

Relationships Between Strategy Use, Listening Proficiency Level, Task Type, and Scores in an L2 Listening Test

Yi-Ching Pan
National Pingtung University

Yo In'nami
Chuo University

Abstract

We examined strategy use in relation to L2 proficiency, types of test task, and test performance in listening assessment. A total of 170 Taiwanese university students completed the Test of English for International Communication (TOEIC®) practice listening test and questionnaires designed to measure cognitive and metacognitive strategies. We found that some strategies—voice and imagery inference and elaboration, approaches, and top-down processing strategies—were used with similar frequency regardless of learners' proficiency, while others—planning, monitoring and evaluation, linguistic inference and elaboration, and bottom-up processing—were more often used by advanced listeners. Additionally, planning (and linguistic inference and elaboration, and top-down processing) strategies were more often used in easier tasks. Finally, the relationship between reported strategy use and test scores was weak, accounting for 7% of the total score variance and 5% to 10% of the score variance for each task type section.

Résumé

Nous avons examiné l'usage de stratégies par rapport à la compétence en L2, le type de tâches de test et la performance au test lors d'une évaluation de l'écoute. Au total, 170 étudiants universitaires d'origine taiwanaise ont complété le test d'entraînement pour l'écoute et les questionnaires mesurant les stratégies cognitives et métacognitives du *Test of English for International Communication* (TOEIC®). Nous avons découvert que certaines stratégies—voix, inférence et élaboration par imagerie, approches, traitement ascendant de stratégies—étaient utilisées à la même fréquence quel que soit le niveau de compétence des apprenants, tandis que d'autres—planification, contrôle et évaluation, inférence linguistique et élaboration, traitement ascendant de stratégies—étaient utilisées beaucoup plus fréquemment par les auditeurs avancés. De plus, les stratégies de planification, tout comme celles de l'inférence linguistique et élaboration, et du traitement ascendant, étaient employées plus souvent quand les tâches étaient faciles. En conclusion, la relation entre le report d'usage de stratégies et les scores des tests était faible, expliquant 7 % de la variation du score total et 5 % à 10 % de la variation du score pour chacune des sections du type de tâche.

Relationships Between Strategy Use, Listening Proficiency Level, Task Type, and Scores in an L2 Listening Test

A large number of variables are considered to affect performance and scores on language tests. Bachman (1990) and later Bachman and Palmer (1996) proposed a unified model of language test performance, pointing out four major influences on language test scores: communicative language abilities, test-method facets, test-taker attributes, and random error of measurement. The first of these variables, communicative language ability, is a reflection of the test construct and is, in fact, what the test intends to measure. Test-taker attributes, the second variable, refers to personal characteristics of test takers, including age, gender, culture, and background knowledge, in addition to cognitive characteristics such as strategy use, attitude, anxiety, and motivation. The third variable, test-method facets, involves the testing environment, test rubrics, test input, and test response. Random error, the fourth and final variable, refers to unpredictable and temporary test conditions such as poor health and fatigue. To date, a substantial number of studies have investigated the relationships between test-takers' attributes (e.g., strategy use and cognitive style) and their test scores (e.g., Lai, 2009; Phakiti, 2003; Purpura, 1999; Song, 2005; Vandergrift, 2003; Wu, J. R., 2008; Yang & Plakans, 2012). However, one area that is under researched is strategy use in relation to second language (L2) proficiency, types of test task, and test performance, particularly in listening assessment. The present study aims to fill this gap by examining these variables using the listening section of the Test of English for International Communication (TOEIC®) practice test.

Literature Review

In order to highlight the research gap explored by the current study, this section reviews previous research related to: (a) language strategy use; (b) strategy use and L2 proficiency; (c) strategy use and types of test task; and (d) strategy use, test tasks, and test performance.

Language Strategy Use

Since the 1970s, there has been a fundamental transformation in L2 research, from examining an array of teaching methods to exploring individual learning behaviour in relation to L2 acquisition. Because of this shift, researchers have started to employ qualitative (e.g., interviews, observation, think-aloud methods, and diaries) and/or quantitative (e.g., questionnaires, test scores, and self-rating of language proficiency) methods to identify the range of strategy use adopted by language learners at different proficiency levels (e.g., Chamot, O'Malley, Kupper, & Impink-Hernandez, 1987; O'Malley, Chamot, Stewner-Manzanares, Küpper, & Russo, 1985; Vandergrift, 2003).

Learning strategies, according to Oxford (1990), are "specific actions taken by learners to make learning easier, faster, more enjoyable, more self-directed, more effective, and more transferable to new situations" (p. 8). O'Malley and Chamot's (1990) three types of learning strategies and Oxford's *Strategy Inventory for Language Learning* are the two most widely adopted sources used to develop instruments to investigate the relationship between strategy use and test performance. O'Malley and Chamot identified three categories of learning strategies—cognitive, metacognitive, and social-affective

strategies—whereas Oxford distinguished six categories: memory, cognitive, compensation, metacognitive, affective, and social. In order to simplify these findings, Plonsky (2011) reclassified various strategies into three major types: cognitive, metacognitive, and social strategies. Cognitive strategies refer to learners' "mental activities to use their language and world knowledge to solve the given tasks" (Phakiti, 2003, p. 30). These strategies include inference, evaluation, bottom-up and top-down processing, summarisation, and translation (O'Malley & Chamot, 1990; Oxford, 1990). Metacognitive strategies involve "preparation prior to or reflection following L2 contact or use" (Plonsky, 2011, p. 998), including planning, monitoring, evaluation, and problem identification. These strategies are usually processed to accomplish cognitive tasks (Macaro, Graham, & Vanderplank, 2007; O'Malley & Chamot, 1990). Thus, metacognitive strategy use has "an executive control over the cognitive strategy use" (Purpura, 1999, p. 6). Social strategies generally involve what learners adopt in interpersonal interactions such as asking questions and cooperating with others (Oxford, 1990). Of the three main types of strategy, we will focus on cognitive and metacognitive strategies in this study because we are interested in what strategies learners adopt when taking the TOEIC practice listening test. We do not address social strategies because, as mentioned above, they are usually adopted in interpersonal interactions, not in examinations. Examples of substrategies under the two broad categories can be found in Table 2.

Learners' cognitive and metacognitive strategy use are considered to influence language test performance, and language testers have expressed growing interest in either proposing models (e.g., Bachman, 1990; Bachman & Palmer, 1996; Buck, 2001; Canale & Swain, 1980) or conducting empirical studies that aim to identify their relationships (e.g., Phakiti, 2003; Purpura, 1998; Song & Cheng, 2006). For example, Phakiti (2003) examined how cognitive and metacognitive strategy use was related to the L2 English reading achievement test scores of 384 Thai university students. It was found that 15% to 22% of test score variance was explained by strategy use. Phakiti described the relationship as weak and reasonable, given that strategies play an ancillary role in helping examinees cope with tasks when their routine approaches do not work. In contrast, studies in listening assessment have been limited in number, but probably the most comprehensive model of strategies in listening assessment has been Buck's (2001) framework for describing listening ability, where cognitive and metacognitive strategies are delineated as part of listening constructs. Cognitive strategies are considered to consist of comprehending processes, storing and memorising processes, and using and retrieving processes. Metacognitive strategies are considered to consist of assessing the situation, monitoring one's own performance, self-evaluating, and self-testing. Strategies do not work independently, however; they need to be organised and integrated for successful listening to take place (e.g., Vandergrift, 2003).

