Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

The Efficacy of Completing Form-focused Tasks Collaboratively vs Individually: Utilizing Interventionist Dynamic Assessment to Quantify Learning Gains

James Scotland
Foundation Program Department of English
Qatar University, Qatar
Email: scotland@qu.edu.qa

Abstract
Many language teachers employ collaborative learning within their classrooms. However, expectations surrounding the efficacy of working collaboratively need to be empirically verified. This study employed dynamic assessment to investigate whether learners who are situated within an undergraduate Qatari EFL context learn second language grammatical structures more effectively by working either collaboratively with their peers or individually. Interventionist dynamic assessment was used to quantify the extent of the learning gains made by male Arabic undergraduate EFL learners (N = 52) three times (pretest, posttest, and delayed posttest) over a 12-week period. In between the pretest and the posttest, six form-focused treatment tasks were administered. The experimental group (n = 20) completed the tasks collaboratively with their peers; a comparison group (n = 16) completed the tasks individually; and a control group (n = 16) did not complete the tasks. The target structures were the simple past passive and the present continuous passive. A Mood's median test (Mood, 1954) found no statistically significant differences between the collaborative condition and the individual condition. Although measuring emergent abilities which are still in the process of developing provided a more complete picture of the efficacy of working collaboratively, the lack of a statistically significant difference between the performances of the experimental and comparison groups for both target structures suggests that working collaboratively is not statistically more effective in facilitating learners' linguistic development than working individually.

Keywords: collaborative learning, focus on form, individual learning, interventionist dynamic assessment

Introduction

Collaborative learning, a set of instructional practices in which students work together to help each other to learn academic content, has been advocated by second language acquisition (SLA) practitioners for decades (Long & Porter, 1985; Pica & Doughty, 1985). For the present study, collaborative learning is defined as "(…) a coordinated synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem" (Roschelle & Teasley, 1995, p. 70). An example of a collaborative task is a collaboratively completed dictogloss task (Wajnryb, 1990). Numerous studies have documented how working collaboratively provides L2 learners with cognitive spaces in which they can pool their linguistic and cognitive resources in order to adjust, refine, and develop their linguistic accuracy and communicative competence, as well as collaboratively perform at a level higher than that at which they could be expected to perform individually (e.g., Anton & DiCamilla, 1999; Donato, 1994; Fernández Dobao, 2012, 2014a, 2014b; Gánem-Gutiérrez, 2008; Storch, 2005; Storch & Wigglesworth, 2007; Swain, 2001; Swain & Lapkin, 1998). Participating in such collaborative spaces is then expected to result in more accurate language use in subsequent individual performance.

The term dynamic assessment describes a range of assessment approaches and uses that are rooted in Vygotskian sociocultural theory. Vygotsky (1978) theorized that although different learners may perform at a similar level, their underlying ability may be vastly different. Focusing exclusively on abilities that have already developed at the time of testing reveals nothing about emergent abilities which are still in the process of developing. Therefore, the purpose of psychological assessment from a Vygotskian perspective is to promote as well as gain an understanding into abilities which have been fully internalized as well as abilities which are still in the process of developing (Poehner, 2008). This belief is reflected in the concept of dynamic assessment. During dynamic assessment, the assessor provides intentional support (i.e., mediation) when the learner encounters difficulties and documents the learner's responsiveness. Analysing a learner's responsiveness to this support can provide a more complete understanding of abilities that are not yet fully developed. As it is able to access emergent abilities, dynamic assessment appears to offer a more nuanced way of assessing a learner's linguistic development than more traditional non-dynamic measures of assessment.

The present study seeks to investigate to what *extent* working collaboratively to complete form-focused tasks impacts on Qatari undergraduate learners' longer-term performance of two complex grammatical structures, the simple past passive and the present continuous passive. The following research question was asked. To what *extent* does working collaboratively to complete form-focused tasks impact on Qatari undergraduate learners' longer-term performance of two complex grammatical structures? Currently, no previous study has employed dynamic assessment to investigate the efficacy of collaboratively learning L2 form in the SLA classroom. A unique aspect of this study is its use of dynamic assessment to assess the impact that working collaboratively has on learner abilities which are not fully formed but are still in the process of developing. Thus, a more complete picture of the effectiveness of learning L2 form collaboratively may be obtained. First, the body of research which has previously investigated the efficacy of collaboratively learning L2 form is reviewed. Then, the principles and procedures of dynamic assessment are explained. Finally, the efficacy of learning L2 form collaboratively is investigated.

Literature Review
Learning L2 Form Collaboratively

Sociocultural theories of learning offer a rational for learning collaboratively. A sociocultural perspective assumes that during collaborative tasks learners get collective help and guided support as a result of interacting with each other (Lantolf & Thorne, 2006). In other words, learners are able to mutually support each other's performance. How working collaboratively allows L2 learners to both provide and receive guided support can be explained through the concept of collaborative dialogue. Swain (2000) defines collaborative dialogue as "dialogue in which speakers engage jointly in problem solving and knowledge building" (p. 102). When learners jointly problem solve and knowledge build, they may overtly use language as a psychological cognitive tool in order to organize and mediate each other's linguistic performance. Swain (2000) explains that collaborative dialogue is an enactment of mental processes. "Through speaking, thought is externalize d. Externalized as an utterance, it becomes an object. As an object it can be scrutinized, questioned, reflected upon, disagreed with, changed, or disregarded" (Swain & Lapkin, 2002, p. 286). Working collaboratively allows learners to both provide and receive peer mediation, providing opportunities for all task participants to restructure and develop features of their linguistic knowledge and L2 production. However, expectations of the efficacy of working collaboratively need to be empirically verified.

To my knowledge, there has been no statistical meta-analysis of the collaborative learning of L2 form. However, Elabdail (2021) examined the claim that collaborative writing is beneficial for L2 learning by synthesising the results of 33 studies. Elabdail (2021) found that collaboratively written texts tended to be more accurate than individually written texts with a medium effect size. Elabdail (2021) suggests that her findings support the position that collaborative writing promotes accuracy as it enables learners to pool expertise and engage in collaborative scaffolding (Donato, 1994).

Several studies have investigated the efficacy of collaboratively learning L2 form (i.e., Adams, 2007; Adams, Nuevo, & Egi 2011; Eckerth, 2008; Kuiken & Vedder, 2002; Lapkin, Swain, & Smith, 2002; McNichol & Lee, 2011; Nassaji & Tian, 2010; Reinders, 2009; Spielman-Davidson, 2000; Swain & Lapkin, 2001; Teng, 2020). These studies have utilized a wide variety of designs, been implemented in a wide variety of contexts, and have findings which differ (see Scotland, 2021).

Although some of these studies found statistical significance within their results (Adams et al., 2011; Eckerth, 2008; McNicoll & Lee, 2011; Spielman-Davidson, 2000; Teng, 2020), many studies contain results which are not statistically significant (Adams et al., 2011; Kuiken & Vedder, 2002; McNicoll & Lee, 2011; Nassaji & Tian, 2010; Reinders, 2009; Spielman-Davidson, 2000; Swain & Lapkin, 2001). The concurrent statistical significance and non-significance reported by some studies can be partly attributed to the implementation of complex designs (e.g., the use of more than one target structure, data collection instrumentation, and/or test type).

