scores.

TABLE 9
Independent Samples t-test; Post-test of Listening

	Levene's Test for Equality of Variances				t-test for Equality of Means				
	F	Sig.	t	df	Sig.	MD	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	9.28	.00	4.67	49	.00	4.20	.758	1.31	4.73

The results of independent samples t-test were $t_{(49)} = 4.67, p = .000 \leq \alpha = .05$ [1.31; 4.73], which indicates that the difference between the control group and the experimental group is significant. Therefore, the first null hypothesis was rejected, and it can be assumed that using C-DA significantly affected Iranian EFL learners' listening comprehension.

Results to Respond to the Second and Third Research Questions

To answer research questions 2 and 3, examining if C-DA has affected the working memory span of the learners and if there has been a change in the listening comprehension scores of Iranian EFL learners as a result of the interactional effect of C-DA and working memory, a factorial ANOVA was conducted.

Prior to the analysis, as observed in Table 9, the assumption of homogeneity of variance was met. To do the analysis, the two post-tests were considered for each group, i.e., post-test of listening comprehension and post-test of memory span. Considering the results of Table 10, Levene's test ($F(3, 98) = .60, p = .71$) indicated that there were no significant differences between the groups' variances.

TABLE 10
Levene's Test of Equality of Error Variances

F	df1	df2	Sig.
.60	3	98	.71

Next, descriptive statistics of the tests were taken into account. As observed in Table 11, the experimental group ($M = 29.94, SE = .41$) had a higher post-test score compared to the control groups ($M = 26.74, SE = .41$).

TABLE 11
Descriptive Statistics; Post-test of Memory Span by Types of Treatment

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control Groups	21.63	.31	20.74	23.15
Experimental Groups	25.82	.31	22.85	28.63

Factorial ANOVA results, i.e., ($F(1, 98) = 34.17, p = .001$, partial $\eta^2 = .33$, representing a large effect size) (Table 12) reveal that there was a statistically significant difference between the post-test scores of the two groups.

TABLE 12
Tests of Between-subjects Effects; Post-test of Listening and Memory Span

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Group	302.81	1	241.71	234.17	.00	.33
Variable	3.71	1	3.71	.59	.60	.01
Group * Variable	.66	1	.66	.16	.83	.00
Error	611.28	98	11.03			
Total	58364.01	102				

As displayed in Table 13, the experimental group ($M = 28.12, SE = .39$) and control group ($M = 27.72, SE = .40$) groups had, more or less, similar means on the post-test of memory span.

TABLE 13
Post-test of Listening by Memory Span

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control	27.72	.40	25.17	29.19
Experimental	28.12	.39	26.90	31.00

Table 12, ($F(1, 98) = .59, p = .60$, partial $\eta^2 = .01$) shows a weak effect size. It means that the difference between the post-test score of the experimental group and the control group of memory span was not significant; thus, the second null-hypothesis was supported, and it can be concluded that C-DA did not affect the participants' memory span.

As for the third research question, based on Table 12 ($F(1, 98) = .16, p = .83$, partial $\eta^2 = .00$ representing a weak effect size), it is concluded that the third null-hypothesis was supported and the interaction of C-DA and memory span did not significantly affect these Iranian EFL learners' listening comprehension.

Discussion

In general, it is believed that C-DA is a helpful approach to language learning and assessment (Poehner, Zhang, & Lu, 2015). In line with this general belief, in this study, it was found that C-DA positively affected the listening comprehension of the participating Iranian EFL learners. As DA, and C-DA have common roots and differ only in terms of utilization of computers in the implementation of DA, the results of studies dealing with these two approaches are used in this section. More specifically, Ableeva (2008) dealt with the effect of DA on the listening comprehension of language learners. In his study with French students, he found out that DA has led to improvement among the students learning French, as interactions with learners and teachers can help teachers diagnose the learners' listening problems. Similarly, in this study, the use of this approach enabled the teacher to initiate dialogues with the learners and diagnose their problems.

Ableeva (2008) and Poehner and Lantolf (2013) believed that the role of the Zone of Proximal Development (ZPD) was missing in listening comprehension in the past, and DA has the advantage of benefiting from others' proximity. However, scholars such as Cheng (2004) note that due to the one-way nature of listening assessment, e.g., in multiple-choice items, it is hard to implement DA in listening classes. Not only was it observed in the current study, but also as stated by Gilbert (2012), the answer to this problem lies in the nature of DA. While implementing DA, the type of assessment may not differ from traditional assessment; however, the purposes and interpretations of the tests are different. Before DA was suggested as an alternative to traditional assessment, formative assessment was used to tackle problems with summative assessment (Taras, 2005). However, formative assessment, per se, could not meet all the needs of all learners who were looking for ways to solve their listening problems. DA uses formative assessment as a stage in process-oriented assessment (Black & Wiliam, 2009), although it seeks more from the learners through dialogue. This was among the main concerns in this study, as the researchers put particular emphasis on the dialogues with the learners.

As for the effect of DA on memory span, Swanson (1994) noted that many educators and even learners are doubtful about the effects of DA on working memory span. However, Swanson (1994) did not deny the fact that DA can positively affect the cognitive abilities of the learners. More recent studies by Tzuril (2001) urged the need for not only new procedures, e.g., DA, but also new testing materials. She believed that DA is not applicable to old materials, and that materials should be in accordance with the new objectives. This consideration seems to be significant, as in the currently study the interaction of C-DA and memory span did not show a significant effect on listening comprehension of the learners.