Strategy Use and L2 Proficiency

Numerous studies have investigated the relationship between strategy use and L2 proficiency. Although there are variations in the findings of the empirical studies, some agreement has been reached: (a) learners used more metacognitive strategies than cognitive strategies in general (Phakiti, 2003; Purpura, 1998; Song, 2005; Wu, J. R., 2008; Zhang, Liu, Zhao, & Xie, 2011); (b) higher proficiency learners appeared to use a greater number of strategies more effectively than lower proficiency learners (Bacon, 1992a, 1992b; Cross,

2009; Green & Oxford, 1995; Vandergrift, 2003; Vandergrift & Goh, 2009); (c) some strategies had a positive effect on test performance, while others produced a negative contribution (Lai, 2009; Purpura, 1998; Wu, J. R., 2008); and (d) the effect of strategy use on language test performance was weak (Park, 1997; Phakiti, 2003; Song, 2005; Wu, J. R., 2008). For example, Vandergrift (2003) examined the relationship between listening strategy use and proficiency for Grade 7 students (who were at a *novice* level of language proficiency according to the proficiency scale of the American Council of Teachers of Foreign Language [ACTFL]) learning French as an L2. His listening test was in a multiple-choice format. He reported that more advanced learners used a greater number of overall and metacognitive strategies. More specifically, his think-aloud protocol analysis demonstrated that metacognitive strategies on the whole and comprehension monitoring in particular were more frequently used by more skilled listeners, cognitive strategies were used as frequently among more skilled listeners as among less skilled listeners, and two of the cognitive strategies—questioning elaboration and translation strategies—were more frequently used by more skilled listeners. Vandergrift also reported that novice learners used elaboration, inferencing, and transfer to compensate for their lack of proficiency with little resource left for metacognitive strategies such as monitoring, whereas learners of higher proficiency used twice as many metacognitive strategies, suggesting that a certain level of proficiency is an essential prerequisite for learners to employ metacognitive strategies. Similarly, Badger and Yan (2012) found, by using a think-aloud protocol analysis, that two metacognitive strategies—directed attention (i.e., monitoring attention and avoiding distraction) and comprehension monitoring (i.e., checking interpretation for accuracy while listening)—were more often adopted by native speakers of English than by speakers of Chinese when taking an International English Language Testing System (IELTS™) listening test. Various task types were used: multiple choice, gap filling, short answer, and information transfer. In both studies (Vandergrift and Badger and Yan), monitoring strategies were more likely to be used by advanced listeners.

Even though a great number of empirical studies have been conducted to investigate the relationships between strategy use and L2 proficiency, most have used reading tasks or a combination of reading and listening tasks to measure language performance. As Vandergrift (2003, 2007) noted, because the majority of research regarding language learning strategies has devoted attention to reading, writing, and speaking, the understanding of listening comprehension strategies remains limited. To better understand how learners of different levels of listening proficiency employ strategies, Vandergrift called for research into the relationship between strategy use and L2 proficiency, particularly with older learners (older than Grade 7). This leads to research question 1 in our study: How do students at different levels of listening proficiency differ in their use of cognitive and metacognitive strategies across the four task types of the TOEIC practice listening section?

Strategy Use and Types of Test Task

Bachman (1990) presented a framework of test method facets comprising five major categories including testing environment, test rubrics, nature of the input the test taker receives, nature of the expected response to that input, and the relationship between input and response. For the purpose of this discussion, the term *test task* is adopted to refer to the

input text and item types (response mode) that listeners will listen to or read on the test (Brindely & Slatyer, 2002).

Strategy use, a set of complicated and complex behaviours that change depending on the nature of tasks, has been examined in many studies (e.g., Phakiti, 2003; Purpura, 1998; Song, 2005). Among the few studies in listening assessment, two are of particular relevance to our study: Buck (1991) and Y. Wu (1998) examined the test-taking processes of L2 learners answering short-answer or multiple-choice items. Buck (1991) examined the think-aloud and interview data of six novice Japanese learners of English engaging in 54 short-answer items based on a 443-word narrative text. Buck (1991) found that short-answer items were not without problems, due, for example, to learners reaching the correct answer for the wrong reason, particularly in items requiring inference. Additionally, evaluation and monitoring strategies were often unsuccessfully employed as learners tended to assume that they would be tested on the information most recently heard, making them feel it was unnecessary to monitor and update their understanding of an aural text. Similarly, Y. Wu analysed the think-aloud data of four Chinese learners of English answering six multiple-choice items based on a 668-word, 3.5-minute interactive radio interview. It was found again that learners scored correctly for the incorrect reasons but, this time, due to uninformed guessing grounded in answer options. Learners also tended to use top-down strategies when they had failed to construct meaning through the bottom-up processing of a text. This led to successful performance only when they did not over rely on these strategies. These findings from Buck (1991) and Y. Wu demonstrate that strategy use changes depending on task type. Our study extends this line of research, focusing on learners' strategy use in the TOEIC practice listening test. As the TOEIC practice listening test is in a multiple-choice format and consists of four different listening tasks, it could be that a wider range of strategies is employed than those reported in Y. Wu (1998). This leads to research question 2 in our study: What are the relationships of those strategies to TOEIC practice listening test task types?

Strategy Use, Test Tasks, and Test Performance

The literature includes studies pertinent to language learners' performance on tests with different types of test task. In reading comprehension tests, Riley and Lee (1996) found significant differences between the test scores of L2 learners of French received from the types of input text given in the form of summaries in contrast to the form of recalls. Shohamy (1984) also found noticeable differences between different proficiency levels of L2 learners of English on reading tests with regard to the test format (multiple-choice or open-ended questions) and the language of presentation (in the target language or the learner's native language). Her finding suggests that items presented in the learners' native language made the test easier; this was especially apparent for learners with low proficiency. Prabhu (1987) discovered that novice L2 students of English in Indian secondary schools found test tasks with an oral input easier than tasks presented in writing. Yang and Plakans (2012) explored the strategy use of L2 learners of various languages and its relationship to test performance on an integrated reading-listening-writing test task. They found that some strategies positively affected test scores whereas others did not. Fulcher (1996) and McNamara and Lumley (1997) also found differences in the performance of L2 learners of English across different test tasks in speaking tests. Barkaoui, Brooks, Swain, and Lapkin (2013) scrutinised the strategic behaviour of Chinese learners of

L2 English in independent and integrated speaking tests and discovered that integrated tasks elicited a wider variety of reported strategy use than did the independent tasks. However, they found no relationship between the total number of reported strategic behaviours and total test scores, regardless of task type. Although the more skills that were involved in a task led to a greater number of reported strategy uses, the relationships between strategy use, task type, and test performance were varied and complex.

These studies have investigated the relationships between test tasks and test performance in reading, writing, or speaking tests, but very few studies exist that explore listening assessment. This leads to research question 3 in our study: What are the relationships of those strategies to TOEIC practice listening test scores?