Two of the above studies are worth further exploration as, similar to the present study, they determined the effectiveness of learning L2 form collaboratively by measuring changes in learners' knowledge of the passive voice. Firstly, Kuiken and Vedder (2002) investigated the

effect of interaction between learners during a dictogloss task and the acquisition of three different classifications of the passive form. The structures targeted were verbal constructions with one auxiliary (e.g., *was owned*), two auxiliaries (e.g., *had been stolen*), and three auxiliaries (e.g., *may have been presented*). 16-18-year-old Dutch high school students completed two dictogloss tasks, either individually (n = 14) or in groups of three or four (n = 20). An analysis of covariance (ANCOVA) found no statistically significant differences between the results of the two groups. Kuiken and Vedder (2002) concluded that "these findings do not support the hypothesis that giving learners the opportunity to interact with each other during a dictogloss task will result in a better score on the post-test or on the delayed post-test" (p. 349).

Secondly, Eckerth (2008) explored the extent to which working collaboratively resulted in the learning of the passive voice, reflexive prepositional verbs, and transitive prepositional verbs. Eckerth (2008) did not specify which passive constructions were targeted. 14 adult lower intermediate and 17 upper intermediate learners of German (N = 31) completed five treatment/test cycles. In each cycle, learners collaboratively completed either a dictogloss task or a text-repair task. A repeated measures two-factorial ANOVA found a statistically significant difference with a large effect size between the results of the pretests and posttests, and the results of the pretests and delayed posttests. Eckerth (2008) concluded that "learners are able to provide each other with feedback rich in acquisitional potential" (p. 133).

Overall, the results of previous studies suggest that collaboratively completing form-focused tasks can, but does not always, result in statistically significant gains in the learning of L2 form. However, all of these studies used non-dynamic testing techniques.

Dynamic Assessment
Dynamic assessment originates from Vygotskian sociocultural theory. Vygotskian sociocultural theory posits that consciousness is derived from the social world (Vygotsky, 1978). Vygotsky believed that knowledge and abilities that once resided in an individual's social interactions on the interpsychological plane (i.e., between people) become internalized and can be used as a resource for new cognitive functions on the intrapsychological plane (i.e., inside a person) (Vygotsky, 1978). This internalization occurs through mediation (Davin & Gómez-Pereira, 2019) with mediation being "the process through which humans deploy culturally constructed artifacts, concepts, and activities to regulate (i.e., gain voluntary control over and transform) the material world of their own and each other's mental activity" (Lantolf & Thorne, 2006, p. 79). Vygotsky makes an important distinction between "unassisted performance, as was and is typically required on static educational and psychological measures, and student performance when instructed by a more knowledgeable other" (Dumas, McNeish, & Greene, 2020, p. 90). The difference between a student's unassisted performance and their assisted performance on a task was termed the Zone of Proximal Development (ZPD) (Dumas et al., 2020). Vygotsky defines the ZPD as, "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). Conceptually based on the ZPD, dynamic assessment is the unification of assessment and instruction into a singular, dialogue-based activity (McNeil, 2018).

Dynamic assessment attempts to understand the scope of a learner's abilities through the promotion of their development (Poehner, 2007). To accurately understand a learner's abilities, dynamic assessment embeds instruction and feedback into the assessment procedure. Though the use of questions, hints, prompts, suggestions, or explanations, an assessor may "guide learners in highlighting important content, making connections, setting goals, planning, regulating and controlling behavior" (Antón, 2009, p. 579). Interactions between the assessor and the learner are not haphazard; they are carefully attuned (i.e., mediated) to the learner's current abilities and provided incrementally. In addition, the assessor focuses "on helping the learner advance their conceptual understanding" (Antón, 2019, p. 119). If a learner experiences difficulty during administration of the dynamic assessment, then "the mediator responds to learners' discourse by adjusting intervention to what is needed in each individual case in order to complete the task and show the full potential of the learners' ability" (Antón, 2009, p. 592). Dynamic assessment can be used to "measure language abilities, intervene in learning, and document learners' growth" (Antón, 2009, p. 576). The two kinds of mediation that researchers can employ when administering dynamic assessment are interventionist and interactionist.

Interactionist dynamic assessment employs mediation which is flexible, open-ended, and emerges from the interactions between the assessor and the learner(s). In interventionist dynamic assessment, mediation is dependent upon the quality of the interaction between the assessor and the learner (Antón, 2009). Mediation is negotiated with a learner by continually being adjusted according to the learner's developmental needs (Lantolf, 2009). Interactionist dynamic assessment is focused on gaining an insight into the kinds of psychological process that the learner might be capable of in the next phase of development and identifying the kinds of instruction, or assistance that will be required if the learner is to realize this potential (Minick, 1987). Examples of interactionist dynamic assessment include implementing a language enrichment program with advanced learners of L2 French (Poehner, 2008), and devising individualized instructional plans for third-year Spanish majors (Antón, 2009). Because this approach it is labour-intensive, non-standardized, and time-consuming, it is more suited for generating detailed qualitative data on individuals or a relatively small number of learners.

Interventionist dynamic assessment relies on standardized administration procedures. Assessors use prescribed clues which are carefully arranged on a scale from implicit (e.g., pausing to indicate a problem) to explicit (e.g., explaining a concept) and are usually assigned a numerical value. This list of pre-scripted prompts is often referred to as a regulatory scale (Rassaei, 2019). Starting with the most appropriate implicit clue, the assessor administers the prescribed clues, until either the learner answers correctly or the final clue is reached. The point at which a learner is able to make use of the assistance provided indicates to the assessor the learner's emergent abilities and supports their development (Lantolf & Poehner, 2011). As the number of predetermined hints that interventionist dynamic assessment uses is fixed and standardized, comparable numerical scores can be generated for each participant (Lantolf, 2009). Examples of interventionist dynamic assessment include promoting L2 development and tracing learner independent functioning post mediation (Lantolf & Poehner, 2011), and diagnosing learner abilities and areas for further instruction (Poehner, Zhang, & Lu, 2015). Interventionist dynamic assessment is often conducted with large numbers of participants to produce quantitative data which can be compared at the group level.

Regulatory Scales and the Learning of L2 Form

A regulatory scale is a graduated continuum which starts with the provision of broad and implicit assistance and progresses to providing more focused and explicit assistance. The use of a regulatory scale allows researchers to provide feedback in a predetermined and structured way. Regulatory scales have been utilized by researchers for decades; examples include: assessing learners' learning potential (Budoff & Friedman, 1964), assessing the learning and transfer of inductive reasoning (Ferrara, Brown, & Campione, 1986), and testing language aptitude (Guthke & Beckmann, 2000).

