Planning Time, Strategy Use, and Written Task Production in a Pedagogic vs. a Testing Context

Seyed Reza Meraji University of Tehran, Tehran, Iran Email: rmeraji@ut.ac.ir

Abstract—As there has been little attempt to delve into performance differentials occasioned as a function of planning time in pedagogic versus testing contexts, the present study, taking up a mixed methods approach, addressed the effects of pre-task planning on the written output of intermediate EFL learners under four conditions: 1) no planning in a pedagogic context, 2) pre-task planning in a pedagogic context, 3) no planning in a testing context, and 4) pre-task planning in a testing context. Results showed that provision of pre-task planning time fostered accuracy, syntactic complexity, and fluency in the pedagogic context. In the testing context, also, more accurate, syntactically complex, and fluent output was generated by planners compared to no-planners. Finally, the data gleaned through the interviews showed that metacognitive and cognitive strategies were the most used strategies and that no significant difference existed between planners in the two contexts in terms of strategy use.

Index Terms—pre-task planning, accuracy, complexity, fluency, metacognitive strategies, cognitive strategies

I. INTRODUCTION

Planning time studies have always garnered the attention of researchers, syllabus designers, language teachers, and language testers. This might have followed from findings and implications of studies which indicate that planning more often than not leads to gains in areas of performance; esp. in fluency and complexity (e.g., Ellis & Yuan, 2004; Foster & Skehan, 1996; Mehnert, 1998; Rouhi & Marefat, 2006; Yuan & Ellis, 2003). Undergirding most of these studies are information processing theories such as those of VanPatten (2002) and Huitt (2003), where it was predicted that in light of limited attentional resources which are at learners' disposal, spoken discourse generated by L2 learners would favor one area of performance (meaning) to the detriment of another (form) (Skehan, 1998; VanPatten, 2002). It was proposed that provision of planning time would unpack some of the real time processing load incurred on L2 learners which buys them time to concentrate more on their performance during on-line execution of the task promoting quality and quantity of performance.

Another account of processes of speech production in L2 was advanced in Anderson's (1983) ACT theory. This theory posits that in skill acquisition, one's explicit knowledge of a skill (declarative knowledge) transforms into implicit knowledge manifest in one's behavior (procedural knowledge). Elder and Iwashita (2005) relate this theory to the beneficial effect of planning time on performance by suggesting that provision of pre-task planning time may ease access to the declarative knowledge through retrieval and rehearsal so that pertinent grammatical forms and lexical items that facilitate production may be retrieved during on-line performance and ready-made plans can be recalled and uttered.

Finally, Skehan (1998) put forth the dual processing mode theory juxtaposing a rule-based system and an exemplarbased system. Unlike the exemplar-based system, the rule-based system is expected to exact a heavy processing burden during on-line language production, especially, in view of learners' limited attentional resources, as "rules need complex processes of construction to underpin their operation...and require detailed attention during comprehension and assembly during production" (p. 89) which make access to such a system time-consuming. It is believed that provision of pre-task planning time makes the rule-based system accessible, rendering language use more accurate.

In short, empirical investigation of such theoretically grounded models are thought to (a) aid task designers select tasks with the appropriate caliber of difficulty which draw upon the full range of learners' abilities; (b) help test developers properly construct test tasks and the conditions under which these tasks are executed; (c) aid development of descriptors of rating scales; and (d) ease interpretation of test scores in accordance with the difficulty of the task (Skehan, 1998).

A. Planning Time Studies in Pedagogic Contexts

In task-based language teaching, planning time has found expression as a task design variable whose contributions to language production cannot be overlooked in pedagogic contexts; especially, with regard to complexity and fluency of production.

In a planning time study, Robinson, Ting and Urwin (1996) investigated the effect of allocation of no planning time vs. 3-minute planning time on written and oral narratives of learners. Results showed that planning time fostered

fluency in speaking but not in writing. However, planners did not generate more accurate discourse in comparison to no-planners in either modality.

Taking up a product- and process-oriented approach toward planning time, Ortega (1999) probed the effects of pretask planning time (10-minute planning and no planning time) on the performance of learners of Spanish using narrative tasks which were ensued by semi-structured retrospective interviews immediately after each oral task so that the researcher could study the cognitive processes implicated during pre-task planning. Planners outperformed no-planners in terms of fluency. Moreover, planners syntacticized their language significantly more than no-planners. The accuracy measures produced significant results in favor of the planning group, too.

As regards the interviews, planners tried to complexify their discourse more than no-planners. Furthermore, provision of pre-task planning time was found to have favored higher-proficiency learners more than their low-proficiency counterparts. In short, it was found that attention to form is not a natural corollary of pre-task planning. Rather, it may come to pass in the course of pre-task planning only for the developmentally ready leaner in conjunction with the communicative demands of the task and learner's proclivity to attend to form and thereby accuracy. At the end, it is suggested that learner's proficiency level should be taken into account in psycholinguistically-oriented theories of attention, as in an earlier similar study (Ortega, 1995, as cited in Ortega, 1999), it was reported through self-reports that for lower-proficiency learners, communicative stress and lexical retrieval problems registered monumentally in the learners' mind during pre-task planning and performance which could have taken a toll on rehearsal and recall strategies.

Furthering their pursuit of bifurcation of planning time, Ellis and Yuan (2004) implemented essentially the same experimental design as that of their previous study (Yuan & Ellis, 2003) around a narrative writing task in order to extend their findings regarding planning time across modalities. As expected, planners produced significantly more fluent discourse in comparison to their no-planning counterparts. Pre-task planners syntacticized their discourse significantly more than no planners and almost significantly compared to on-line planners. Accuracy measures revealed that the difference between on-line planners and no planners verged on significance with on-line planners being the most accurate among all.

In a rare attempt at exploring the composite effects of task design indices, Gilabert (2004) crossed planning time (50second planning time vs. 10-minute planning time) with +/- Here-and-Now (keeping the visual stimuli vs. removal of visual stimuli). Four different oral narrative tasks which were manipulated along the task features were used. It was found that provision of planning time aided fluency in terms of pruned speech rate in both Here-and-Now and Thereand-Then planning conditions. As regards lexical complexity assessed by Guiraud's Index of lexical richness, planning conditions lexicalized their speech more than no-planning conditions.

Contesting the previous research which has not taken account of the bearings of pre-task planning on L2 learners' allocation of attention to form, meaning, or both, Sangarun (2005) examined the effects of form-fused, meaning-focused, and meaning/form-focused pre-task planning on EFL learners' pre-task cognitive processes, on the way L2 learners put their strategic plans into effect, and on the measures of spoken production thus affected (fluency, complexity, and accuracy). Having provided think-alouds during the strategic planning phase, the learners in the planning groups executed an instruction task ensued by an argumentative task following the same procedure.

In the main, regardless of their planning condition, learners stood to focus on meaning, when engaged in pre-task planning. Overall, the instructions led the learners to focus on either form or meaning related to their respective groups, i.e., the form-focused group planned vocabulary on a large scale, the meaning-focused group planned meaning intensively, and the meaning/form-focused group divided their attention between form (grammatical corrections) and meaning.

Also, regarding application of strategic plans, it was found that learners in the meaning-focused and meaning/form-focused group used more planned ideas and fewer unplanned ideas than the other planning group concerning the argumentative task and the instruction task, respectively. To boot, the meaning/form-focused group used more planned forms in both tasks. The researcher concluded that pre-task planning directed toward meaning/form is more effective than either form-focused or meaning-focused strategic planning.

Regarding accuracy assessed through the percentage of error-free clauses and the number of errors per 100 words, positive accuracy gains were found for the meaning-focused and form-focused groups for the argumentative task and in the meaning/form-focused group for both tasks. Regarding complexity measured by the number of S-nodes per T-unit and the number of clauses per T-unit, the hypothesis that the meaning-focused and meaning/form-focused groups would promote speech complexity was borne out. In terms of speech fluency calculated by speech rate and the percentage of total pausing time, positive effects for all three planning groups were found for the instruction task and for the form-focused group for the argumentative task.