In summary, empirical studies on the relationships between strategy use and L2 proficiency level, strategy use and types of test task, and test tasks and test performance, have been primarily devoted to the areas of speaking, writing, or reading skills, with only a few studies in listening. Strategy use may change depending on learners' proficiency levels (Vandergrift, 2003, 2007) and the nature of different tasks, and may contribute differently to language performance (Phakiti, 2003; Purpura, 1998; Song, 2005). Due to the lack of research exploring the relationships between strategy use, L2 proficiency, test tasks, and performance, more studies in this area are needed to better understand the relationships among these variables (Plonsky, 2011).

Context of the Current Study

Since 2003, Taiwan's Ministry of Education has encouraged universities and colleges of technology to set English thresholds for graduates. This is aimed at increasing their level of English proficiency sufficiently to meet the anticipated needs of both domestic and international job markets. The Ministry established a list of recommended tests to fulfill this requirement, which are thought to reflect either the B1 or A2 level of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2014). The list includes the TOEIC, the Test of English as a Foreign Language (TOEFL®), the IELTS, and two local tests: the General English Proficiency Test (GEPT; Taiwan Executive Yuan, 2006) and the College Student English Proficiency Test (CSEPT; Taiwan Executive Yuan, 2006). To address the educational policy of enhancing students' English proficiency for strengthening global competitiveness, as of the 2011 academic year, 39% of technical colleges/universities and 94% of universities had established a graduation benchmark for English. As a result of this policy, many institutes of tertiary education have begun offering courses related to the English proficiency certification tests and have set an objective to increase their students' pass rate on these tests. In addition, statistics collected from the official curricula of tertiary institutions clearly indicate that the number of TOEIC-related courses has increased from 758 in the 2008 academic school year to 1046 in 2009 (Curricula Web, Taiwan Universities of Technology, 2015). Moreover, more than one quarter of Taiwan's 416 largest corporations has adopted TOEIC scores as a criterion for screening and promoting employees (104 Job Bank, 2012), and this has almost certainly encouraged an increased number of university students to take the TOEIC. According to TOEIC Score Data Summaries from 2006 to 2011, the number of university students who took the TOEIC has doubled, from 12,546 to 25,142 (Educational Testing Service [ETS] Taiwan Representative, 2012). In light of these facts, performing better on the TOEIC has become a major concern for both teachers and students in Taiwan. It is essential to

investigate the array of strategies that students utilise on the different parts of the TOEIC to assist English teachers and improve students' English proficiency.

Methods

Participants

The participants in this research study were 170 Taiwanese first-year technological university students enrolled in business-related courses. There were 31 males and 139 females. Three quarters of these participants ($n = 110$) had been studying English as a foreign language in Taiwan for at least 10 years. The institution at which the participants were studying has established English graduation requirements. To meet these requirements, students must choose from an array of English proficiency tests and achieve a certain level on their chosen test in order to graduate. In order to help students pass the TOEIC, which many of them chose to take, TOEIC preparation instruction was offered during their first year of English classes. At the time of this study, the participants had received approximately 7 months of such instruction, 3 hours per week. However, only 14% ($n = 24$) of these students had any experience taking the official TOEIC. The participants were at a *novice high* level of proficiency on the ACTFL scale.

Instruments

The TOEIC practice listening test.

Due to the limited amount of time available in which to complete our study, the participants answered the TOEIC practice minitest of 10 questions each from the Photographs and Question-and-Response parts and 12 questions each from the Conversations and Short Talks parts that were adopted from the textbook *New TOEIC* (Zhang, 2006), rather than the 100 listening questions of the full official TOEIC. In Part I, Photographs, the participants listened to four statements regarding a picture in their test book, and selected one statement that best described what they saw in the picture. In Part II, Question-Response, they listened to a question or statement and three responses, all of which were aurally presented and not printed in their test book. The participants then chose the best response to the question or statement. In Part III, Conversations, they listened to conversations between two people, and answered three questions regarding what the speakers had said in each conversation. In Part IV, Short Talks, the participants listened to talks given by a single speaker, and answered three questions regarding what the speaker had said in each talk. Questions and responses in Parts III and IV were printed in the test book.

Table 1 outlines the structure of the TOEIC practice listening test. The duration of each question and the response time for each part of the test were designed to replicate the official TOEIC. The duration of each question in Parts I and II is the shortest, and the duration of each short talk is the longest. In contrast to the question-response task, the remaining three listening tasks include prompts such as photographs and written questions. As with major standardised language tests, items that appear in the first section are easier and those in the latter sections are more difficult.

Table 1
Structure of the TOEIC Practice Listening Test

Task type	Duration of questions (sec/per question)	Number of questions	Prompt	Response time (sec/per question)
Part I: Photographs	15-17	10	Photographs	5-6
Part II: Question-Response	15-17	10	None	5-6
Part III: Conversations	28-30	12	Written questions & responses	8-9
Part IV: Short Talks	50-55	12	Written questions & responses	8-9

Cognitive and metacognitive strategy questionnaire.

We developed a questionnaire related to the four task types in the TOEIC practice listening test. Questionnaire items were based on O'Malley and Chamot (1990), Oxford's (1990) Strategy Inventory for Language Learning, and Vandergrift's (2003) metacognitive and cognitive strategies questionnaire. The items consisted of those designed to measure cognitive strategies (18 items) and metacognitive strategies (10 items). They were revised as necessary to better suit the needs of the current study. The structure of the items is presented in Table 2 and the actual items are shown in Appendix A. They were originally written in Chinese—the participants' L1—to avoid misunderstanding or misinterpretation. The questionnaire used a 5-point Likert scale: 1 (*never*), 2 (*sometimes*), 3 (*often*), 4 (*usually*), and 5 (*always*). Due to logistical constraints, it was not feasible to use interviews, observation, or think-aloud methods.

Table 2
A Taxonomy of the Cognitive and Metacognitive Strategy Questionnaire

Strategy	Subscale	Items Used
Metacognitive	Planning (directed attention, selective attention, self-management: reading the directions, selecting what to pay more attention to, trying to reduce anxiety, listening with eyes closed)	1, 2, 3, 4, 27, 28
	Monitoring and evaluation (encouraging oneself to keep going when one does not understand the listening text or when one finds oneself not attentive during the test)	21, 26, 29, 30
Cognitive	Linguistic inference (using linguistic information within the text to guess what is unknown)	13, 17
	Linguistic elaboration (using prior knowledge to fill in the missing information)	12, 16, 18
	Voice inference (using tone and intonation to guess the meaning of unknown words/utterances)	11, 15
	Voice elaboration (using tone and intonation to fill in the missing information)	14
	Imagery elaboration (using a mental picture to understand the whole text)	19
	Prediction (predicting what is going to be heard)	20
	Top-down processing (paying attention to the main idea and not the details of the text)	9, 10
	Bottom-up processing (paying attention to individual words/phrases and details)	7, 8
	Translation (translating what is heard into one's native language)	25
	Note taking (taking notes of what is heard)	24
	Summary (summarising what is heard)	23
Words/phrases repetition (repeating the words/phrases heard)	22	

Data Collection Procedure

The data were collected every other week during the 2012 academic year. Over a period of 8 weeks, the participants spent approximately 7 to 10 minutes each answering the four sections of the TOEIC practice listening test in the same order as arranged in the official test, and approximately 10 minutes each responding to the four identical versions of the cognitive and metacognitive questionnaires.