In the field of SLA, perhaps the most influential study involving a regulatory scale was carried out by Aljaafreh and Lantolf (1994). Although not specifically framed as dynamic assessment, Aljaafreh and Lantolf (1994) investigated the relationship between error correction and language learning. Three intermediate ESL learners volunteered for eight extra tutorial sessions that focused on reoccurring grammatical problems (i.e., articles, tense marking, prepositions, and modal verbs) within their writing. The researcher in this study endeavoured to diagnose areas of difficulty and to help learners gain control over the relevant structures. The mediation between the researcher and the learner emerged spontaneously from their interaction. An analysis of the tutorial sessions led to the creation of a 13-point regulatory scale (Aljaafreh & Lantolf, 1994) that captures relative degrees of explicitness. This regulatory scale was employed when analysing the interactions that took place. Aljaafreh and Lantolf (1994) were able to show how independent linguistic production alone does not provide a full picture of L2 learning; it is also essential to acknowledge the existence of, and attempt to understand emergent knowledge.

Two SLA studies have applied a regulatory scale to the learning of grammar. Firstly, van Compernolle and Zhang (2014) employed interventionist dynamic assessment to investigate one learner's implicit grammatical competence through elicited imitation. Elicited imitation requires learners to repeat strings of words, phrases, or sentences. van Compernolle and Zhang (2014) targeted three word-final morphological features (i.e., plural –s, past tense –ed, and third-person singular –s). The learner was required to listen to a statement, identify whether the statement is true or false, and then repeat the statement. The learner had up to four attempts to correctly repeat the statement. If the learner was not successful, graduated mediation as provided, ranging from more implicit (i.e., listening to the statement again) to more explicit (i.e., provision for the correct form and explanation). The less assistance the learner required, the higher they scored for each statement. The results were used to calculate a learning potential score and to create a detailed diagnostic profile of the learner's current and emerging abilities.

Secondly, Mohammadimoghadam (2015) employed interventionist dynamic assessment to investigate the effects of mediation on one EFL learner's development of tag questions. After an initial pretest, a mediation phase was administered in which the learner answered questions pertaining to the target structure. Feedback was provided in the form of interventionist dynamic assessment, which was guided by an eight-point regulatory scale. Each point on the scale was allocated a score. Scores ranged from seven for the most implicit feedback (i.e., a pause) to zero for the most explicit feedback (i.e., identification and explanation of the correct answer). These scores were then compared to the pretest results and used to calculate a learning potential score. Mohammadimoghadam (2015) concluded that participating in the dynamic assessment procedure improved the learner's performance of English tag questions. This conclusion is

expected. As well as revealing the depth of a learner's abilities, dynamic assessment also helps learners to realize their future by acting as a catalyst for development (Poehner, 2007). In summary, both van Compernolle and Zhang (2014) and Mohammadimoghadam (2015) utilized a regulatory scale within the framework of interventionist dynamic assessment in order to gain insights into learners' emergent grammatical abilities.

Previous studies which have attempted to measure the efficacy of collaboratively learning L2 form have only attempted to measure linguistic knowledge that was already relatively developed at the time of testing through the use of non-dynamic testing. These studies did not attempt to take into account the impact that working collaboratively may have on learner abilities which are still in the process of developing. Thus, the treatment activities used in these studies may have been more effective than initially thought.

The present study seeks to provide a deeper understanding on the efficacy of learning L2 form collaboratively. The following research question was asked. To what *extent* does working collaboratively to complete form-focused tasks impact on Qatari undergraduate learners' longer-term performance of two complex grammatical structures? To quantify emergent learning gains, a regulatory scale was employed within the framework of interventionist dynamic assessment.

Methods
The methodology presented in this article is taken from a larger study (see Scotland, 2017); what follows is the methodology relevant to the data presented.

Participants
Six classes of male students were invited to participate in the study. The participants were recruited over two semesters. Three parallel classes participated in the first semester; three parallel classes participated in the second semester. Convenient sampling was employed. All classes were taught by the teacher who agreed to help with the study. Thus, the participants were students who happened to find themselves in one of these classes. Initially, 105 students volunteered to participate in this study; however, this number was later reduced to 52 participants (see Procedures). All participants were between 17-44 years old, nationals of Arabic speaking countries, and shared Arabic as their L1. Approval to carry out the study was given by the institute's review board. Participation in the study was voluntary. Written informed consent was given by all participants.

This study is situated within a B1+ level (Council of Europe, 2001) general proficiency English course which is part of the core curriculum of a Qatari institute of higher education. This student-centred course provides learners with an opportunity to develop their skills in reading, writing, speaking, and listening. Readings and listening are taken from a diverse range of texts. In addition, there is a focus on both semi-formal and formal writing skills. The teaching of vocabulary, grammar, and critical thinking skills is embedded into the curriculum. The course requires five contact hours per week over a 15-week period.

Research Instruments
A quasi-experimental pretest-treatment-posttest design was employed. Three groups were utilized. Due to the use of pre-existing classes, the sampling was convenient. An experimental

group completed the treatment tasks collaboratively; a comparison group completed the treatment tasks individually; and a control group did not complete the treatment tasks. All groups completed the pretest, posttest, and delayed posttest. Quantitative data was collected three times (i.e., pretest, posttest, and delayed posttest) using interventionist dynamic assessment over a 12-week period. Between the pretest and the posttest, six separate treatment tasks were administered (figure one).

The experiment was carried out twice over two consecutive semesters. The data generated was combined, producing one data set for each condition (i.e., learning collaboratively, learning individually, and the control). By examining group level changes in performance across tests, the present study attempted to determine to what extent completing the treatment tasks collaboratively impacted upon the participants' knowledge of the target structures.
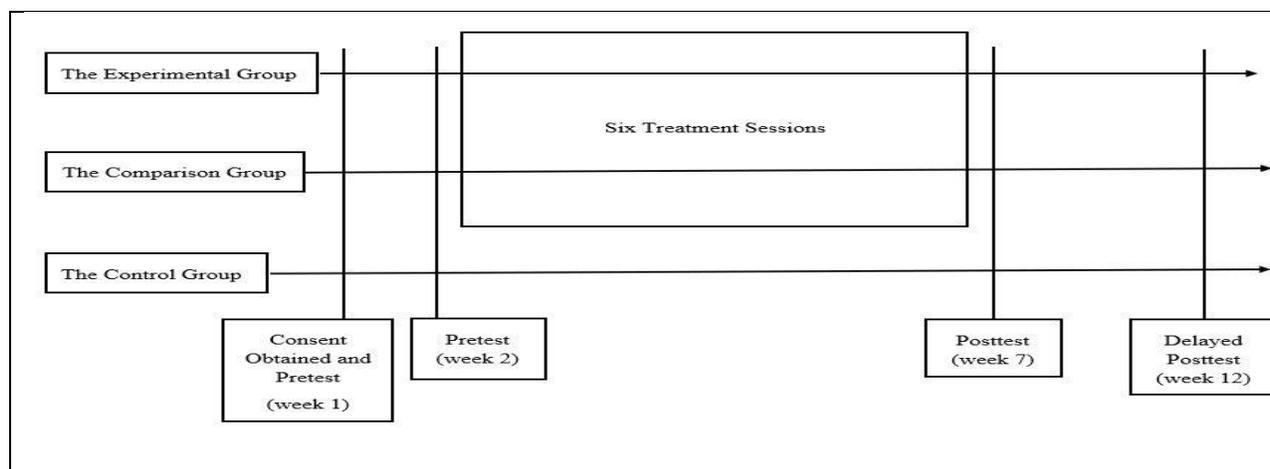
Figure 1. Design Overview

The simple past passive (e.g., *The question was answered by the student*) and the present continuous passive (e.g., *The question is being answered by the student*) were chosen as the target structures. These target structures were selected because they were part of the curriculum of the course in which this study is situated and complex enough to allow for a dynamic assessment procedure to detect incremental changes in the participants' performance of grammatical form.