B. Planning Time Studies in Testing Contexts

A number of studies in testing contexts have also homed in on the bearings of planning time on performance. Interestingly, relatively attenuated gains in performance areas were reported in these contexts. To compound the problem, the operationalizations of testing contexts have differed across studies.

Using an array of tasks, Wigglesworth (1997) investigated planning (no planning time and 1-minute planning time) across proficiency levels (high and low). Both discourse-analytic measures and analytic ratings were implemented. It was found that planning time led to the complexification of higher-proficiency testees' speech in the more difficult tasks.

Also, planners produced more fluent discourse compared to no-planners. It was hypothesized that the easier the task, the less likely it is that planning time advantages performance. Analysis of accuracy gains revealed that performance was significantly promoted by planning time for the high-proficiency group in the summary of conversation and picture tasks. Regarding the low-proficiency group, performance in terms of accuracy on all tasks except for picture tasks was improved by planning time. As regards analytic ratings, the researcher found that the trained raters did not find any differences across the proficiency groups.

Iwashita, McNamara and Elder (2001) explored the differential effects of manipulation of planning time (3.5 mins vs. 0.5 min). No significant differences were found in terms of the discourse-analytic measures of accuracy, complexity, and fluency, findings which did not accord well with the existing findings pertaining to Limited Attentional Capacity Model (Skehan, 1998; Skehan and Foster, 1999, 2001).

Elder and Iwashita (2005) probed the effect of pre-task planning on oral performance in a testing context and on test takers' perceptions of task difficulty. Measures of accuracy, complexity, and fluency did not show any difference between the planning conditions. In addition, it was found that pre-task planning exerted no influence on perceptions of task difficulty and task enjoyment. At the end, the researchers questioned the contribution of pre-task planning to language testing and came to the conclusion that results yielded may be "due to the fact that language behavior in a test situation is fundamentally different from language of the classroom or of real world encounters and therefore that findings of SLA classroom research may not be generalizable to testing contexts" (p. 235).

Tavakoli and Skehan (2005) crossed degree of structure (4 levels) and strategic planning across elementary and intermediate proficiency levels. Planners generated more fluent discourse than no-planners. Accuracy was also promoted as a function of planning time. The planning group produced significantly more complex discourse than the no-planning group.

Grounding his speculation in the findings of planning studies in pedagogic and testing contexts, Ellis (2005) contends that "planning seems to have less effect on performance on tasks in tests than in other contexts" (p. 217). By punctuating the high-stakesness of a test environment, and therefore, the significance of testees' performance in a testing context, Ellis proposes that apparently a testing environment calls for attention to accuracy while steering attention away from other areas of performance and that the high-stakesness of the test environment heightens attention to accuracy during task execution which in turn eschews any effect of pre-task planning on all three areas of performance.

By and large, the above cited research indicates that planning favors fluency and complexity of production in pedagogic contexts; however, results regarding the effect of planning on performance areas in testing contexts are mixed. Therefore, a single study of planning in both pedagogic and testing contexts bears special significance, as the differentials of performance in these two different contexts have not been explored in one single study. To compound the problem, in addition to the fact that the writing mode has received scant attention in planning research in pedagogic contexts, no studies of planning have targeted writing in a testing context. This seems bizarre since the role of pre-task planning cannot be denied in writing. Finally, what planners do when they plan have gone unheeded in previous research save in a handful of oral studies (Ortega, 1999, 2005). Therefore, the present study sought to ascertain the effects of pre-task planning on written performance in pedagogic and testing contexts in terms of discourse-analytic measures of accuracy, syntactic complexity, lexical complexity, and fluency, and also to explore the strategies utilized during pre-task planning time through post-task interviews.

Research Questions:

The following research questions led the course of the present investigation:

1. Does manipulation of pre-task planning time and context affect written performance in terms of accuracy?

2. Does manipulation of pre-task planning time and context affect written performance in terms of syntactic complexity?

3. Does manipulation of pre-task planning time and context affect written performance in terms of lexical complexity?

4. Does manipulation of pre-task planning time and context affect written performance in terms of fluency?

5. Does manipulation of context affect strategies adopted during pre-task planning time?

II. METHOD

A. Participants

The participants for the present study were 75 Iranian intermediate EFL learners, 43 females and 32 males, aged between 18 and 41, chosen from 101 English learners from twelve intact mixed sex classes. The participants had been placed in the same level of English class by either an institutional placement test or a final examination composing of a written examination and an oral interview at a language school, namely Farhikhtegane Daneshgah Institute. At the time of data collection, the participants were studying New Interchange 2. In order to ensure homogeneity of the groups of participants in terms of language proficiency, Nelson Language Proficiency test (Fowler & Coe, 1976) was administered; the participants whose scores ranged between 50 and 60 were selected.

There were four groups of participants; learners from every three classes were grouped together. Each of these groups was assigned to one of the conditions in the study. The composition of each group was as follows: The noplanning group in the pedagogic context consisted of 16 language learners; the pre-task planning group in the pedagogic context consisted of 18 participants; the no-planning group in the testing context consisted of 18 language learners; and the pre-task planning group in the testing context consisted of 23 participants. Further analysis based on the proficiency test showed no significant differences across the groups of participants in terms of language proficiency.

B. Material

The task used in the present study was a tightly structured picture narrative consisting of eleven pictures taken from Salvador (1991). A narrative task was chosen because first, in addition to being used by some international testing organizations (e.g., Test of Spoken English, University of Cambridge's YLE), narrative tasks have a long tradition in task-based research (Ortega, 1999) in light of their being non-interactive and thus openness to greater control (Tavakoli & Skehan, 2005) and the resultant suitability for standardization (Tavakoli, 2009); second, the use of narrative tasks enables comparison with studies of planning which have used similar tasks (e.g., Foster & Skehan, 1996; Gilabert, 2007); third, learners are familiar with their macrostructure (Skehan, 2001); fourth, they are deemed reasonably demanding on the part of learners (Yuan & Ellis, 2003), provided that some interpretation of the pictures bears down upon the learners (e.g., Skehan & Foster, 1999); fifth, the task used involves a small number of characters in the story (Gilabert, 2007); sixth, narrative tasks are reliable and authentic elicitation devices (Tavakoli & Skehan, 2005); seventh, they are tightly structured so as to eschew individual variation and facilitate comparison (Tavakoli & Skehan, 2005); eighth, they typify tasks used in EFL course books; finally, they are wordless so as not to clue learners in on the words that participants can use in their writing (Gilabert, 2007).

C. Procedures

Planning was defined at two levels of no planning and pre-task planning and context, on the other hand, was operationalized at two levels of pedagogic and testing contexts. These operationalizations generated the following conditions (Table 1): no planning in the pedagogic context; pre-task planning in the pedagogic context; no planning in the testing context; and pre-task planning in the testing context.

In the no planning & pedagogic context condition, the participants performed the task under normal classroom settings in the presence of their normal teacher with the researcher present all along. Upon receiving the picture set, the participants wrote the story embedded in the picture series in 14 minutes and had to write at least 150 words. The time limit for task completion was established based on a pilot study in which participants who had been at a proficiency level comparable to that of the participants in the main study had to write out the same narrative as the one used in the main study. No time limit was fixed. All times were jutted down and then, following Rouhi and Marefat (2006), averaged to set a time limit for the study in the pedagogic context. The resultant mean, and thereby the time limit, was 14 minutes. Following Ellis and Yuan (2004) and Yuan and Ellis (2003), this procedure was implemented so as to curb large scale on-line planning.