Participants were required to answer each questionnaire with reference to the task they had just completed. For example, after they had finished the 10 listening questions in the Photographs part, they were asked to answer the questionnaire, which read: "Please choose the strategies you have adopted based on the 10 listening questions in the Photographs part you have just finished, and tick them in the appropriate boxes." The different parts of the listening tasks were highlighted with underlining and bolding on the questionnaire, to gain the attention of the participants. Having completed all of the tests and questionnaires, participants were rewarded with a voucher. This ensured that there was no missing or invalid information in the data.

Data Analysis

Data analysis consisted of preliminary and main analyses. A preliminary analysis was conducted in order to group the 170 participants into three proficiency levels based on their total scores on the TOEIC practice listening test. Those whose total scores were ranked within the top one third were considered high-proficiency students ($n = 58$ [34.12%]; scores between 55 and 73 points). Students in the low-proficiency group were those whose total scores were in the bottom third ($n = 59$ [34.71%]; scores between 20 and 43 points). Those whose scores were between the top third and bottom third were considered intermediate-proficiency students ($n = 53$ [31.18%]; scores between 44 and 54 points). Table 3 provides overall descriptive statistics of the total scores on the TOEIC practice listening test. The scores on the Photographs task were the highest, followed by those on the Question-Response and Conversation tasks. The scores on the Short Talk task were the lowest. These gradual decreases in mean scores were expected and desirable as the latter sections in the TOEIC practice listening test included more difficult items.

Table 3

A Breakdown of the TOEIC Practice Listening Test Scores

Task type	<i>M</i>	<i>SD</i>	Min	Max
Photographs	63.94	18.56	10	100
Question-Response	50.06	15.32	10	90
Conversations	48.07	20.28	10	100
Short Talks	33.24	13.53	10	70
Total	48.83	10.24	20	73

Note. In this study, each section was worth 100 points, and the total score was the sum of the scores in the four sections.

A subsequent preliminary analysis, exploratory factor analysis using principal component analysis with varimax rotation, was conducted to examine the structure of the questionnaire. The normality of the score distribution was inspected using skewness and kurtosis values. No outliers were found (skewness ranged from -0.64 to 0.72 and kurtosis from -0.79 to 0.59). The data across the four task types were combined and analysed using factor analysis. The factorability of the items was examined using the Kaiser-Meyer-Olkin measure (i.e., KMO = .89) and Bartlett's test of sphericity (i.e., $p < .001$). Factor loadings lower than .3 (i.e., items 1 and 27) were deleted and were not counted toward any factors. The factor structure of the questionnaires of strategy use, alongside Cronbach's α and the items used, are illustrated in Table 4.

Table 4

Composites of Cognitive and Metacognitive Strategies with Reliability Estimates

Strategy	Factor	Items Used	Reliability	Cumulative %
Metacognitive	Planning	2, 3, 4, 28	.676	
	Monitoring and evaluation	21, 26, 29, 30	.747	17.715
Cognitive	Linguistic inference and Elaboration	12, 13, 16, 17, 18	.740	
	Voice & imagery inference and Elaboration	11, 14, 15, 19, 20	.705	
	Top-down	9, 10	.722	
	Bottom-up	7, 8	.684	
	Approach	22, 23, 24, 25	.817	40.444

Note. For the whole test, the reliability was .886 and the cumulative percentage of the variance explained was 58.159% (17.715% + 40.444%).

Following the two preliminary analyses, we conducted two main analyses. First, to examine research questions 1 and 2, repeated-measures multivariate analysis of variance (MANOVA) was employed to investigate the effects of independent variables (proficiency levels and task types) on dependent variables (metacognitive and cognitive strategy uses). The proficiency levels are a between-subjects variable and the task types and strategies are repeated-measures variables. Since the three proficiency groups were slightly unequal in sample size and Box's test of equality of covariance matrices were significant ($p = .003$), suggesting the inequality of those matrices of the dependent variables, we used Pillai's criterion to interpret results from the MANOVA analyses (Tabachnick & Fidell, 2007). Significant interaction effects were decomposed and analysed using analysis of variance (ANOVA). To examine research question 2, the strategy by task interaction effects from the MANOVA analysis above were examined. For both research questions, analyses were conducted using the means of the totals of frequencies for all items for each factor and not

factor scores. This was because the former method retains the metric of a scale and is easy to calculate and interpret. Furthermore, as an effect size measure, partial eta squared was calculated to examine the proportion of variance attributable to a particular variable when excluding the effects of other variables in the analysis. Moreover, sphericity was not satisfied in most of the cases as judged by Mauchly's test, and we reported Greenhouse-Geisser corrected estimates of degrees of freedom. Post-hoc pairwise comparisons were tested using the Bonferroni correction.

Second, stepwise regression was utilised to examine research question 3. We used stepwise regression because it was not clear from previous studies which strategy was more important and thus should be first included in the regression. This choice made the current study exploratory and model building, and limited the generalisability of the findings. Again, we used the means of the totals of frequencies for all items for each factor for the same reason as above. Further, as an effect size measure, the *R* squared values of .02, .13, and .26 were interpreted as small, medium, and large effects, respectively (Cohen, 1988). All data for research questions 1 to 3 were analysed using PASW Statistics Version 18.

Results

Descriptive Statistics

Table 5 illustrates students' reported use of cognitive and metacognitive strategies. The means of the seven strategies across the four task types ranged from highest (3.72) for planning to lowest (2.83) for approaches, with 3.64 for bottom-up processing, 3.61 for top-down processing, 3.60 for monitoring and evaluation, 3.43 for linguistic inference and elaboration, 3.05 for voice and imagery inference and elaboration, and 2.83 for approaches. The participants as a whole reported using all seven types of strategy with quite similar frequency across the four listening tasks, while a closer examination demonstrates that planning strategies were used most frequently and that approaches strategies were employed the least often.

Cognitive and Metacognitive Strategy Use and Listening Proficiency

A repeated-measures MANOVA test was conducted to test the effect of proficiency levels on strategy use across four different task types. Table 6 illustrates that there were significant multivariate effects for task types, strategy use, proficiency levels, the interaction between strategy use and proficiency levels, and the interaction between strategy use and task types. There was no significant interaction between task types and proficiency levels or between task types, strategy use, and proficiency levels.