Linguistic knowledge is operationalized as the ability to accurately write two predetermined structures of the passive voice at the sentence level with the aid of prescripted assistance (i.e., through the use of interventionist dynamic assessment). Linguistic development is operationalized as a reduction in the explicitness of assistance required during testing to accurately write these two predetermined structures.

Tests

The tests were designed around the principles of interventionist dynamic assessment. To create an obligatory context for the production of the target structures, each test item created a

scenario which required the participants to produce a predetermined target sentence (see figure two).

Figure 2. Example Test Item

Each test item was a sentence level written production task which required a constrained constructed response (Norris & Ortega, 2000). A stem sentence begins each target sentence. The main verb and the agent are supplied in parentheses. The participants were expected to use and modify the words in the parentheses as well as adding their own function words to complete the sentence. Participants were given four attempts to correctly write each target sentence. After each incorrect attempt, prescripted assistance was provided. For each target structure, a bank of test items was created.

Several principles guided the construction of each test item. Firstly, the answers that the participants needed to produce were standardized. Only regular verbs were used; all of the sentences required an agent; the verb and the agent in the stem sentence never shared the same root word; and no phrasal verbs or modal passives were used. Secondly, the composition of the paragraphs was controlled. Most words were within the first three thousand words of the British National Corpus and the Corpus of Contemporary American English (BNC/COCA); each paragraph contained four to seven sentences and between 35-53 words; all of the paragraphs had a minimum Flesch-Kincaid readability test score of 70; each target sentence was placed in the middle of its paragraph; and all sentences within a paragraph, except the target sentence, were in the active voice.

Due to time constraints, each test item was piloted non-dynamically. The facility index and the discrimination index of each test item was calculated. Six test items were removed from the test banks. A Cronbach alpha was calculated for all of the remaining items. The test bank of simple past passive test items scored 0.843 and consisted of 21 items; the test bank of present continuous passive test items scored 0.887 and consisted of 24 items. Due to the need to comply

with a time limit (see Procedures) when administering a test, each test contained two test items, one for the structure of the simple past passive and one for the structure of the present continuous passive. A randomization program was used to select the items for each test.

The Regulatory Scale

In order to quantify the explicitness of assistance required to accurately write a target structure, prescripted clues were created. These clues were adapted from Aljaafreh and Lantolf's (1994) thirteen-point regulatory scale and were organized into four categories of support (table one). The categories consisted of: indication of the presence of (a) problem(s), indication of the location of (a) problem(s), provision of specific information about (a) problem(s), and provision of the correct answer. Each category is based around the explicitness of the assistance required to write a target structure during testing. The categories were arranged from most implicit to most explicit and were each assigned a score (see table one).

If a participant writes the correct answer on their first attempt, then the participant is considered to be able to perform independently within the context of the test and receives a score of four. If a participant initially writes an incorrect answer, the assessor provides the appropriate clue within the first category of assistance; the participant is alerted to the existence of their mistake(s). In the next category of assistance, the participant is shown the location of their mistake(s). Then, the participant is given specific information about the nature of each mistake and alerted to its specific location. Finally, the correct answer is revealed. Due to the need to comply with a time limit (see Procedures) when administering a test, explanations of incorrect answers were not given, and linguistic terms such as past participle were not explained. Scores only represent the explicitness of the assistance provided to write a target structure not the amount of assistance provided or the content of the assistance provided.

Table 1. *The prescripted assistance*

Categories of assistance	Example of phrasing	Physical action of assessor	Score
Indication of the presence of (a) problem(s)	There is a mistake./There are mistakes.	None	3
Indication of the location of (a) problem(s)	There is a mistake here.	To indicate location of mistake(s)	2
Provision of specific information about (a) problem(s)	The sentence needs to be in the simple past/present continuous tense.	None	1
	This word needs to be in the past/present.	To indicate the word which is incorrect	
	The main verb should be a past participle.	To indicate the word which is incorrect	
	You need to use a different preposition.	To indicate the word which is incorrect	
	A/an be verb/ing be verb/past participle/preposition is missing here.	To indicate the location of the missing word	

	These words are in the wrong order.	To indicate which words are in the incorrect order	
	You have (an) extra word(s).	To indicate which word(s) are extra	
Provision of the correct answer	Here is the correct answer.	To show the correct answer	0

The Treatment Tasks

There were six treatment sessions. For each target structure, the following tasks were created: a guided learning task, a text-editing task, and a dictogloss task. See appendices A, B, and C for an example of each. Research has shown that guided learning can be more effective than didactic instruction (Alfieri et al., 2011). In addition, text-editing and dictogloss tasks have been used previously when investigating the efficacy of collaborative learning the passive voice (Eckerth, 2008; Kuiken & Vedder, 2002).

The construction of each treatment task was controlled. Excluding proper nouns, words depicted by accompanying pictures, or contemporary words that were very familiar to the participants, all words used for the text-editing and dictogloss task were within the first three thousand words of the British National Corpus and the Corpus of Contemporary American English (BNC/COCA). The paragraphs used for the text-editing task and the dictogloss task have a Flesch-Kincaid readability score of 70 or higher.

Research Procedures

Data collection followed the timeline in table two.

Table 2. *Timeline of events*

Group	Week 1	2	3	4
Control	demonstration of testing procedure, consent form, and pretest	pretest		
Comparison		pretest and guided learning task (simple past passive)	guided learning task (present continuous passive)	text-editing task (simple past passive)
Experimental				

Table 2 Continued

Group	Week 5	6	7	12
Control			posttest	delayed posttest
Comparison	text-editing task (present continuous passive)	dictogloss task (simple past passive)	dictogloss task (present continuous passive) and posttest	text-editing task (simple past passive)
Experimental				

Evidence exists that interleaved learning conditions are more effective than blocked learning conditions (Carpenter & Mueller, 2013; Rohrer, 2012). Therefore, the sequence of the treatment sessions alternated between target structures. Although the target structures are part of the course in which this study took place, the official course content associated with them was not delivered to all participants until just after the delayed posttest.

Testing took place during class time. On a day of testing, the class teacher proceeded with the scheduled lesson. Each participant then stepped out of the classroom and was tested. In order to cause as little disruption as possible to the academic lives of the participants, a five-minute time limit for testing was imposed. When being tested, each participant was told the following, '*Read the paragraph. Use these words to write the missing sentence*'. If required, the assessor gave additional instruction on an ad hoc basis on how to complete a test. No time limit for writing the first sentence was imposed; however, for each subsequent sentence a time limit of 30 seconds was imposed. If a sentence was not complete after this time, then the assessor administered the appropriate clues within the next category of assistance. If requested, a clue was repeated. Apart from the initial sentence, the participants did not have to write in complete sentences; they could alter existing sentences by writing replacement words underneath. Participants who were not able to complete a test within the prescribed five-minute time limit were still provided with the correct answer(s). Due to the time constraints of administering tests during scheduled lessons, testing took place over several consecutive days.