In the planning & pedagogic context condition, the participants performed the task under normal classroom settings in the presence of their teacher with the researcher present all the time. Like the no planning & the pedagogic context condition, they were required to finish the task in 14 minutes and write at least 150 words so that on-line planning was partially curtailed. Having been provided with the picture set, they were given a piece of paper to write notes during the pre-task planning time which was set at 10 minutes in the study. The provision of 10-minute pre-task planning time was based on previous research (Ellis & Yuan, 2004; Foster & Skehan, 1996; Gilabert, 2007; Ortega, 1999, to name but a few). Following Ellis and Yuan (2004) and Foster and Skehan (1996), no detailed instruction was given to the participants; they were asked to plan their output in terms of content, organization, and language. Upon the completion of 10 minutes of pre-task planning time, the notes were removed, with participants' prior knowledge. According to Ellis and Yuan (2004), the removal of written notes serves dual purposes: first, it ensures that the language generated during task completion is produced solely during that spell; secondly, notes can be used to probe how the participants plan and to help stimulate recall in retrospective interviews.

In the no planning & testing context condition, with the institute's consent, the researcher posing as an administrative in the institute informed the learners, who did not have any prior knowledge of an upcoming examination, that they would be taking an examination in the form of a pop quiz in that session. They were also told that the test results would have a direct impact on their final score, and therefore, on their entry into the following level. Upon receiving the picture set, the testees commenced writing the story embedded in the picture series. They had 16 minutes to complete the task and were asked to write at least 150 words. This time limit was fixed based on a pilot run in which participants had to narrate the same story as the one used in the main study. No time limit was fixed. All times were noted and then averaged to set a time limit for the study in the testing context which was clocked at 16 minutes. Following Ellis and Yuan (2004) and Yuan and Ellis (2003), this procedure was implemented so as to limit on-line planning.

In the planning & testing context condition, which was formulated the same as the no planning & the testing context condition in terms of testing context establishment, having been provided with the picture set, the testees were given a piece of paper to write notes during the pre-task planning time which was set at 10 minutes in the study. Upon the completion of 10 minutes of pre-task planning time, the notes were collected with testees' prior knowledge. Like the no planning & pedagogic context condition, they were required to finish the task in 16 minutes and write at least 150 words.

	I HE EXPERIMEN	NIAL DESIGN	
No planning & pedagogic context	Planning & pedagogic context	No planning & testing context	Planning & testing context
n = 16	n = 18	n = 18	n = 23
No pre-task planning time	10 minutes for planning	No pre-task planning time	10 minutes for planning
14 minutes to perform the task	14 minutes to perform the task	16 minutes to perform the task	16 minutes to perform the task
Picture set accessible during execution			
	Interview		Interview

TABLE 1 THE EXPERIMENTAL DESIGN

Subsequently, the 41 planners were interviewed by the researcher and a research assistant who had a degree in TEFL and was fully aware of research purposes and familiar with the research study. The interviews were conducted in Persian in order to gain better insights into what strategies were employed by the participants over the course of pre-task planning time, what differentials, if any, existed between the two planning groups in terms of their use of strategies while planning for an upcoming task, and finally what planners did during pre-task planning time. The participants were not previously let in on the fact that they would be interviewed. The reason for opting for stimulated retrospective interviews rather than introspective verbal reports was the dissenting accounts of validity of verbal reports (See Alavi, 2005; Cohen, 1994).

The interviews lasted approximately 15 minutes per interviewee. The purpose of the semi-structured interviews was to delve into the a priori categories pertaining to the taxonomies advanced by O'Malley and Chamot (1996), Oxford (1990), Ortega (2005) and Dornyei (1995).

The interviews were done with a strong focus on the possible differentials in terms of strategy use between pedagogic and testing planners. Accordingly, the picture task, the participants' notes collected after task performance, and the researcher's notes formed the basis for interview questions. To this end, an interview protocol was developed which was to be modified through follow-up questions as needs be. Subsequently, the researcher and the research assistant discussed the protocol and their approach to questioning the interviewees. The researcher and the research assistant did their utmost to eschew from asking leading questions which would conduce to researcher bias. All interviews were recorded by an MP3 recorder and subsequently transcribed for analysis of strategy use during pre-task planning time.

D. Data Analysis

In order to detect any possible variations across conditions, discourse-analytic measures of grammatical accuracy, syntactic complexity, lexical complexity, and fluency were run.

Accuracy was operationalized as the percentage of error-free clauses (EFC), the number of errors per 100 words (NER), and the percentage of target-like use of English articles (TLU). In order to implement EFC and NER, Polio's (1997) guidelines with regard to errors were strictly observed save misspellings (punctuation and capitalization). T-units were taken as the clauses which encompassed a main clause in addition to any subordinate clauses (Richards, Platt, & Weber, 1985). Sentence fragments were not deemed T-units. Clauses were defined as those which had finite verbs (Polio, 1997). As regards TLU, misspellings between *a* and *an* were overlooked.

As regards syntactic complexity, two production measures were used: the number of S-nodes per T-unit (S-nodes/T) and the number of clauses per T-unit (C/T). T-units and C-units have been both opted as measures of complexity in task-based research; however, Robinson (2001) views the former as a more viable index for measuring complexity of monologic narrative performance and the latter as a variant for assessing complexity of interactional production, inasmuch as dialogic performance is suffused with elliptical and fragmentary utterances. Following Mehnert (1998), S-nodes included both finite and nonfinite clauses. S-nodes are also thought to be a better measure of syntactic complexity in comparison to subordinate clauses (Crookes, 1989).

Lexical complexity was assessed through one measure of lexical variation, namely MSTTR (Mean Segmental Type-Token Ratio), springing from Ellis and Yuan (2004), Malvern, Richards, Chipere, and Duran (2004), and Yuan and Ellis (2003). Type-Token ratio (TTR) has been the mainstay of a good deal of task-based researches; however, many recent calls have brought the validity of this measure into question, as myriads of studies have not dealt in the fact that type-token ratio is overly sensitive to text length (Carrol, 1967; as cited in Wolfe-Quintero, Inagaki & Kim, 1998), i.e., the longer a text is, the lower the score for TTR. In order to mitigate this problem, MSTTR was implemented. As such, following Ellis and Yuan (2004) and Yuan and Ellis (2003), for the calculation of this index of lexical complexity, each narrative was divided into segments of 40 words and the Type-Token Ratio of each segment was computed by dividing the total number of words by the total number of different words in the segment. Subsequently, the MSTTR was calculated by adding the mean score of the segments and dividing the total by the total number of segments in the text for each participant.

Fluency was assessed through the calculation of the number of dysfluencies, i.e., the total number of words a participant reformulated (crossed out and changed) (Ellis & Yuan, 2004). Hyphenated words were counted as one word (Ishikawa, 2007).

All performances were coded by the researcher and an independent coder whose reliability index for measures of accuracy and fluency reached above .90. However, the reliability coefficients for the complexity measures registered lower. To remedy this problem, after a discussion session between the researcher and the independent coder, the data for these measures were re-coded. This time, the inter-coder reliability reached .84 and .88 for syntactic complexity measures (number of S-nodes/T and C/T, respectively) and .89 for the lexical complexity measure.

As regards the dataset, the normal distribution of the four groups' scores was tested in terms of skewness and kurtosis. In cases of normal distribution of scores, two-way between subjects ANOVAs were run for which main effects and partial eta squared (Cohen, 1988) were reported (0.01 = small, 0.06 = medium, 0.14 = large).

Concerning interviews, all interviews were coded in relation to the a priori categories anchored, in the main, in the taxonomy of O'Malley and Chamot (1996), and secondarily in that of Oxford (1990), of Ortega (2005) and one category from Dornyei (1995).

Selected from O'Malley and Chamot (1996), these strategies included meta-cognitive strategies (advance organizing, directed attention, selective attention, self-monitoring, functional planning, self-evaluation, and problem identification), cognitive strategies (translation, note taking, outlining, summarizing, imagery, elaboration, and rehearsing, the last of which was chosen from Oxford, 1990), avoidance strategies (topic avoidance and one category from Dornyei, 1995, namely message abandonment) and affective strategies (self-talk) and also three emergent themes in Ortega's (2005) study, namely writing for retrieval, writing for recall, and cross-language monitoring, and finally compensation strategies, grounded in Oxford's (1990) classification, namely circumlocution, approximation, and lexical transfer. To wit, writing for retrieval collapsed into lexical retrieval and grammatical retrieval and also two other categories were considered as avoidance strategies (grammatical and lexical avoidance) which were all studied in order to further elucidate the use of the given superordinate categories.