Table 5

Distributions for the Cognitive and Metacognitive Strategies Employed by Students at Different Proficiency Levels Across the Four Listening Tasks

Test Task	Strategy Use	M				SD			
		Low n = 59	Mid n = 53	High n = 58	Total	Low n = 59	Mid n = 53	High n = 58	Total
Photographs	P	3.53	3.86	3.81	3.73	.64	.64	.64	.65
	ME	3.40	3.76	3.83	3.66	.68	.56	.60	.64
	LIE	3.34	3.48	3.63	3.48	.69	.53	.55	.60
	VIIE	2.99	3.01	3.16	3.05	.52	.61	.61	.58
	TDP	3.49	3.82	3.81	3.70	.92	.81	.76	.84
	BUP	3.49	3.63	3.87	3.66	.75	.79	.69	.75
	A	2.75	3.00	2.98	2.91	.71	.74	.82	.76
Question-Response	P	3.62	3.91	3.96	3.83	.65	.54	.59	.61
	ME	3.28	3.66	3.80	3.58	.68	.58	.58	.65
	LIE	3.29	3.49	3.57	3.45	.62	.60	.65	.63
	VIIE	2.98	3.05	3.11	3.05	.68	.63	.64	.65
	TDP	3.29	3.75	3.83	3.62	.83	.80	.81	.84
	BUP	3.50	3.67	3.92	3.70	.74	.82	.68	.76
	A	2.62	2.83	2.84	2.76	.65	.71	.65	.67
Conversations	P	3.43	3.73	3.92	3.67	.74	.82	.68	.66
	ME	3.36	3.61	3.71	3.56	.64	.58	.60	.62
	LIE	3.26	3.44	3.58	3.42	.46	.58	.64	.58
	VIIE	3.04	3.00	3.18	3.08	.66	.63	.75	.69
	TDP	3.31	3.64	3.85	3.60	.72	.73	.73	.76
	BUP	3.37	3.56	3.78	3.57	.71	.74	.72	.74
	A	2.79	2.83	2.89	2.83	.61	.63	.73	.66
Short Talks	P	3.53	3.54	3.73	3.60	.58	.52	.76	.63
	ME	3.33	3.66	3.69	3.55	.69	.62	.71	.69
	LIE	3.24	3.40	3.37	3.33	.56	.50	.67	.58
	VIIE	3.07	2.93	3.04	3.00	.74	.65	.70	.70
	TDP	3.25	3.54	3.71	3.50	.82	.73	.74	.79
	BUP	3.49	3.60	3.75	3.61	.75	.69	.81	.76
	A	2.73	2.82	2.81	2.78	.72	.73	.72	.72
Total	P	3.53	3.77	3.85	3.72	.63	.59	.67	.63
	ME	3.35	3.68	3.76	3.60	.67	.59	.63	.63
	LIE	3.29	3.46	3.54	3.43	.59	.56	.64	.60
	VIIE	3.02	3.00	3.13	3.05	.66	.63	.68	.66
	TDP	3.34	3.69	3.81	3.61	.83	.77	.77	.79
	BUP	3.47	3.62	3.84	3.64	.74	.76	.73	.74
	A	2.72	2.88	2.88	2.83	.68	.71	.74	.71

Note. P = Planning, ME = Monitoring and evaluation, LIE = Linguistic inference and elaboration, VIIE = Voice and imagery inference and elaboration, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

Table 6
Multivariate Tests

Effect	<i>V</i>	<i>F</i>	Hypothesis <i>df</i>	Error <i>df</i>	<i>p</i>	η_p^2
Task	.09	5.43	3.00	165.00	.001	.09
Task x Proficiency Level	.049	1.37	6.00	332.00	.225	.02
Strategy	.77	91.54	6.00	162.00	< .001	.77
Strategy x Proficiency Level	.13	1.81	12.00	326.00	.046	.06
Strategy x Task	.23	2.52	18.00	150.00	.001	.23
Task x Strategy x Proficiency Level	.18	.82	36.00	302.00	.755	.09
Proficiency Level	75.60	8.21	2.00	167.00	< .001	.09

The interaction effects were decomposed into simple effects to identify the source of the interaction. With regard to the interaction between strategy use and proficiency levels, significant differences were found in the frequency of strategy use for each group: $F(5.08, 1194.82) = 63.02, p < .001, \eta_p^2 = .21$ for the low proficiency group; $F(4.72, 995.75) = 91.20, p < .001, \eta_p^2 = .30$ for the intermediate proficiency group; and $F(5.46, 1262.22) = 119.02, p < .001, \eta_p^2 = .34$ for the high proficiency group. A post-hoc test in Table 7 shows that the frequency of reported strategy use differed in most cases among the strategy types according to learner proficiency. For example, planning was more often used than the remaining strategies, with the exception of bottom-up processing, among low-proficiency learners. In contrast, the prevalent use of planning declined among intermediate- and high-proficiency learners, and it was used as often as monitoring and evaluation and top-down processing strategies.

Table 7
The Post Hoc Test of Differences in Strategy Use

(I) strategy	(J) strategy	Low			Intermediate			High		
		Mean Difference (I-J)	SE	<i>p</i>	Mean Difference (I-J)	SE	<i>p</i>	Mean Difference (I-J)	SE	<i>p</i>
P	ME	.18	.04	< .001	.09	.046	.617	.08	.04	.536
	LIE	.24	.05	< .001	.31	.04	< .001	.30	.05	< .001
	VIIE	.52	.04	< .001	.76	.05	< .001	.72	.05	< .001
	A	.81	.05	< .001	.89	.06	< .001	.96	.05	< .001
	TDP	.19	.06	.015	.08	.05	1.000	.04	.05	1.000
	BUP	.07	.05	1.000	.15	.06	.191	.01	.05	1.000
ME	LIE	.06	.04	1.000	.22	.04	< .001	.22	.04	< .001
	VIIE	.33	.04	< .001	.67	.05	< .001	.63	.05	< .001
	A	.62	.04	< .001	.80	.06	< .001	.88	.04	< .001
	TDP	.01	.06	1.000	-.01	.05	1.000	-.05	.05	1.000
	BUP	-.12	.05	.201	.06	.03	1.000	-.08	.05	1.000
LIE	VIIE	.27	.05	< .001	.45	.04	< .001	.42	.04	< .001
	A	.56	.05	< .001	.58	.05	< .001	.66	.05	< .001
	TDP	-.05	.06	1.000	-.24	.05	< .001	-.26	.05	< .001
	BUP	-.18	.05	.016	-.16	.05	.036	-.29	.05	< .001
VIIE	A	.29	.05	< .001	.13	.06	.430	.24	.05	< .001
	TDP	-.32	.06	< .001	-.69	.056	< .001	-.68	.06	< .001
	BUP	-.45	.05	< .001	-.62	.06	< .001	-.71	.06	< .001
A	TDP	-.61	.06	< .001	-.82	.07	< .001	-.92	.06	< .001
	BUP	-.74	.05	< .001	-.75	.06	< .001	-.93	.05	< .001
TDP	BUP	-.13	.06	1.000	.07	.06	1.000	-.03	.06	1.000

Note. P = Planning, ME = Monitoring and evaluation, LIE = Linguistic inference and elaboration, VIIE = Voice and imagery inference and elaboration, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

Further, significant differences were found in the frequency of strategy use for most of the tasks: $F(2, 167) = 4.37, p = .014, \eta_p^2 = .05$ for planning; $F(2, 167) = 7.86, p = .001, \eta_p^2 = .09$ for monitoring and evaluation; $F(2, 167) = 3.21, p = .043, \eta_p^2 = .04$ for linguistic inference and elaboration; $F(2, 167) = 1.49, p = .229, \eta_p^2 = .02$ for voice and imagery inference and elaboration; $F(2, 167) = 2.01, p = .137, \eta_p^2 = .02$ for approaches; $F(2, 167) = 2.96, p = .055, \eta_p^2 = .03$ for top-down processing; and $F(2, 167) = 4.06, p = .019, \eta_p^2 = .05$ for bottom-up processing. A post-hoc test in Table 8 shows that the frequency of reported strategy use differed across learners of different proficiency according to types of strategy. For example, voice and imagery inference and elaboration, approaches, and top-down processing strategies were used with similar frequency across learners of different proficiency, whereas planning, monitoring and evaluation, linguistic inference and elaboration, and bottom-up processing were more often used by higher proficiency learners. The results are visually presented in Figure 1.