The treatment sessions occurred during class time and were administered by the participants' regular class teacher in a whole class setting. The comparison group completed the treatment tasks individually. The experimental group completed the treatment sessions in self-selected groups of two or three. Learner roles within the experimental group (e.g., scribe) were not prescribed and no L1 restrictions were imposed. Post-task feedback was given by the teacher in a whole class setting.

In an attempt to standardize opportunities for learning for all participants, it was desirable that each participant complete all tests and all treatment tasks. Fifty-three participants were absent for either one or more tests and/or one or more treatment sessions, and were thus excluded from the present study. In total, there were 52 complete data sets. These data sets were distributed as follows: control group ($n = 16$), comparison group ($n = 16$), and experimental group ($n = 20$).

Results
Each participant's score was quantified based on the explicitness of the clues required to complete a test item (see table one). The scores ranged from zero to four. If on a subsequent test participants showed a reduction in the explicitness of assistance required to produce a target structure and thus received a higher score, then linguistic development is considered to have taken place. The test score data are ordinal in nature; thus, the most appropriate measure of central tendency is the median (Cohen, Manion, & Morrison, 2007).

Descriptive Statistics
Table three provides the median and range scores. Descriptive differences exist between the results of the target structures. Nearly all median scores for the simple past passive are higher than their equivalent score for the present continuous passive.

Table 3. *Median and range scores*

| Group | Simple past passive | | | | | |
| | Pretest | | Posttest | | Delayed Posttest | |
	Mdn	Range	Mdn	Range	Mdn	Range
Control	3	4	3	4	3	4

(n = 16)						
Comparison (n = 16)	2	4	3	4	3	4
Study (n = 20)	2	4	3	3	3	2

Table 3 Continued

Group	Present continuous passive					
	Pretest		Posttest		Delayed Posttest	
	Mdn	*Range*	*Mdn*	*Range*	*Mdn*	*Range*
Control (n = 16)	0	4	0	4	0	4
Comparison (n = 16)	0	3	0	4	1.5	3
Study (n = 20)	0	3	1.5	4	3	4

To better understand the effectiveness of the treatment tasks, the median score differences between consecutive tests were calculated for each participant by subtracting earlier scores from latter scores at the individual level. Table four provides the median score differences and respective ranges for each group.

Table four shows that only the comparison and experimental groups were able to achieve median score gains. The comparison group achieved a median score gain from the pretest to the posttest for the structure of the simple past passive, and the experimental group achieved median score gains from the pretest to the posttest for both target structures. Furthermore, the collaborative condition was responsible for the largest median score gains. In contrast, the control group did not achieve any median score gains, and for all groups no median score gains were achieved from the posttest to the delayed posttest for both target structures. However, for both target structures, no groups recorded median score declines between consecutive tests. Thus, all groups either improved or maintained their level of performance between tests. Table four also indicates that the highest range was seven and the lowest range was three. Since a participant could score a maximum of four points on any given test, a range of over four indicates that the scores of some participants decreased from one test to the next.

Table 4. *Median score differences and respective ranges*

Group	Simple past passive				Present continuous passive			
	Pretest-Posttest		Posttest-Delayed Posttest		Pretest-Posttest		Posttest-Delayed Posttest	
	Mdn	*Range*	*Mdn*	*Range*	*Mdn*	*Range*	*Mdn*	*Range*
Control (*n* = 16)	0	7	0	6	0	3	0	4
Comparison (*n* = 16)	0.5	6	0	6	0	7	0	5
Study (*n* = 20)	2	6	0	3	0.5	4	0	5

Trends Within the Data

A relatively large proportion of each group's participants were unable to improve on their previous test score. In total, there are 101 unchanged scores (table five).

Table 5. *Amount and location of unchanged scores*

| Group | Unchanged scores | | | |
| | Simple past passive | | Present continuous passive | |
	Pretest - Posttest	Posttest – Delayed Posttest	Pretest - Posttest	Posttest – Delayed Posttest
Control (*n* = 16)	12	8	12	13
Comparison (*n* = 16)	6	5	8	6
Experimental (*n* = 20)	5	5	10	11

Table five shows that the control group has the highest number of unchanged scores (45). Furthermore, for both target structures a relatively large proportion of the participants in the comparison and experimental groups received the treatment and yet were unable to improve on a previous test score.

There are thirty instances of a participant's score declining. This breaks down into seven for the control group, eleven for the experimental group, and twelve for the comparison group (table six). Across the groups, the majority of these performance declines occurred between the posttests and the delayed posttests (twenty-two); however, eight score declines also occurred between the pretests and the posttests.

Table 6. *Amount and location of score declines*

| Group | Score declines | | | |
| | Simple past passive | | Present continuous passive | |
	Pretest - Posttest	Posttest – Delayed Posttest	Pretest - Posttest	Posttest – Delayed Posttest
Control (*n* = 16)	1	4	1	1
Comparison (*n* = 16)	2	5	2	3
Experimental (*n* = 20)	2	8	0	1

Although much individual variation exists within the data, the largest proportion of the recorded performance gains for both target structures can be attributed to the experimental group and occurred after the treatment condition of collaborative learning had been administered.

Effectiveness of Intervention
The effectiveness of the treatment tasks was determined by analysing individual participants' score changes between tests (i.e., between score gains/declines at the individual level). In total, there are twelve data sets. This is due to having three conditions (i.e., control, comparison, and experimental), two target structures (i.e., simple past passive, and present continuous passive), and two test score changes (i.e., pretest to posttest, and posttest to delayed posttest).

A Shapiro–Wilk test revealed that seven of the data sets significantly deviate from normality and that there is no robust evidence that five data sets differ from normality. Additionally, a Non-parametric Levene F-test revealed that the data sets for the structure of the simple past passive have a statistically similar variance; whilst, the data sets for the structure of the present continuous passive violate the homogeneity of variance. Taking into consideration the results of the Shapiro–Wilk test and the Non-parametric Levene F-test, the data is analysed using non-parametric tests. This results in the data analysis being more robust.

A Mood's median test was employed in order to analyse the median score differences (Mood, 1954). Mood's median test was selected because it can be used with three independent groups, can be used with ordinal data, and it does not make assumptions about distribution (i.e., whether the data is normally distributed and whether the variance of the data is approximately equal across samples). For each target structure, the score differences between consecutive tests were compared across treatment conditions (Table seven).

Table 7. *Mood's median test*

Consecutive tests	Mood's median test					
	Simple past passive			Present continuous passive		
	M	df	p-value	M	df	p-value
Pretest-Posttest	7.79	2	0.02*	3.75	2	0.15
Posttest-Delayed Posttest	0.05	2	0.72	4.34	2	0.11

*$p \leq .05$

For the target structure of the present continuous passive, the results of the Mood's median test suggest that completing the treatment tasks, either individually, collaboratively, or not at all, did not have a statistically significant effect on the performance of the participants. However, a statistically significant effect was found between the pretest and posttest for the target structure of the simple past passive ($M = 7.79$, $df = 2$, $p = 0.02$). Additionally, the non-significance from the posttests to the delayed posttests for both target structures confirms that learning was maintained. The Mood's median test does not identify between which groups the statistically significant differences occurred. Thus, a post-hoc analysis is required.