In addition, a number of items from O'Malley and Chamot's (1996) emergent themes were not dealt in in light of the assumption that they did not pertain to the tightly structured narrative writing which was used in the study (comprehension monitoring, auditory monitoring, personal elaboration, academic elaboration, resourcing, grouping, substitution, between parts elaboration, questioning elaboration, self-evaluative elaboration, creative elaboration, inferencing, questioning for clarification, cooperation, and self-reinforcement), overlapped with other categories (transfer and visual monitoring), or those which could not be applied to strategy use during pre-task planning time (strategy monitoring, plan monitoring, and strategy evaluation) were excluded from the interviews. The same was applied to Oxford's (1990) classification of rehearsing (memory) and compensation strategies. As for the former, physically acting out words, using rhymes, using flashcards, placing new words in a context were excluded due to their assumed irrelevance to a writing investigation. Also, reviewing well and elaboration were also crossed out, as they overlapped with other categories. As regards the latter, guessing, getting help, using mime or gesture, and coining words were excluded, as they were deemed unrelated to the nature of the writing task in the study.

The transcribed interviews were then subjected to thematic content analysis, which was based on the a priori strategies expounded earlier. The researcher coded all the qualitative data. However, the independent coder coded 30% of the data. The intercoder agreement using Cohen's kappa reached .83.

III. RESULTS

In what follows, first the results of the analysis of the participants' output is presented followed by the quantitative and qualitative analysis of the interviews whose main purpose was to tap into the strategies employed by planners during planning time and also to find differences, if any, between pedagogic planners and testing planners in their use of strategies during that time spell.

A. Measures

The descriptive statistics for the measures of grammatical accuracy, syntactic complexity, lexical complexity, and fluency are presented in Table 2.

Dependent variables		Pedagogi	Pedagogic context Testing		context	
		No planning	Planning	No planning	Planning	
Accuracy						
% of EFC	Mean	58.18	61.16	61.83	68.13	
	SD	7.52	6.97	9.88	7.30	
NER	Mean	8.31	7.16	6.88	5.47	
	SD	2.65	2.22	2.32	2.10	
% of TLU	Mean	64.51	71.68	73.66	67.57	
	SD	9.84	9.22	12.47	10.90	
Syntactic complexity						
S-nodes per T-unit	Mean	1.33	1.52	1.28	1.60	
	SD	.17	.18	.16	.14	
Clauses per T-unit	Mean	1.15	1.20	1.17	1.24	
	SD	.07	.07	.05	.08	
Lexical complexity						
MSTTR	Mean	.77	.80	.78	.78	
	SD	.06	.05	.07	.05	
Fluency						
Dysfluencies	Mean	8.18	4.77	8.94	6.17	
	SD	2.88	1.86	2.64	1.74	

DESCRIPTIVE STATISTICS FOR GRAMMATICAL ACCURACY, SYNTACTIC COMPLEXITY, LEXICAL COMPLEXITY, AND FLUENCY ACROSS THE FOUR GROUPS

TABLE 2

Note: EFC = error-free clauses; NER = number of errors per 100 words; TLU = target-like use of English articles; MSTTR = Mean Segmental Type-Token Ratio.

Skewness and kurtosis showed that the data for the percentage of error-free clauses, the number of errors per 100 words, the number of S-nodes per T-unit, MSTTR, and the number of dysfluencies in the text were normal. That leaves the data for the percentage of target-like use of English articles and the number of clauses per T-unit, for which data were found to be nonnormal. Therefore, with regard to the former group, two-way between subjects ANOVAs were conducted (Table 3). As regards the latter, the two-way between subjects ANOVA's equivalent, namely the Kruskal-Wallis test (Table 4) was conducted ensued by the Mann-Whitney U tests (Table 5).

TABLE 3

 $Results \ of \ two-way \ between-subjects \ ANOVAS \ for \ grammatical \ accuracy, \ syntactic \ complexity, \ lexical \ complexity, \ and \ accuracy \ syntactic \ complexity, \ lexical \ complexity, \ and \ syntactic \$

	FLUENCY A	ACROSS THE	FOUR	GROUPS			
Independent variables	Measures	SS	df	MS	F-Value	Sig.	Partial Eta Squared
Planning time	% of EFC	396.377	1	396.377	6.240	.015*	.081
-	NER	30.105	1	30.105	5.633	.020*	.074
	% of TLU	n.a	n.a	n.a	n.a	n.a	n.a
	S-nodes per T-unit	1.155	1	1.155	41.766	.000*	.370
	Clauses per T-unit	n.a	n.a	n.a	n.a	n.a	n.a
	MSTTR	0.006	1	0.006	1.591	.211	.022
	Dysfluencies	145.253	1	145.253	26.555	.000*	.269
Context	% of EFC	518.513	1	518.513	8.163	.006*	.103
	NER	44.611	1	44.611	8.347	.005*	.105
	% of TLU	n.a	n.a	n.a	n.a	n.a	n.a
	S-nodes per T-unit	0.008	1	0.008	0.305	.583	.004
	Clauses per T-unit	n.a	n.a	n.a	n.a	n.a	n.a
	MSTTR	0.000	1	0.000	0.081	.777	.001
	Dysfluencies	0.500	1	0.500	0.091	.763	.001
Planning time *	% of EFC	50.710	1	50.710	0.798	.375	.011
Context	NER	0.323	1	0.323	0.60	.807	.001
	% of TLU	n.a	n.a	n.a	n.a	n.a	n.a
	S-nodes per T-unit	0.089	1	0.089	3.218	.077	.043
	Clauses per T-unit	n.a	n.a	n.a	n.a	n.a	n.a
	MSTTR	0.007	1	0.007	1.671	.200	.023
	Dysfluencies	0.000	1	n.a	n.a	n.a	.000

* = Statistically significant at p < .05

TABLE 4

RESULTS OF KRUSKAL-WALLIS TESTS FOR THE PERCENTAGE OF TLU AND THE NUMBER OF CLAUSES PER T-UNIT ACROSS THE FOUR GROUPS

Dependent variables	CIII-Square	ui	Asymp. sig.
% of TLU	6.261	3	.100
C/T	13.485	3	.004*
		-	

* = Statistically significant at p < .05

RESULTS OF MANN-WHITNEY U TESTS FOR THE PERCENTAGE OF TLU AND THE PERCENTAGE OF C/T ACROSS THE FOUR GROUPS					
Dependent variables	Comparison	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2- tailed)
% of TLU	Planning vs. No planning	682.500	1543.500	154	.877
% 01 1LU	Pedagogic context vs.Testing context	652.000	1247.000	479	.632
C/T	Planning vs. No planning	384.500	979.500	-3.332	.001*
	Pedagogic context vs.Testing context	543.000	1138.000	-1.642	.101
* - Statistically significant at $n < 05$					

 TABLE 5

 Results of Mann-Whitney U tests for the percentage of TLU and the percentage of C/T across the four groups

* = Statistically significant at p < .05

Null Hypothesis 1. Manipulation of pre-task planning time and context does not affect written performance in terms of grammatical accuracy.

As regards the percentage of EFC, a significant main effect for pre-task planning time (F $_{(1,71)} = 6.240$, p = .015) with a moderate effect size ($\eta_p^2 = .081$) was detected. On the other hand, the effect of manipulation of context reached statistical significance (F $_{(1,71)} = 8.163$, p = .006) with a moderate effect size ($\eta_p^2 = .103$). On the other hand, the effect of the interaction between the two independent variables was shown to be nonsignificant (Planning time * Context: F $_{(1,71)} = 0.798$, p = .375).