Table 8

The Post Hoc Test of Differences in Strategy Use Across the Three Proficiency Levels

Strategy	Proficiency Levels (I-J)	Mean Difference (I-J)	SE	p
Planning	Low-Mid	-.33	.12	.023
	Low-High	-.28	.12	.059
	Mid-High	.05	.12	1.000
Monitoring and evaluation	Low-Mid	-.36	.12	.008
	Low-High	-.43	.12	.001
	Mid-High	-.06	.12	1.000
Linguistic inference and elaboration	Low-Mid	-.14	.11	.686
	Low-High	-.28	.11	.037
	Mid-High	-.14	.11	.627
Voice & imagery inference and elaboration	Low-Mid	-.02	.11	1.000
	Low-High	-.17	.11	.337
	Mid-High	-.15	.11	.531
Approach	Low-Mid	-.26	.14	.237
	Low-High	-.24	.14	.282
	Mid-High	.02	.15	1.000
Top-down processing	Low-Mid	-.33	.16	.118
	Low-High	-.33	.16	.108
	Mid-High	.00	.16	1.000
Bottom-up processing	Low-Mid	-.14	.14	.959
	Low-High	-.39	.14	.016
	Mid-High	-.25	.14	.247

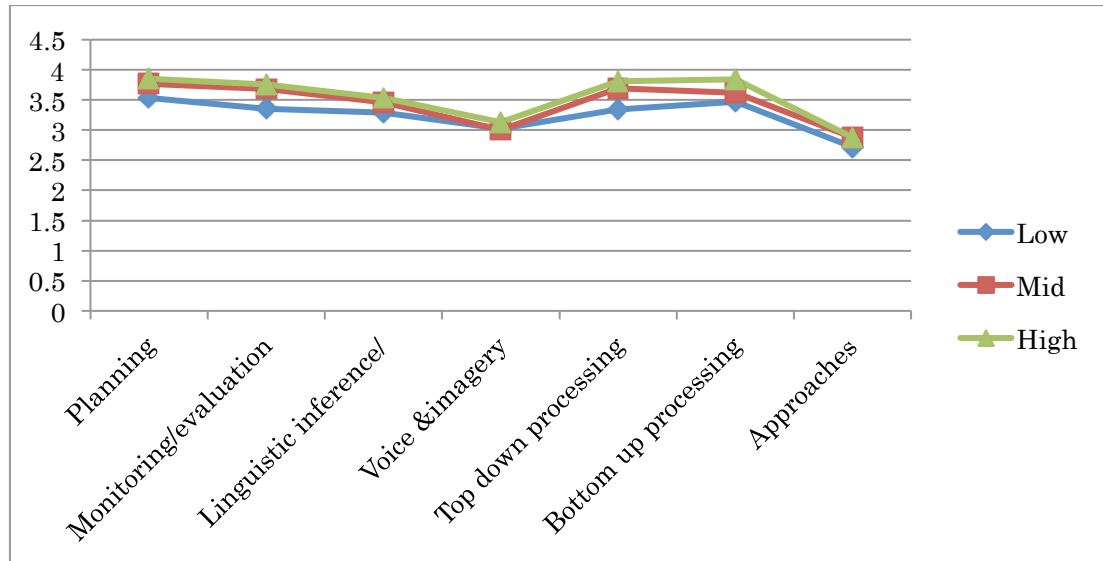


Figure 1. Interactional effects between strategy use and proficiency levels.

Cognitive and Metacognitive Strategy Use and TOEIC Listening Task Types

With regard to the interaction between strategy use and task types as seen in the MANOVA analysis, significant differences were found in the frequency of strategy use for each task type: $F(5.15, 870.61) = 58.42, p < .001, \eta_p^2 = .26$ for the Photographs tasks; $F(5.25, 887.71) = 82.02, p < .001, \eta_p^2 = .33$ for the Question-Response tasks; $F(5.23, 883.85) = 64.87, p < .001, \eta_p^2 = .28$ for the Conversation tasks; and $F(5.18, 875.82) = 60.03, p < .001, \eta_p^2 = .26$ for the Short Talk tasks. A post-hoc test in Table 9 shows that the frequency of reported strategy use was, in most cases, different across tasks. For example, in the Photographs tasks, planning was more often used than linguistic inference and elaboration, voice and imagery inference and elaboration, and approaches. This was also true in the remaining three tasks. Additionally, in the Question-Response tasks, planning was more often used than monitoring and evaluation, and top-down processing.

Table 9
The Post Hoc Test of Differences in Strategy Use

(I) strategy	(J) strategy	Photographs			Question-Response			Conversations			Short Talks		
		Mean Difference (I-J)	SE	<i>p</i>	Mean Difference (I-J)	SE	<i>p</i>	Mean Difference (I-J)	SE	<i>p</i>	Mean Difference (I-J)	SE	<i>p</i>
P	ME	.07	.05	1.000	.25	.05	< .001	.12	.04	.132	.05	.05	1.000
	LIE	.24	.06	< .001	.38	.06	< .001	.25	.05	< .001	.27	.05	< .001
	VIIE	.67	.06	< .001	.78	.05	< .001	.60	.05	< .001	.60	.06	< .001
	A	.82	.07	< .001	1.07	.06	< .001	.84	.06	< .001	.82	.06	< .001
	TDP	.03	.07	1.000	.21	.07	.033	.08	.05	1.000	.10	.06	1.000
	BUP	.07	.07	1.000	.13	.06	.583	.11	.06	1.000	-.01	.06	1.000
ME	LIE	.18	.05	.016	.13	.05	.144	.13	.05	.079	.22	.05	< .001
	VIIE	.61	.06	< .001	.53	.06	< .001	.48	.05	< .001	.55	.06	< .001
	A	.75	.06	< .001	.82	.06	< .001	.72	.05	< .001	.77	.06	< .001
	TDP	-.04	.07	1.000	-.04	.07	1.000	-.04	.06	1.000	.06	.06	1.000
	BUP	-.00	.06	1.000	-.12	.06	.883	-.01	.06	1.000	-.06	.06	1.000
LIE	VIIE	.43	.05	< .001	.40	.05	< .001	.35	.05	< .001	.33	.05	< .001
	A	.58	.06	< .001	.69	.06	< .001	.59	.05	< .001	.55	.06	< .001
	TDP	-.22	.06	.018	-.17	.07	.206	-.17	.06	.056	-.16	.06	.198
	BUP	-.18	.06	.093	-.25	.06	.001	-.14	.06	.199	-.28	.06	< .001
VIIE	A	.15	.06	.327	.29	.06	< .001	.24	.06	.001	.22	.06	.008
	TDP	-.65	.07	< .001	-.57	.07	.001	-.52	.06	< .001	-.50	.06	< .001
	BUP	-.61	.07	< .001	-.65	.07	< .001	-.49	.06	< .001	-.61	.07	< .001
A	TDP	-.80	.08	< .001	-.86	.07	< .001	-.76	.07	< .001	-.72	.07	< .001
	BUP	-.76	.07	< .001	-.94	.06	< .001	-.73	.06	< .001	-.83	.06	< .001
TDP	BUP	.04	.07	1.000	-.08	.07	1.000	.03	.06	1.000	-.12	.08	1.000