In order to conduct a post-hoc analysis, the Mood's median test was repeated on the results of the simple past passive from the pretest to posttest with a different group omitted from the analysis each time (Mood, 1954) (see table eight). A Bonferroni adjustment was made by dividing the alpha level (0.05) by the number of between group comparisons (three), resulting in a post-hoc alpha level of 0.02. In order to better understand the strength of any association, the effect size was calculated using Cramer's coefficient (Cramér's V).

Table 8. *Post-hoc analysis for the simple past passive between the pretest and posttest*

Groups Compared	M	df	p-value	Cramér's V
Control - Comparison	3.46	1	0.06	0.33
Comparison - Experimental	0.82	1	0.36	0.15
Experimental - Control	7.70	1	0.01*	0.46

*$p \leq .02$

The results of the post-hoc analysis suggest that there is a statistically significant difference between the performance of the experimental group and the performance of the control group for the structure of the simple past passive between the pretest and the posttest ($M = 7.70$, $df = 1$, $p = 0.01$). Furthermore, the difference between these groups is approaching a large effect (Cramér's $V = 0.46$). A moderate effect exists between the control and comparison groups (Cramér's $V = 0.33$); however, the *p*-value suggests that this effect is not statistically significant ($M = 3.46$, $df = 1$, $p = 0.06$). No statistically significant difference was found between the experimental and comparison groups for the structure of the simple past passive between the pretest and the posttest ($M = 0.82$, $df = 1$, $p = 0.36$).

Discussion

This study utilized interventionist dynamic assessment when investigating whether L2 learners who are situated within an undergraduate Qatari EFL context learn L2 grammatical structures more effectively by completing a series of treatment tasks either collaboratively or individually. The descriptive statistics suggest that the treatment condition of working collaboratively had a greater impact on the participants' linguistic development than either completing the treatment tasks individually or not completing them at all. Furthermore, the absence of median score declines between the posttests and the delayed posttests suggests that the collaborative condition's gains were stable over the duration of the study. However, only one statistically significant difference was found. A Mood's median test shows a pretest to posttest statistically significant difference, which is moderate to large in size (Cramér's $V = 0.46$), between the performance of the collaborative condition and the performance of the control group for the structure of the simple past passive (Mood, 1954). Thus for the target structure of the simple past passive, there is a moderate to large association between completing the treatment tasks collaboratively and the resulting linguistic development when compared to not completing the treatment tasks. No other statistically significant differences were found.

No statistically significant differences were found between the collaborative condition and the individual condition. This suggests that for the participants in this study completing the treatment tasks collaboratively was not significantly more effective in facilitating linguistic development than completing the treatment tasks individually. The absence of statistically significant differences between the results of these two groups suggests that the descriptive differences between them may be due to random variation, measurement error, or a lack of statistical power. This finding is largely in line with previous SLA research which has investigated working collaboratively and the attainment of specific grammatical outcomes. When employing a pretest-treatment-posttest design to investigate the effectiveness of working collaboratively in order to learn predetermined grammatical structures, two studies found statistically significant differences at posttesting between the conditions of working collaboratively and working individually (Spielman-Davidson, 2000; Teng, 2020), whilst all other studies found that although descriptive differences were present between the two learning conditions, statistically significant differences were absent (Kuiken & Vedder, 2002; Nassaji & Tian, 2010; Reinders, 2009). As these studies all employed the attainment of specific grammatical outcomes as their dependent variables, it is possible that working collaboratively may be better suited to tasks which have a more open-ended outcome.

The median score gains for the control group show no group-level development. As well as obtaining a nuanced understanding of the participants' linguistic development, the use of interventionist dynamic assessment also had the potential to promote the participants' linguistic knowledge of the target structures (Mohammadimoghadam, 2015). However, at the group-level, there is no evidence that participants who did not complete the treatment tasks developed their knowledge of either target structure. Several explanations exist. Firstly, the assistance provided during testing was predetermined and standardized; it was not intentionally tailored to the needs of each participant. Secondly due to the five-minute testing time limit, it is probable that some participants rushed reading the scenarios contained with each test item as well as writing their initial answers. Thirdly, participants who were unable to correctly write a target structure were shown the correct answer but were not provided with a corresponding explanation. Fourthly, in comparison to the other two groups the control group has the highest median pretest score for the structure of the simple past passive. Thus, the scope for gains for this target structure was more limited in comparison to the other two groups. Finally, the control group only contained 16 participants; thus, the results could have been affected by learner variation.

Within the data, there were 101 unchanged between test scores. Twenty-nine scores of participants who received the treatment, either individually or collaboratively, remained unchanged from their pretest to their posttest. One explanation is that the treatment tasks as well as the assistance provided during the pretest were ineffective for some participants. Another explanation concerns the sensitivity of the data collection tools. In order to score points, a participant's answer needed to be entirely correct. Thus, the performance of these participants may have improved but not enough to register a score on a data collection tool.

Within the data, there are 30 instances of a participant's score declining between consecutive tests. Several explanations exist. Firstly, a participant may have correctly guessed the answer on a preceding test. Due to time constraints, for both target structures only a single test item was administered; there was no verification from a second test item. Secondly, a participant may not have understood the scenario created by an item on a subsequent test. Thirdly, a participant may have not taken a subsequent test seriously. Finally, completing the treatment sessions may have negatively affected the performance of some participants.

The high levels of individual variation within the data indicate that some learners benefited more from participating in the study than others. Other studies which have investigated learner-learner interaction have also reported considerable individual variation (Adams, 2007; Fernández Dobao, 2012; Storch, 2005; Swain & Lapkin, 1998). The large amount of individual variation contained within the data suggests that the sample size may be too small to conduct research with this design.

The design of this study contains several limitations. Firstly, this study's design is too complex. This study's design contains two target structures; the intention was to provide two dependent variables. However, the target structures have a similar syntactic structure and share some of the same parts of speech (e.g., past participle and the preposition 'by'). Consequently, linguistic input received when competing a test or treatment task pertaining to one of the target structures could potentially aid the development of the other target structure. Secondly, this study has a sample of 52. The small sample size limited the power of the statistics. Repeating this

study with a larger sample size may result in the emergence of more statistically significant differences. Thirdly, the testing procedure contained time limits. Due to the five-minute testing time limit, it is probable that some participants rushed reading the scenarios that were contained within the test items as well as writing their initial answers. Fourthly, the post-task feedback given by the teacher was not carefully controlled. Thus, the post-task feedback given to the experimental and comparison groups could have differed. Finally, neither in its design or in its data analysis did this study control for the intervening factors of age and proficiency level.