In terms of NER, a significant main effect for pre-task planning time (F $_{(1, 71)} = 5.633$, p = .020) with an almost moderate effect size ($\eta_p^2 = .074$) was found. Moreover, manipulation of context significant influenced accuracy gains (F $_{(1, 71)} = 8.347$, p = .005) with a moderate effect size ($\eta_p^2 = .105$). To wit, the effect of provision of pre-task planning time did not impinge on the context in which it was provided (Planning time * Context: F $_{(1, 71)} = 0.60$, p = .807).

The last measure of accuracy, the percentage of target-like use of English articles revealed no significant differences across the groups of participants ($x^2 = 6.261$, df = 3, p = .100). Further, two Mann-Whitney U tests were run to gain a better understanding of the differences across the conditions which registered no differences across the groups either as a function of planning time or that of context (z = 682.500, p = .877, z = 652.000, p = .632, respectively).

Accordingly, in light of the substantial effect of manipulation of planning time and context on the accuracy of written output, the first null hypothesis was disconfirmed.

Null Hypothesis 2. Manipulation of pre-task planning time and context does not affect written performance in terms of syntactic complexity.

Pre-task planning time significantly increased the number of S-nodes per T-unit (F $_{(1,71)} = 41.766$, p = .000), an effect which was very large in magnitude ($\eta_p^2 = .370$). However, manipulation of context did not exert a significant influence on syntactic complexity (F $_{(1,71)} = .305$, p = .583). Finally, the effect of the interaction between the two independent variables was shown to be nonsignificant (F $_{(1,71)} = 3.218$, p = .077).

The number of clauses per T-unit, the second measure of syntactic complexity, demonstrated that planners significantly complexified their written output in comparison to no-planners ($x^2 = 13.485$, df = 3, p = .004). Further analysis revealed that increments along planning time significantly fostered syntactic complexity (z = 384.500, p = .001) although the effect of manipulation of context fell shy of significance (z = 543.000, p = .101).

In the wake of the significant effect of manipulation of planning time on the syntactic complexity of written output, the third null hypothesis was disconfirmed.

Null Hypothesis 3. Manipulation of pre-task planning time and context does not affect written performance in terms of lexical complexity.

The lexical complexity measure, MSTTR, did not show any significant main effects as a function of pre-task planning time or context (F $_{(1, 71)} = 1.591$, p = .211, F $_{(1, 71)} = 0.081$, p = .777, respectively). Moreover, the interaction between the two independent variables was found to be nonsignificant (F $_{(1, 71)} = 1.671$, p = .200). Hence, the third hypothesis had to be accepted.

Null Hypothesis 4. Manipulation of pre-task planning time and context does not affect written performance in terms of fluency.

Finally, Two-way between subjects ANOVA showed a significant main effect for pre-task planning time (F $_{(1,71)}$ = 26.555, p = .000) with a very large effect size (η_p^2 = .26). Manipulation of context, on the other hand, did not significantly influence fluency (F $_{(1,71)}$ = 0.091, p = .763). Therefore, the fourth null hypothesis was disconfirmed.

B. The Interviews

In this section, the quantitative analyses of the interviews are presented coupled with tidbits of translated interview responses from the participants, deemed paramount, in quotations.

The interviews were conducted in order to ferret out the metacognitive, cognitive, affective, compensation, and avoidance strategies employed during the pre-task planning phase of the task (Table 6).

Strategy	types	Total sam	nle	Pedagogic	planners	Testing	planners $(n = 23)$
Strategy	cypes	(N = 41)	-pro	(n = 18)		61 ···· ()	
		Raw	Percent	Raw	Percent	Raw	Percent
Metacog	nitive strategies	1					
1	Directed attention	12	29%	8	44%	4	17%
2	Problem identification	15	36%	4	22%	11	48%
3	Advance organizing	19	46%	10	55%	9	39%
4	Functional planning	30	73%	12	67%	18	78%
5	Selective attention	37	90%	16	89%	21	91%
6	Evaluation	29	18%	9	12%	20	22%
6.1	Performance evaluation	9	22%	3	17%	6	26%
6.2	Production evaluation	5	12%	0	0%	5	22%
6.3	Repertoire evaluation	6	15%	4	22%	2	9%
6.4	Ability evaluation	9	22%	2	11%	7	30%
7	Monitoring	112	45%	48	44%	64	46%
7.1	Production monitoring	88	71%	36	67%	52	75%
7.1.1	Vocabulary	20	49%	7	39%	13	56%
7.1.2	Grammar	33	80%	13	72%	20	87%
7.1.3	Error correction	35	85%	16	89%	19	83%
7.2	Cross-language monitoring	9	22%	6	33%	3	13%
7.3	Style monitoring	5	12%	5	28%	0	0%
7.4	Double-check monitoring	10	24%	1	6%	9	39%
Cognitiv	e strategies	10	2.70	-	070	7	0,7,0
1	Note taking	29	71%	11	61%	18	78%
2	Translation	22	54%	11	61%	11	48%
3	Outlining	18	44%	5	28%	13	56%
4	Summarizing	3	7%	3	17%	0	0%
5	Induction/deduction	22	54%	12	67%	10	43%
6	Writing for recall	35	85%	15	83%	20	87%
7	Writing for retrieval	23	28%	11	30%	12	26%
71	Writing for retrieval (lexical items)	18	43%	6	33%	12	52%
7.2	Writing for retrieval (grammatical	5	12%	5	28%	0	0%
/	items)	5	12/0	5	2070	Ű	0,0
8	Imagery	15	36%	5	28%	10	43%
9	Elaboration	8	19%	2	11%	6	26%
10	Rehearsal	53	43%	19	35%	34	49%
10.1	Rereading	34	83%	15	83%	19	83%
10.2	Mental	16	39%	4	22%	12	52%
10.3	Subarticulatory	3	7%	0	0%	3	13%
Affectiv	e strategy		.,.			-	
1	Self-talk	12	29%	3	17%	9	39%
Compen	sation strategies	12	2270	5	1770	7	0,7,0
1	Circumlocution	12	29%	6	33%	6	26%
2	Approximation	7	17%	4	22%	3	13%
3	Lexical transfer	11	27%	2	11%	9	39%
Avoidan	ce strategies	**	2170	-	11/0	1 1	2770
1	Concept avoidance	8	19%	4	22%	4	17%
2	Avoidance (lexical items)	13	32%	4	22%	9	39%
3	Avoidance (grammatical items)	8	19%	2	11%	6	26%
4	Message abandonment	3	7%	3	17%	0	0%
				1 -		~	

TABLE 6 STRATEGIES DOCUMENTED IN THE INTERVIEWS

In general terms, the most documented strategies employed by at least half of all 41 planners included selective attention (90%), writing for recall (85%), error correction (85%), rereading (83%), grammar monitoring (80%), functional planning (73%), production monitoring in general (71%), note taking (71%), translation (54%), induction/deduction (54%), and vocabulary monitoring (49%). In effect, the oft-mentioned strategies were the metacognitive and cognitive ones which direct attention to the primacy of monitoring, planning, rehearsal, and retrieval strategies during pre-task planning time.

With respect to the differences in terms of strategy use between pedagogic and testing planners over the course of task preparation, testing planners more often than not seem to have resorted to more strategies than pedagogic planners with these discrepancies, if 20% difference or higher is deemed large, being strongest in double-check monitoring (33%), mental rehearsal (30%), writing for retrieval of grammatical items (28%), style monitoring (28%), outlining (28%), lexical transfer (28%), directed attention (27%), problem identification (26%), production evaluation (22%), cross-language monitoring (20%), ability evaluation (19%), writing for retrieval of lexical items (19%). The thread of preponderance of monitoring, rehearsal, and retrieval strategies hand in hand with the evaluation ones runs through the participants' interview responses which, in effect, reveal the relatively divergent nature of strategies adopted by testing planners in comparison to the pedagogic planners. In essence, tweaking the nature of a task environment may have exerted an influence on the way planners put their time to use.