Note. P = Planning, ME = Monitoring and evaluation, LIE = Linguistic inference and elaboration, VIIE = Voice and imagery inference and elaboration, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

Further, significant differences were found in the frequency of strategy use for most of the tasks: $F(3, 507) = 9.91, p < .001, \eta_p^2 = .06$ for planning; $F(2.54, 428.68) = 2.79, p = .049, \eta_p^2 = .02$ for monitoring and evaluation; $F(2.74, 462.83) = 3.99, p = .010, \eta_p^2 = .02$ for linguistic inference and elaboration; $F(2.80, 472.74) = 1.06, p = .363, \eta_p^2 = .01$ for voice and imagery inference and elaboration; $F(2.76, 466.20) = 3.51, p = .018, \eta_p^2 = .02$ for approaches; $F(2.57, 434.54) = 3.18, p = .031, \eta_p^2 = .02$ for top-down processing; and $F(2.75, 465.10) = 1.78, p = .154, \eta_p^2 = .01$ for bottom-up processing. A post-hoc test in Table 10 shows that the frequency of reported strategy use was similar across tasks except in a few cases. For example, planning was more often used in the Photographs tasks than in the Short Talk tasks, in the Question-Response tasks than in the Conversation tasks, and in the Question-Response tasks than in the Short Talk tasks. Planning, linguistic inference and elaboration, and top-down processing strategies were consistently more often used in the Photographs tasks than in the Short Talk tasks. The results are visually presented in Figure 2.

Cognitive and Metacognitive Strategy Use and TOEIC Listening Test Scores

A stepwise regression analysis was conducted to investigate whether listening strategies had an effect on TOEIC listening scores. Table 11 indicates that monitoring and evaluation and top-down processing strategies had a significant and small effect on the prediction of TOEIC listening total scores, accounting for 7% of the total variance. The remaining regression models explained 5% to 10% of the scores for each task type section except the Short Talks. The significant predictor of the Photographs section score was monitoring and evaluation strategies. These strategies also had a positive effect on the Question-Response scores, whereas approaches demonstrated a negative effect on those scores. The significant contributor to the Conversation scores was the bottom-up processing strategy. There was no significant predictor contributing to the Short Talk score, which implies the relationship between strategy use and test performance on the Short Talks was the weakest.

It should be noted that top-down processing was a significant predictor of total scores, but not of the scores of any individual task type. This was probably due to our decision to use stepwise regression. As stepwise regression is a procedure where the independent variable that most highly correlates with the dependent variable is entered into the equation first, other independent variables that correlate less with the dependent variable are omitted from the equation. In the Photographs tasks, for example, monitoring and evaluation strategy most strongly correlated with the section score ($r = .25$) and was entered into the regression (see Appendix B). Planning—the next highly correlated variable ($r = .20$)—was not entered into the regression. This meant that planning remained an important predictor but less so than monitoring and evaluation. See Appendices C to F for the remaining correlations matrices between each strategy and scores.

Table 10
The Post Hoc Test of Differences in Strategy Use Across the Four Task Types

Strategy	Listening Tasks (I-J)	Mean Difference (I-J)	SE	<i>p</i>
Planning	Photo-QR	-.10	.04	.131
	Photo-Conversation	.05	.05	1.000
	Photo-Short talk	.13	.04	.022
	QR-Conversation	.15	.04	.003
	QR-Short talk	.23	.04	< .001
	Conversation-Short talk	.07	.04	.325
Monitoring and evaluation	Photo-QR	.08	.04	.243
	Photo-Conversation	.10	.05	.207
	Photo-Short talk	.11	.05	.166
	QR-Conversation	.02	.04	1.000
	QR-Short talk	.02	.04	1.000
	Conversation-Short talk	.00	.03	1.000
Linguistic inference and elaboration	Photo-QR	.03	.04	1.000
	Photo-Conversation	.06	.05	1.000
	Photo-Short talk	.15	.05	.009
	QR-Conversation	.03	.05	1.000
	QR-Short talk	.12	.04	.047
	Conversation-Short talk	.09	.04	.116
Voice & imagery inference and elaboration	Photo-QR	.01	.04	1.000
	Photo-Conversation	-.02	.05	1.000
	Photo-Short talk	.05	.05	1.000
	QR-Conversation	-.03	.04	1.000
	QR-Short talk	.04	.04	1.000
	Conversation-Short talk	.07	.04	.262
Approach	Photo-QR	.15	.05	.013
	Photo-Conversation	.07	.05	1.000
	Photo-Short talk	.12	.06	.172
	QR-Conversation	-.08	.05	.698
	QR-Short talk	-.02	.05	1.000
	Conversation-Short talk	.05	.04	1.000
Top-down processing	Photo-QR	.08	.06	1.000
	Photo-Conversation	.11	.08	1.000
	Photo-Short talk	.20	.08	.043
	QR-Conversation	.02	.07	1.000
	QR-Short talk	.12	.06	.300
	Conversation-Short talk	.10	.05	.453
Bottom-up processing	Photo-QR	-.04	.07	1.000
	Photo-Conversation	.09	.07	.898
	Photo-Short talk	.05	.07	1.000
	QR-Conversation	.13	.05	.090
	QR-Short talk	.09	.06	.919
	Conversation-Short talk	-.04	.05	1.000

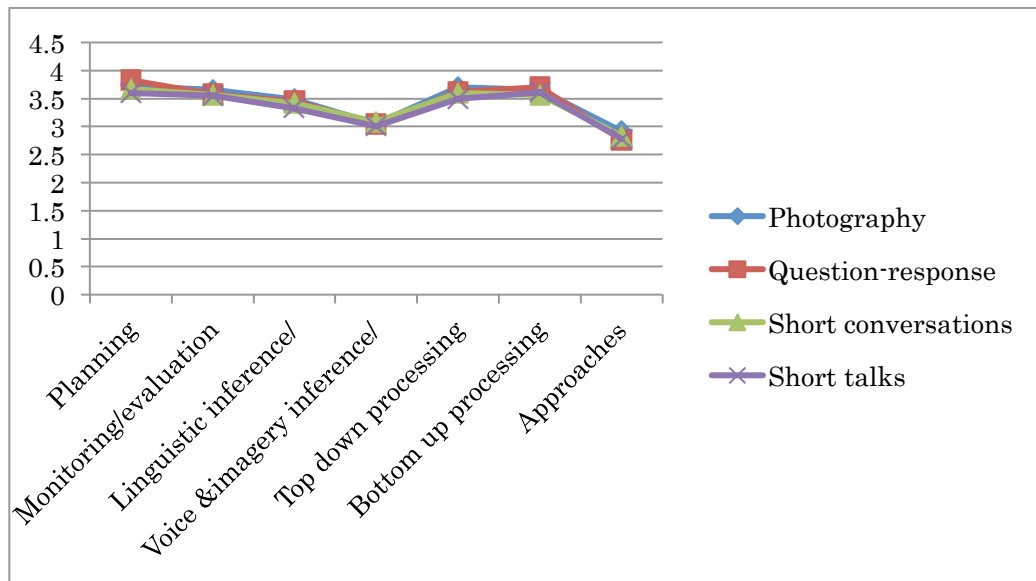


Figure 2. Interactional effects between strategy use and task types.