Conclusion

The present study attempted to investigate a possible relationship between collaboratively completing form-focused tasks and linguistic development for Qatari undergraduate learners. Although measuring emergent abilities which are still in the process of developing provided a more complete picture of the efficacy of working collaboratively, no statistically significant differences were found between the results of the collaborative condition and the results of the individual condition. The lack of a statistically significant difference between the performances of the experimental and comparison groups for both target structures suggests that working collaboratively is not statistically more effective in facilitating learners' linguistic development than working individually. However due to the major limitations of this study, caution needs to be exercised when coming to any conclusions. Future research could include repeating a version of this study which addresses its limitations. In addition, the dynamic assessment procedure utilized within this study could be computerized.

About the Author

James Scotland currently teaches English in the Foundation Program at Qatar University. James has previously taught in Australia, Canada, England, South Korea, and Turkey. He has a Cambridge DELTA, a Master's in Education, and a Doctorate in Education. His professional interests include Vygotskian sociocultural theory and second language acquisition.
ORCID: https://orcid.org/0000-0003-1728-3839

References

Adams, R. (2007). Do second language learners benefit from interacting with each other. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 29-51). Oxford University Press.

Adams, R., Nuevo, A. M., & Egi, T. (2011). Explicit and implicit feedback, modified output, and SLA: Does explicit and implicit feedback promote learning and learner–learner interactions? *The Modern Language Journal, 95*(s1), 42-63. https://doi.org/10.1111/j.1540-4781.2011.01242.x

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1-18. https://doi.org/10.1037/a0021017

Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal, 78*(4), 465-483. https://doi.org/10.1111/j.1540-4781.1994.tb02064.x

Antón, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals, 42*(3), 576-598. https://doi.org/10.1111/j.1944-9720.2009.01030.x

Antón, M. (2019). Expanding the role of dynamic assessment in language education. *Language and Sociocultural Theory, 6*(1), 116-131. https://doi.org/10.1558/lst.38913

Antón, M., & DiCamilla, F. J. (1999). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. *The Modern Language Journal, 83*(2), 233-247. https://doi.org/10.1111/0026-7902.00024

Budoff, M. & M. Friedman. (1964). "Learning potential" as an assessment approach to the adolescent mentally retarded. *Journal of Consulting Psychology, 28*(5), 434–439. https://doi.org/10.1037/h0040631

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition, 41*(5), 671-682. https://doi.org/10.3758/s13421-012-0291-4

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Davin, K. J., & Gómez-Pereira, D. (2019). Evaluating instruction through classroom dynamic assessment: A sandwich approach. *Language and Sociocultural Theory, 6*(1), 6-31. https://doi.org/10.1558/lst.38914

Donato, R. (1994). Collective scaffolding in second language learning. In J. P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 33-56). Ablex.

Dumas, D., McNeish, D., & Greene, J. A. (2020). Dynamic measurement: A theoretical–psychometric paradigm for modern educational psychology. *Educational Psychologist, 55*(2), 88-105. https://doi.org/10.1080/00461520.2020.1744150

Eckerth, J. (2008). Investigating consciousness-raising tasks: Pedagogically targeted and non-targeted learning gains. *International Journal of Applied Linguistics, 18*(2), 119-145. https://doi.org/10.1111/j.1473-4192.2008.00188.x

Elabdali, R. (2021). Are two heads really better than one? A meta-analysis of the L2 learning benefits of collaborative writing. *Journal of Second Language Writing, 52*, 1-16. https://doi.org/10.1016/j.jslw.2020.100788

Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing, 21*(1), 40-58. https://doi.org/10.1016/j.jslw.2011.12.002

Fernández Dobao, A. (2014a). Vocabulary learning in collaborative tasks: A comparison of pair and small group work. *Language Teaching Research, 18*(4), 497-520. https://doi.org/10.1177/1362168813519730

Fernández Dobao, A. (2014b). Attention to form in collaborative writing tasks: Comparing pair and small group interaction. *Canadian Modern Language Review, 70*(2), 158-187. https://doi.org/10.3138/cmlr.1768

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies of proximal development. *Child Development, 57*(5), 1087-1099. https://doi.org/10.2307/1130433

Gánem-Gutiérrez, G. A (2008). Microgenesis, method and object: A study of collaborative activity in a Spanish as a foreign language classroom. *Applied Linguistics, 29*(1), 120-148. https://doi.org/10.1093/applin/amm032

Guthke, J. & J. F. Beckmann. (2000). The learning test concept and its applications in practice.

In C. S. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp.17-69). Elsevier.

Kuiken, F., & Vedder, I. (2002). The effect of interaction in acquiring the grammar of a second language. *International Journal of Educational Research, 37*(3), 343-358. https://doi.org/10.1016/s0883-0355(03)00009-0

Lantolf, J. P. (2009). Dynamic assessment: The dialectic integration of instruction and assessment. *Language Teaching, 42*(3), 355-368. https://doi.org/10.1017/s0261444808005569

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press. https://doi.org/10.2167/le127b.0

Lantolf, J. P., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research, 15*(1), 11-33. https://doi.org/10.1177/1362168810383328

Lapkin, S., Swain, M., & Smith, M. (2002). Reformulation and the learning of French pronominal verbs in a Canadian French immersion context. *The Modern Language Journal, 86*(4), 485-507. https://doi.org/10.1111/1540-4781.00157

Long, M. H., & Porter, P. A. (1985). Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly, 19*(2), 207-228. https://doi.org/10.2307/3586827

McNeil, L. (2018). Understanding and addressing the challenges of learning computer-mediated dynamic assessment: A teacher education study. *Language Teaching Research, 22*(3), 289-309. https://doi.org/10.1177/1362168816668675

McNicoll, J., & Lee, J. H. (2011). Collaborative consciousness-raising tasks in EAL classrooms. *English Teaching: Practice and Critique, 10*(4), 127-138.

Minick, N. (1987). Implications of Vygotsky's theories for dynamic assessment. In C. Lidz (Ed.), *Dynamic Assessment* (pp. 116–140). Guilford Press.

Mohammadimoghadam, M. (2015). Effects of mediation on an EFL learner's grammar development: A case study of an EFL beginner student. *Procedia-Social and Behavioral Sciences, 192*, 101-106. https://doi.org/10.1016/j.sbspro.2015.06.015

Mood, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *The Annals of Mathematical Statistics, 25*(3), 514-522. https://doi.org/10.1214/aoms/1177728719

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research, 14*(4), 397-419. https://doi.org/10.1177/1362168810375364

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528. https://doi.org/10.1111/0023-8333.00136

Pica, T., & Doughty, C. (1985). The role of group work in classroom second language acquisition. *Studies in Second Language Acquisition, 7*(02), 233-248. https://doi.org/10.1017/s0272263100005398

Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal, 91*(3), 323-340. https://doi.org/10.1111/j.1540-4781.2007.00583.x

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer Science & Business Media.