As can be seen in Table 6, the metacognitive themes highly punctuated in the interviews were allocation of attention, planning, monitoring, and evaluation. Both groups capitalized on the idea of selective attention over the course of pretask planning time. This "considering the details" was more often than not entwined with the forethought to detect problems; in particular, regarding grammar and vocabulary during performance. Furthermore, it seems that a focus-onform concern in addition to the need to "laden the text with details", to be "organized in writing", to "write in good English" and to "write correctly" prodded the planners into selectively attending to, without being prompted to do so, areas of performance in which they felt they were falling behind. It seems that the demands of the task and the context goaded them into choosing selective allocation of attention, an approach which was more stressed upon by testing planners. On the other hand, a smattering of pedagogic planners reported using directed attention. It seems that these planners had a penchant to merely focus on meaning and message conveyance.

A second oft-mentioned theme was the participants' concern for planning their performance, the comments regarding which ranged from planning the development in the simplest form, i.e., previewing and thinking about the storyline to linking it to their life, in O'Mally and Chamot's (1996) term, what is known as personal elaboration, thinking about the purpose of the task and the goals to achieve (esp., in the case of testing planners), changing the perspective, creative interpretations and concoctions, according to O'Malley and Chamot, what is referred to as creative elaboration, vocabulary, grammar, phrases or chunks that come in handy, and even in one case spelling.

Likewise, self-monitoring figured prominently in the interviews, among which production monitoring found striking expression. In addition to the need to be correct in terms of grammar and lexis, testees took a mention of monitoring as a necessary strategy in exams as it leads to higher scores and more self-confidence, and also "it makes one look good". It seems that the demands of the test environment occasioned a stronger sort of focus on form during pre-task planning time.

The cognitive themes underscored in the interviews revolved around how the planners interacted with the task in advance of task execution, namely note-taking, outlining, summarizing, writing for recall, writing for retrieval, imagery, induction/deduction, translation, and rehearsal.

Being a frequently used strategy by both groups of planners, note-taking was in the main implemented to focus attention on the main content of the picture series, help "gather one's thoughts", "help one remember the main points", and improve the organization of the text. This strategy was more preferred by the testing planners, as they were in the belief that taking notes and memorizing them could help them develop a layout of the whole plotline and expedite their production, a conviction which grew out of a concern for time management and this in its own right lowered their anxiety and upped their self-confidence under the stressful examination conditions.

Outlining was another strategy characteristically favored by testing planners which was opted in order to provide a blueprint for action, as the fact that they were going to be tested seemed to have taken a toll on their attentional resources and prompted them to come up with ways to counter the pressure of task demands which may have been amplified by the manipulation of context. In contrast, pedagogic planners did not "see a point" in outlining. Summarizing, on the other hand, was adopted solely by few pedagogic planners. Other planners in both contexts thought the planning time was almost sufficient to write out the whole task. Hence, they preferred to complete the task during strategic planning time as "a dry run" of the main task so that they could "sail through it" during task execution and in the case of testing planners remember the vocabulary that they had used in their notes more easily during the task proper.

A good deal of planners took to writing for the purpose of recall in advance of task execution. In effect, they believed that planning time helped them recall the storyline better during task execution, boosted their self-confidence, lowered their anxiety, helped them manage time better, helped them take risks over the course of task execution, in particular, with regard to grammatical structures, and helped them unpack the processing load of the task. Writing for recall was more often than not executed in the hope of later recall and that production monitoring could be set in motion, as the output was "in print", thereby rendering task proper easier to manage.

On the other hand, the planners' responses contained remarks of writing for lexical retrieval and also grammatical retrieval to a lesser extent. The former typified testing planners' approach, while the latter was in the main employed by a few pedagogic planners, as other planners in both groups thought if they "had the right words in mind", they could "put them into sentences without thinking too deeply about grammar". Nevertheless, lexical retrieval problems and grammatical retrieval problems were rife. This resonates with Ortega's (1999) ceiling for the effect of planning.

The next strategy was imagery which was doubly preferred by the testing planners in comparison to pedagogic planners. Visualization was thought to help one remember the storyline better and more vividly, engage with the story and relate to it which helped the planners add more details (personal elaboration) and come up with more vocabulary items and more interpretations.

Another type of documented strategy was deduction which was mostly employed by pedagogic planners, as they reported having cogitated grammatical structures while producing the language. This theme was in the main followed by double-check monitoring and reading rehearsal. It seemed that conscious attention to rules raised more doubts about the incorrectness of those rules. However, testees believed that giving too much attention to grammatical rules could have "steered attention away from the message", would have led to more mistakes, and could have "jumbled the

storyline" This might have sprung from their preoccupation with time management and preparation for the upcoming performance.

Translation was also favored more by pedagogic planners than by their testing counterparts. This strategy was usually coupled with production monitoring. Translation happened at word level by all the participants that reported using it, at phrase level by 4 pedagogic planners (36%) and 2 testing planners (10%), and at sentence level by only 2 pedagogic planners (18%).

Finally, the rehearsal strategy was used; particularly, rereading and mental rehearsal in order to facilitate recall and retrieval. The former was utilized to improve development, organization, grammar, and vocabulary, while the latter was used to enhance vocabulary and grammar.

Table 7 reports the descriptive statistics of the five strategy types employed by each pedagogic and testing planner. Overall, out of the approximately 11 or 12 strategies which each planner on average availed themselves of, 6 to 7 were metacognitive strategies, 5 to 6 cognitive ones, 0 to 1 affective strategy, 0 to 1 compensation strategy, and finally 0 to 1 avoidance strategy. In addition, testing planners used more strategies than pedagogic planners with the differences being largest in terms of cognitive and metacognitive strategies.

TABLE 7

STRATEGIES REPORTED PER PLANNING WRITER						
	All strategies	Metacognitive	Cognitive	Affective	Compensation	Avoidance
All participants						
Mean	11.75	6.19	5.56	.29	.73	.78
SD	2.67	1.76	1.37	.46	.80	.85
Pedagogic planners						
Mean	10.27	5.94	5.22	.16	.66	.72
SD	2.24	1.76	1.39	.38	.90	.89
Testing planners						
Mean	12.91	6.39	5.82	.39	.78	.82
SD	2.42	1.77	1.33	.49	.73	.83

In order to find out if these differences were significant, a series of independent-samples t tests was run which
revealed no significant differences between pedagogic and testing planners in their use of metacognitive, cognitive,
affective, compensation, and avoidance strategies (t $_{(1, 39)} = -0.801$, p = .428; t $_{(1, 39)} = -1.408$, p = .167; t $_{(1, 39)} = -1.578$,
$p = .123$; $t_{(1, 39)} = -0.452$, $p = .654$; $t_{(1, 39)} = -0.383$, $p = .704$, respectively). Hence, there does not seem to be any
significant difference between planners in terms of strategy use in pedagogic and testing contexts which renders the last
null hypothesis probable.

IV. DISCUSSION

The findings of the present study almost chime well with those of previous research which have shown standard complexity and fluency gains accrued from opportunities for planning (e.g., Crookes, 1989; Ellis & Yuan; 2004; Foster & Skehan, 1996; Kawauchi, 2005; Ortega, 1999; Sangarun, 2005; Skehan & Foster, 1997, 2005; Tavakoli & Skehan, 2005; Yuan & Ellis, 2003).

In the planning literature, findings with regard to grammatical accuracy have been mixed across studies. In the main, the findings of the present study with respect to planning in the pedagogic context were in tandem with those which have found blurred results for planning on the accuracy of EFL output (e.g., Foster and Skehan, 1996; Ortega, 1999; Gilabert, 2007). Like the pedagogic planners' discourse, testing planners' accuracy of written discourse benefitted from opportunities for planning, a finding which was consonant with those of Tavakoli and Skehan (2005) and Wigglesworth (1997).