Shabani (2012), who investigated the effect of C-DA on reading comprehension of language learners, contends that the interactive aspect of C-DA is among the most significant dimensions of this method, and should not be taken for granted. This issue was highlighted in the current study through the utilization of C-DA and initiating conversations with the participants based on their mistakes. The researchers found that such interactions would assist the participants when diagnosing the depth of their problems and correcting them accordingly.

Conclusion

Based on the results, while C-DA has had a significant positive effect on the listening comprehension of these Iranian EFL learners, it did not significantly affect the working memory span of the learners, and the interactional effect of listening comprehension and the participants' working memory span was negligible. This study has some pedagogical implications for language teachers who may be interested in using C-DA in their classes. Future researchers can also focus on the effects of C-DA on other aspects of language learning such as productive language skill.

The Authors

Masoomeh Estaji (corresponding author) is an associate professor of Applied Linguistics at Allameh Tabataba'i University (ATU), Tehran, Iran. She holds a Ph.D. in Applied Linguistics from Allameh Tabataba'i University. She has presented and published numerous papers on methodology, testing, and second language acquisition (SLA). Her research interests include language testing and assessment, teacher education, and ESP.

Department of English Language and Literature
Allameh Tabataba'i University
Faculty of Persian Literature and Foreign Languages, South Allameh St, Modiriati Fly-over, Chamran Highway, Tehran, Iran
Tel: +9821 88692346-9 Ext: 522
Email: mestaji74@gmail.com

Fatemeh Safari is a Ph.D. candidate of TEFL at Allameh Tabataba'i University (ATU), Tehran, Iran. Her areas of interest are language testing and assessment, educational psychology, and teacher education.

Email: fatemeh.safari.1989@gmail.com

References

- Ableeva, R. (2008). The effects of dynamic assessment on L2 listening comprehension. In J. P. Lantolf & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 57-86). London: Equinox Press.
- Atasheneh, N., & Izadi, A. (2012). The role of teachers in reducing/increasing listening comprehension test anxiety: A case of Iranian EFL learners. *English Language Teaching*, 5(3), 178-194. <https://doi.org/10.5539/elt.v5n3p178>
- Azarian, F., Nourdad, N., & Nouri, N. (2016). The effect of dynamic assessment on elementary EFL learners' overall language attainment. *Theory and Practice in Language Studies*, 6(1), 203-216. <https://doi.org/10.17507/tpls.0601.27>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5-25. <https://doi.org/10.1007/s11092-008-9068-5>
- Butler, S. M., & McMunn, N. D. (2006). *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning*. Indianapolis, USA: Jossey-Bass, an Imprint of Wiley.
- Cheng, H. F. (2004). A Comparison of multiple-choice and open-ended response formats for the assessment of listening proficiency in English. *Foreign Language Annals*, 37(4), 544-553. <https://doi.org/10.1111/j.1944-9720.2004.tb02421.x>
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-183. [https://doi.org/10.1016/s0160-2896\(01\)00096-4](https://doi.org/10.1016/s0160-2896(01)00096-4)
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769-786. <https://doi.org/10.3758/bf03196772>
- Dochy, F. (2001). A new assessment era: Different needs, new challenges. *Learning and Instruction*, 10(1), 11-20. [https://doi.org/10.1016/s0959-4752\(00\)00022-0](https://doi.org/10.1016/s0959-4752(00)00022-0)

- 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* (Doctoral dissertation). University of San Francisco, USA.
- Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55-75. [https://doi.org/10.1016/s0346-251x\(99\)00060-3](https://doi.org/10.1016/s0346-251x(99)00060-3)
- George, D., & Mallery, M. (2003). *Using SPSS for Windows step by step: A simple guide and reference*. Boston, USA: Allyn and Bacon.
- Gilbert, J. B. (2012). *Clear speech teacher's resource and assessment book: Pronunciation and listening comprehension in North American English*. England: Cambridge University Press.
- Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory-based approach to the measurement of foreign language learning ability: The CANAL-FT theory and test. *The Modern Language Journal*, 84(3), 390-405. <https://doi.org/10.1111/0026-7902.00076>
- Karpicke, J., & Pisoni, D. B. (2000). Memory span and sequence learning using multimodal stimulus patterns: Preliminary findings in normal-hearing adults. *Research on Spoken Language Processing*. Progress Report No. 24. Bloomington, IN: Speech Research Laboratory, pp. 393-406.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics*, 1(1), 49-72. <https://doi.org/10.1558/japl.1.1.49.55872>
- Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology*, 130(2), 199-207. <https://doi.org/10.1037/0096-3445.130.2.199>
- 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>
- 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>
- Shabani, K. (2012, February). *Computerized dynamic assessment: An interventionist procedure to assess L2 reading*. Paper presented at the Third International Conference on E-Learning and E-Teaching (ICELET) (pp. 15-24). IEEE.
- Swanson, H. L. (1994). The role of working memory and dynamic assessment in the classification of children with learning disabilities. *Learning Disabilities Research & Practice*, 4(4), 188-199.
- Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466-478. <https://doi.org/10.1111/j.1467-8527.2005.00307.x>
- Tzuriel, D. (2001). *Dynamic assessment of young children*. New York: Kluwer Academic/Plenum Press.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40(03), 191-210. <https://doi.org/10.1017/s0261444807004338>
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.

(Received August 14, 2020; Revised August 26, 2020; Accepted September 12, 2020)