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

Rassaei, E. (2019). Tailoring mediation to learners' ZPD: Effects of dynamic and non-dynamic corrective feedback on L2 development. *The Language Learning Journal, 47*(5), 591-607. https://doi.org/10.1080/09571736.2017.1343863

Reinders, H. (2009). Learner uptake and acquisition in three grammar-oriented production activities. *Language Teaching Research, 13*(2), 201-222. https://doi.org/10.1177/1362168809103449

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24*(3), 355-367. https://doi.org/10.1007/s10648-012-9201-3

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69-97). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-85098-1_5

Scotland, J. (2017). *Participating in a shared cognitive space: An exploration of working collaboratively and longer-term performance of a complex grammatical structure*, (Unpublished Doctoral Dissertation). University of Exeter, United Kingdom. http://hdl.handle.net/10871/32739

Scotland, J. (2021). The efficacy of collaboratively completing form-focused tasks: A research synthesis. In M. D. P. G. Mayo (Ed.), *Working collaboratively in second/foreign language learning* (pp. 35-57). De Gruyter Mouton. https://doi.org/10.1515/9781501511318-003

Spielman-Davidson, S. J. (2000). *Collaborative dialogues in the zone of proximal development, grade eight French immersion students learning the conditional tense*, (Unpublished Doctoral Dissertation). Ontario Institute for Studies in Education of the University of Toronto, Canada.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing, 14*(3), 153-173. https://doi.org/10.1016/j.jslw.2005.05.002

Storch, N., & Wigglesworth, G. (2007). Writing tasks: The effects of collaboration. In M. D. P. García Mayo (Ed.), *Investigating tasks in formal language settings* (pp. 157-177). Multilingual Matters. https://doi.org/10.21832/9781853599286-011

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural Theory and Second Language Learning* (pp. 97-114). Oxford University Press.

Swain, M. (2001). Integrating language and content teaching through collaborative tasks. *Canadian Modern Language Review, 58*(1), 44-63. https://doi.org/10.3138/cmlr.58.1.44

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal, 82*(3), 320-337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99-118). Routledge. https://doi.org/10.4324/9781315838267-14

Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research, 37*(3/4), 285-304. https://doi.org/10.1016/s0883-0355(03)00006-5

Teng, M. F. (2020). The effectiveness of group, pair and individual output tasks on learning

phrasal   verbs.   *The   Language   Learning   Journal*,   *48*(2),   187-200.
doi.org/10.1080/09571736.2017.1373841

van Compernolle, R. A., & Zhang, H. (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

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wajnryb, R. (1990). *Grammar dictation*. Oxford University Press.

Appendices
Appendix A
The Guided Learning Task for the Structure of the Simple Past Passive

Part 1

Example Sentences

Sentence 1: The student took the exam.

Sentence 2: The exam was taken by the student.

1a)   In   sentence   1,   who   took   the   exam?
_____

1b)   In   sentence   2,   who   took   the   exam?
_____

1c)   Is the meaning of the sentences different? Yes/No
         If   yes,   then   how?
_____

Circle the correct word

2a)   Sentence   1   is   in   the   *past/present/future*.   How   do   you   know?
_____

2b)   Sentence   2   is   in   the   *past/present/future*.   How   do   you   know?
_____

Fill in the missing number

3a)   How   many   words   are   in   sentence   1?
_____

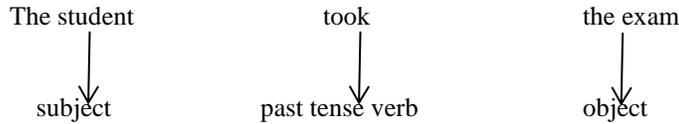3b)   How   many   words   are   in   sentence   2?
_____

3c)   Which   words   are   different?
_____

Circle the correct word
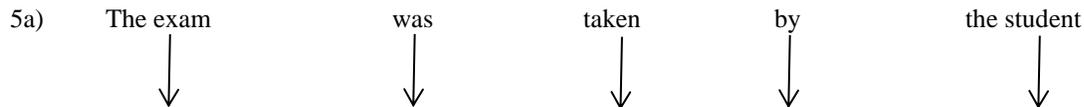
4a)   Sentence 1 is in the *active/passive* voice.

How	do	you	know?

_____

4b)   Sentence 2 is in the *active/passive* voice.

How	do	you	know?

_____

The structure of sentence 1 is

The student	took	the exam
↓	↓	↓
subject	past tense verb	object

Use the words in the box to make the structure of sentence 2

| subject |
| object |
| past participle |
| be verb |
| by |

5a)   The exam   was   taken   by   the student

       ↓   ↓   ↓   ↓   ↓

Complete the following sentence. Use the words in the box to help you.

| past participle |
| be verb |

5b)   The verb phrase of the simple past passive is made using a _____ followed by the _____ of the main verb.

Part 2
Please change the following sentences to the active voice

Example
       Passive: The game was played by the girl
       Active: The girl played the game.

1a)   *Passive:*   The email was deleted by the worker.
       Active: The worker _____

1b)   *Passive:* The student was tested by the teacher.
       Active: _____

1c)   *Passive:*   The letters were delivered by the worker.
       Active: _____

Please change the following sentences to the passive voice

Example

  Active: The girl played the game.
  Passive: The game was played by the girl

2a) *Active:* The scientist researched the idea.
   Passive: The idea _____

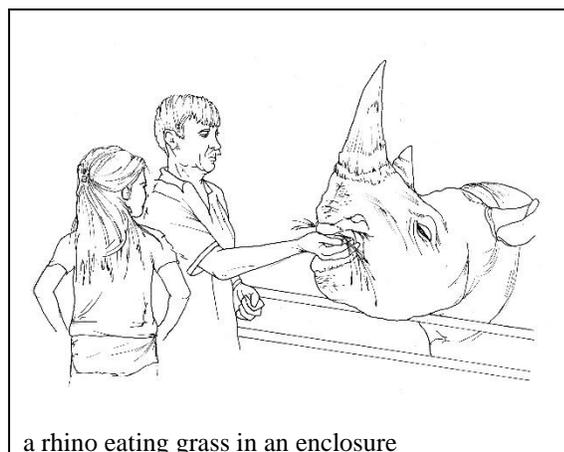2b) *Active:* The football player scored the goal.
   Passive: _____

2c) *Active:* The police officer investigated the crimes.
   Passive: _____

Appendix B
The Text-editing Task for the Structure of the Present Continuous Passive

Directions: Please read the following text. As you read, please underline and correct any mistakes that you find.

  Hint – There are six mistakes. Three mistakes are in the active voice and three mistakes are in the passive
  voice.

Task 1

Right now, Maha is at the zoo. The weather is very nice and the sun is shine brightly. Maha observing a zoo keeper feed two rhinos. The rhinos is being fed fresh grass by the zoo keeper. The grass is chewed by the rhinos. Maha watch by one rhino as it slowly chews the grass. Maha press her hands onto the glass of the enclosure.

a rhino eating grass in an enclosure

Appendix C
The Dictogloss Task for the Structure of the Simple Past Passive

The administration procedures for the dictogloss tasks

- The teacher introduced and explained each stage of the dictogloss task to the participants.
- The teacher played each audio recording thrice. The students took notes during the second and third times.
- The participants were given around 10 minutes to recreate the text and check their reconstruction against the original text.
- The teacher provided feedback in a whole class setting.

Oil was used by people thousands of years ago. In ancient times, it was burned by people in oil lamps for light at night. Also, people covered boats with oil to keep water out and the Chinese used oil as a surface for roads. About a hundred years ago, far more oil was needed as a modern transport industry developed. Luckily, people found large amounts of oil in many parts of the world, including the Middle East and North America.