As regards the propitious effect of manipulation of context on accuracy of production in terms of percentage of errorfree clauses and the number of errors per 100 words for testees, citing Hulstijn and Hulstijn's (1984) study, Wigglesworth (1997) argues that a testing context may engender a focus-on-form environment in which planners channel their attention toward accuracy (p. 103). Therefore, it might be the case that in a test context, "cognitive capacity is reached" (p. 102), as testees, being more oriented to focus on accuracy of their production, try to rise to the high-stakes occasion. This was further championed by Sangarun (2005) who observed that accuracy can be promoted when learners are accuracy-oriented. Although this proposition may ring true for the present study, this is not to say that testing planners do not direct their attention towards content, either, as the syntactic complexity scores and the data from the interviews speak otherwise. Furthermore, the fact that testing planners reported using more production monitoring than pedagogic planners and that they rehearsed their output more may have led to their significantly higher scores. On the other hand, that the attenuated performance of pedagogic planners in terms of these two accuracy measures may have resulted from the assumption that they may have monitored the accuracy of their production to a smaller extent due to the lower stakes imposed by the situation; hence, pre-task planning time may less facilitate focus on form on a large scale in the pedagogic contexts.

As regards the syntactic complexity of production, the data from the interviews revealed that planning, i.e., previewing, thinking about the story, identifying the purpose, setting goals, and giving own interpretations in addition to writing for recall, outlining (for testing planners), and note-taking abounded in planners' responses with regard to what they did during the strategic planning time. In effect, planners spent their time conceptualizing and practicing what they wanted to say during task execution which might have led to the higher gains in syntactic complexity scores of planners in comparison to their no-planning counterparts. Therefore, it can be postulated that planning time may have given learners some leeway to conceptualize what they wanted to say in terms of content, to ponder on their newly assimilated cutting-edge knowledge of grammar, and to test their hypothesis regarding those grammatical items; and therefore, during task completion they were less burdened by the processing load of the task and were able to deploy those items, a finding which was also corroborated by the syntactic complexity measures.

That both syntactic complexity and grammatical accuracy were promoted seems to run counter to the tradeoffs portended by the limited attentional capacity model. However, a closer look at the data reveals that the effect size for grammatical accuracy (percentage of error-free clauses and number of errors per 100 words) was lower than that for syntactic complexity (number of S-nodes per T-unit) which suggests that tradeoffs between accuracy and syntactic complexity transpired with detrimental consequences for accuracy; Nevertheless, it seems that the learners in the testing context shifted priorities (Tavakoli and Skehan, 2005, p. 268) to focus on accuracy more than pedagogic planners.

The present investigation also found a standard effect for fluency in accord with that of previous research (e.g., Ellis & Yuan, 2004; Foster & Skehan, 1996; Gilabert, 2004; Kawauchi, 2005; Mehnert, 1998; Ortega, 1999; Rouhi & Marefat, 2006; Sangarun, 2005; Skehan & Foster, 2005; Tavakoli & Skehan, 2005; Yuan & Ellis, 2003), as fluency was promoted by the provision of planning time in both contexts.

That the lexical complexity measure did not yield significant results was in line with findings of and Ellis & Yuan (2005), Ortega (1999), and Wigglesworth (1997). In contrast, some studies have found favorable gains for the effect of pre-task planning on lexical complexity (e.g., Crookes, 1989; Foster & Skehan, 1996; Gilabert, 2004; Kawauchi, 2005; Mehnert, 1998; Ortega, 1995, as cited in Ortega, 2005; Yuan & Ellis, 2003).

Another possibility can be entertained which pertains to the nature of writing. Writing seems to impose less cognitive load than speaking does (Skehan, 1998); therefore, even robbed of pre-task planning time, L2 learners may be able to ponder on the retrieval and rehearsal of lexical items which may enable them to be on a par with planners in this respect. This can also be observed in the planning studies conducted to date, as most studies which have found favorable gains for lexical complexity as a result of pre-task planning have been oral studies (e.g., Crookes, 1989; Foster & Skehan, 1996; Gilabert, 2004; Kawauchi, 2005; Mehnert, 1998; Ortega, 1995, as cited in Ortega, 2005; Yuan & Ellis, 2003) with one exception Ellis & Yuan (2004), a writing investigation, which found no significant differences in terms of lexical complexity scores across the groups of participants. Similarly, the performance of the participants of the four groups in the present study did not differ remarkably. This may have arisen from the proposals regarding the nature of the picture task which may have fettered the use of more diverse vocabulary or from the conjecture that there was a ceiling for the retrieval of lexical items (Ortega, 1999, 2005) to which pre-task planning may have been of no avail, since the participants were from the same proficiency level in the study at which they were not deemed to possess a well-stocked battery of lexical items and, as a result, may have encountered lexical retrieval problems. This proposition was further endorsed by the planners' comments in the interviews that although half of the testing planners and onethird of pedagogic planners wrote to facilitate the retrieval of lexical items later on and rehearsed what they wanted to say during pre-task planning time, they failed to lexicalize their output more than no-planners.

The favorable effects of pre-task planning time on fluency of written output might have happened, as learners tried to plan, rehearse, review, and write to remember the content of what they wanted to say during strategic planning time. Therefore, during the on-line performance, learners were less burdened by the processing load of the task and since they were equipped with newly developed plans, they were able to generate more fluent discourse. Building up on these assumptions, the data from the interviews revealed that learners tried to conceptualize what they wanted to say (macroplanning) and formulate what they wanted to say (micro-planning) which may have rendered their production more fluent.

As far as the use of strategies in advance of task execution is concerned, selective attention, writing for recall, error correction, rereading, grammar monitoring, functional planning, production monitoring, note-taking, translation, induction/deduction, and vocabulary monitoring were predominant strategies. It seems that conceptualization, formulation, monitoring, rehearsal, and translation activities were given prominence which points to the predominance of metacognitive and cognitive strategies among writing planners in pedagogic and testing contexts. Despite the fact that these findings partially diverged from those of Ortega (2005) who found writing/outlining/summarizing, production monitoring, functional planning, lexical compensation strategies, translation, empathizing with the listener, and rehearsing to be the oft-used strategies in her oral studies, the preponderance of metacognitive and cognitive strategies over the course of strategic planning time can be traced. Nevertheless, similar to Ortega who found metacognitive and cognitive and cognitive and cognitive and cognitive and cognitive and strategies by planners. This is at odds with O'Malley, Chamot, Stewner-Manzanares, K üpper, and Russo's (1985) finding who found that more metacognitive strategies were used by intermediate learners in comparison to cognitive strategies in writing. The fact that at odds with Ortega's study, affective strategies did not draw much attention among planners in the present

study may have stemmed from the assumption that in oral studies, as Ortega puts it, the presence of an authentic listener puts the use of socioaffective strategies on the front burner which is not the case for the present study, which examined the writing modality. This finding was also in keeping with that of Chamot, O'Malley, Küpper, & Impink-Hernandez (1987), in which the use of socioaffective strategy use was limited at all proficiency levels. Avoidance strategies were also minimally used which is consonant with Ortega's findings. On the other hand, lexical compensation strategies were found to have played a minor role in the formulation of the message during pre-task planning time which runs counter to Ortega's investigation; the storyline of the picture narratives may have warranted the use of different vocabulary which led to the different findings in each study.

As regards the difference in strategy use during pre-task planning time, both groups of planners made use of almost the same number of metacognitive and cognitive strategies. Nonetheless, pedagogic planners' use of strategies partially differed from that of testing planners with the former using remarkably more writing for retrieval of grammatical items, style monitoring, directed attention, and cross-language monitoring and the latter employing sizably more double-check monitoring, mental rehearsal, outlining, lexical transfer, problem identification, production evaluation, ability evaluation, and writing for retrieval of lexical items. These differences may have followed from the different demands of a testing situation and a focus-on-form concern which may have called for one's outlining the message to be conveyed, evaluation of one's ability to complete the task, identification of the problem spots, writing for retrieval of vocabulary items, rehearsal of the message, monitoring, resorting to compensation strategies to make up for their lack of linguistic resources, and checking one's completed work a second time more than pedagogic planners, whereas in the pedagogic context, planners were at liberty to more attend to the sheer content of the stories denuded of details, heed formality/informality of language, monitor the use of Persian-English equivalents, and write to remember the grammatical items more than what transpired during strategic time in the testing context. Therefore, the difference between pedagogic and testing planners in terms of strategy use seems to be more qualitative than quantitative.

V. CONCLUSIONS

The main purpose of the present study was to build up on previous planning research by examining: a) the effect of provision of strategic planning time on accuracy, complexity, and fluency of written production of EFL learners in a pedagogic vs. a testing context; b) the strategy use of EFL learners during strategic planning time in a pedagogic vs. a testing context.

The present study endorses the findings of the foregoing research in terms of discourse-analytic measures of accuracy, complexity, and fluency with respect to the effect of pre-task planning in a pedagogic context. As regards the testing context, it tentatively suggests that accuracy and complexity are also in competition for attention, as although both syntactic complexity and accuracy were promoted as a function of provision of pre-task planning in the study, the effect size for accuracy was smaller than that for syntactic complexity which may speak to the assumption that heightened awareness in a test situation focuses learners' attention on language form. These results, however, run counter to a number of claims and speculations with regard to the effect of testing context on the propitious effects of pre-task planning time on learners' production (Elder & Iwashita, 2005; Iwashita, McNamara, & Elder, 2001) that in testing contexts, the beneficial effects of pre-task planning time are attenuated. On the other hand, it seems that more focus on accuracy is occasioned (Ellis, 2005) and that learners' attention is shifted from complexity towards accuracy (Tavakoli & Skehan, 2005). Furthermore, the findings converge on a limited attentional capacity view of learners in which a competition for attention has been predicted between accuracy and complexity to the detriment of one of the two.

As regards the strategy use of planners, it was found that broadly speaking, planners made use of more metacognitive and cognitive strategies than the other types of strategies with pedagogic planners and testing planners employing approximately the same number of strategies.

The driving force behind the present study was the investigation of the bearings of process-oriented research on providing insights into what transpires over the course of pre-task planning time, the idea whose pacesetter was Ortega who explored the role of pre-task planning time and proficiency level on what learners do when they plan and on their perceptions toward pre-task planning time through retrospective interviews (Ortega, 1995, as cited in Ortega, 1999, Ortega, 1999, 2005). In a related vein, the present investigation champions this research tradition, as it throws more light on how planning works to learners' advantage. A second approach to look into what learners do when they plan would be to use think-alouds and verbal protocol analysis (Kawauchi, 2005). It seems that the process-oriented research holds considerable promise in this regard.

APPENDIX THE TASK

Look at the picture story below starting with picture 1 going through to picture 11. Write the story down.



ACKNOWLEDGEMENT

The author wishes to thank the research assistant for his unstinting assistance in interviewing the participants, the teachers who helped with data collection, and the learners who served as the participants in the study.

REFERENCES

- [1] Alavi, S. M. (2005). On the adequacy of verbal protocols in examining an underlying construct of a test. *Studies in Educational Evaluation*, 31, 1–26.
- [2] Anderson, J. (1983). Architecture of Cognition. Cambridge MA: Harvard University Press.
- [3] Chamot, A., O'Malley, J. M., Küpper, L., & Impink-Hernandez, M. V. (1987). A study of learning strategies in foreign language Instruction: First year report. Rosslyn, VA: InterAmerican Research Associates.
- [4] Cohen, A. D. (1994). Verbal report on learning strategies. TESOL Quarterly, 28(4), 678-682.
- [5] Cohen, J. (1988). Statistical power analysis for the behavioral sciences. New York: Academic Press.
- [6] Crookes, G. (1989). Planning and interlanguage variation. Studies in Second Language Acquisition, 11(4), 367–383.
- [7] Dörnyei, Z. (1995). On the teachability of communication strategies. *TESOL Quarterly*, 29(1), 55–85.
- [8] Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference?. In R. Ellis (Ed.), *Planning and task performance*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 219–238.
- [9] Ellis, R (Ed.). (2005). Planning and task performance in a second language. Amsterdam/ Philadelphia: John Benjamins.
- [10] Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in second Language acquisition*, 26, 59–84.
- [11] Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323.
- [12] Fowler, W. S. & Coe, N. (1976). Nelson English language texts. London: Thomas Nelson and Sons Ltd.
- [13] Gilabert, R. (2004). Task complexity and L2 narrative oral production. Ph.D. dissertation, University of Barcelona.
- [14] Gilabert, R. (2007). The simultaneous manipulation of task complexity along planning time and [+/- Here-and-Now]: Effects on L2 oral production. In M. del Pilar Garcia-Mayo (Ed.), *Investigating tasks in formal language learning*. Clevedon, Avon: Multilingual Matters, 44–68.
- [15] Givon, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition*: Vol 1. Hillsdale, NJ: Lawrence Erlbaum, 1008–1025.

- [16] Huitt, W. (2003). The information processing approach. Educational Psychology Interactive. Valdosta, GA: Valdosta State University. http://www.edpsycinteractive.org/topics/cogsys/infoproc.html (accessed 20, 9, 2008).
- [17] Ishikawa, T. (2007). The effect of manipulating task complexity along the [+/- Here-and-Now] dimension on L2 written narrative discourse. In M. del Pilar Garcia-Mayo (Ed.), *Investigating tasks in formal language learning*. Clevedon, Avon: Multilingual Matters, 136–156.
- [18] Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning*, 51(3), 401-436.
- [19] Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (Ed.), *Planning and task performance*. Amsterdam/ Philadelphia: John Benjamins, 143–164.
- [20] Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). Lexical diversity and language development: Quantification and assessment. New York: Palgrave Macmillan.
- [21] Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 52–83.
- [22] O'Malley, J. M., & Chamot, A. U. (1996). Learning Strategies in Second Language Acquisition. New York: the Press Syndicate of the University of Cambridge.
- [23] O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Kupper, L, & Russo, R. (1985). Learning strategies used by beginning and intermediate ESL students. *Language Learning*, 35, 21–46.
- [24] Ortega, L. (1999). Planning and focus on form in L2 Oral Performance. Studies in Second Language Acquisition, 21, 109–148.
- [25] Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 77–109.
- [26] Oxford, R. (1990). Language learning strategies: What every teacher should know. Heinle & Heinle Publishers.
- [27] Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. Language Learning, 47, 101–143.
- [28] Richards, J., Platt, J., & Weber, H. (1985). Longman dictionary of applied linguistics. London: Longman.
- [29] Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22 (1), 27–57.
- [30] Robinson, P. Ting, S. C. C & Urwin, J. (1996). Three dimensions of second language task complexity. *The University of Queensland Working Papers in Language and Linguistics*,1 (1), 16–32.
- [31] Rouhi, A. & Marefat, H. (2006). Planning time effect on fluency, complexity and accuracy of L2 output. *Pazhuhesh-e Zabanha-ye Khareji*, 27, 123–141.
- [32] Salvador, J. (1991). Humano se nace. Barcelona: Lumen.
- [33] Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 111–141.
- [34] Skehan, P. (1998). A cognitive approach to language learning. Oxford: Oxford University Press.
- [35] Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan, M. Swain (Eds.), Res*earching pedagogic tasks: second language learning, teaching and testing*. Harlow: Pearson Education Limited, 167–185.
- [36] Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1 (3), 185–211.
- [37] Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49 (1), 93–120.
- [38] Skehan, P., & Foster, P. (2001). Cognition and Tasks. In P. Robinson. (Ed.) Cognition and second language instruction. Cambridge: Cambridge University Press, 183–205.
- [39] Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 193–216.
- [40] Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 239–273.
- [41] VanPatten, B. (2002). Processing instruction: An update. Language Learning, 52 (4), 755–803.
- [42] Wendel, J. (1997). Planning and second language narrative production. Ph.D. dissertation, Temple University.
- [43] Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14 (1), 85–106.
- [44] Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). Second language development in writing: Measures of fluency, accuracy, and complexity. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawaii at Manoa.
- [45] Yuan F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics*, 24 (1), 1–27.

Seyed Reza Meraji got his B.A. in the English Language and Literature from the University of Tehran. He got his M.A. in TEFL from the University of Tehran. His areas of interest include task-based language teaching, feedback, and writing.