Table 11

A Model Summary for the Four TOEIC Listening Task Types

Task type	Strategy use	R^2	B	β	t	p
Photographs	(constant)	.06	37.69		4.74	< .001
	ME		7.16	.25	3.35	.001
Question-Response	(constant)	.10	36.11		7.85	< .001
	ME		7.50	.32	4.08	< .001
	A		-4.67	.21	-2.63	.009
Conversations	(constant)	.05	26.93		3.58	< .001
	BUP		5.92	.22	2.87	.005
Short Talks	(constant)	.02	34.38		4.54	< .001
	(constant)	.07	32.59		14.13	< .001
	ME		3.06	.20	4.88	.001
	TDP		1.46	.12	2.89	< .001

Note. ME = Monitoring and evaluation, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

Discussion

This study aimed to explore the relationships between strategy use, L2 proficiency, task types, and listening test performance. Over a period of 8 weeks, 170 technological university students in Taiwan completed the four sections of the TOEIC practice listening test and four identical questionnaires. The questionnaires were designed to measure cognitive and metacognitive strategy use and were adopted from O'Malley and Chamot (1990), Oxford (1990), and Vandergrift (2003).

Research Question 1: How Do Students at Different Levels of Listening Proficiency Differ in Their Use of Cognitive and Metacognitive Strategies Across the Four Task Types of the TOEIC Practice Listening Section?

As interaction effects were found between strategy use and proficiency levels, we deconstructed these effects and found that reported strategy use differed in frequency among the type of strategies according to learner proficiency. More specifically, we found that bottom-up processing was used most often among low-proficiency learners, followed by planning. However, among intermediate- and high-proficiency learners, planning, bottom-up processing, monitoring and evaluation, and top-down processing strategies were all used equally.

Of greater interest were analyses of each strategy use according to proficiency level. We found that the use of each strategy differed depending on learners' proficiency. For example, voice and imagery inference and elaboration, approaches, and top-down processing strategies were used with similar frequency regardless of learners' proficiency. In contrast, planning, monitoring and evaluation, linguistic inference and elaboration, and bottom-up processing were more often used by higher proficiency learners. These results partly concur with Vandergrift (2003), who reported that metacognitive strategies on the whole and comprehension monitoring in particular were more frequently used by more skilled listeners and that more advanced learners used a greater number of overall and metacognitive strategies. Consistent with Vandergrift (2003), planning and monitoring and evaluation were metacognitive strategies that were more likely to be used by advanced learners in the current study. Vandergrift's (2003) participants were Grade 7 students learning French as an L2 whereas the participants in the current study were first-year university students learning English as an L2. However, participants in both studies were at a *novice high* level of proficiency on the ACTFL scale. Obtaining similar results across the two studies appears to indicate the generalisability of the finding—planning and monitoring and evaluation are metacognitive strategies that more advanced learners are more likely to use.

Research Question 2: What are the Relationships of Those Strategies to TOEIC Practice Listening Test Task Types?

Because interaction effects were found between strategy use and task types, we examined these effects and found that reported strategy use differed in frequency according to task types. More specifically, in the Photographs tasks, planning was more often used than linguistic inference and elaboration, voice and imagery inference and elaboration, and approaches. This was also true in the remaining three tasks. This frequent use of planning strategy compared with other strategies suggests the relative importance of planning when engaging in listening tasks. Additionally, in the Question-Response tasks, planning was more often used than monitoring and evaluation and top-down processing.

More importantly, we found that the frequency of reported strategy use was similar across tasks except in a few cases: Planning strategy was more often used in the Photographs tasks than in the Short talk tasks, in the Question-Response tasks than in the Conversation tasks, and in the Question-Response tasks than in the Short Talk tasks. Recall that the four task types were easier in the order of Photographs, Question-Response, Conversation, and Short Talk. This suggests that planning strategy was more often used when participants worked on easier task types. Similarly, planning, linguistic inference and elaboration, and top-down processing strategies were

consistently more often used in the Photographs tasks than in the Short Talk tasks. Given that the Photographs tasks were the easiest and the Short Talk tasks the most difficult, it seems reasonable to infer that planning, inference/elaboration, and top-down processing strategies were more likely to be used in easier tasks.

The finding that planning (and linguistic inference and elaboration, and top-down processing) strategies were more likely to be used in easier tasks seems contradictory to the study reported in Barkaoui et al. (2013), in which test-takers tended to have a higher frequency of reported strategy use when faced with difficult tasks. One possible reason could be that when facing a task that test-takers believed they could handle, they would attempt to complete it using various strategies; when the task demand far exceeded their proficiency, however, they would focus on the strategies that they could make the most sense of, such as planning, bottom-up/top-down processing, and monitoring and evaluation strategies. In other words, participants reverted to less demanding strategies when faced with more difficult listening tasks, as was found in studies conducted by Bacon (1992a, 1992b). Bacon (1992a) discovered that “when faced with faster speech and a less salient topic, they [the participants] responded by retreating into much less cognitively demanding processing strategies” (p. 408).

Research Question 3: What are the Relationships of Those Strategies to TOEIC Practice Listening Test Scores?

The relationship between reported strategy use and test scores was rather weak, accounting for 7% of the total score variance ($R^2 = .07$, see Table 11) while the effects of the other variables remained constant. The remaining regression models explained 5% to 10% of the scores for each task type section except the Short Talks. This result is consistent with Phakiti (2003), Purpura (1997, 1998), Song (2005), and Barkaoui et al. (2013). This finding is understandable because strategy use is merely one variable that affects test performance; other variables are communicative language ability and various personal characteristics. Meanwhile, test methods and random measurement errors all affect learners' test performance (Bachman, 1990). In particular, certain strategies had positive and other strategies had negative effects on students' listening scores on certain task types. For example, when students had a higher frequency use of the monitoring and evaluation strategy, they were likely to refuse to give up when they noticed their concentration flagging during the test, evaluate how much they understood about the test, and attempt to determine problems they faced in the test. This led to better performance on the whole listening test, the Photographs task, and the Question-Response task. Another example is that the bottom-up processing strategy, where students paid attention to the individual words, phrases, and details of the text, positively affected their scores on the Conversation task, whereas the top-down processing strategy, where students paid attention only to the main idea, positively affected their scores on the whole TOEIC practice listening test. The approach strategy, in which students translated, repeated, summarised, and took notes of what they heard while listening, negatively affected their scores on the Question-Response tasks. This finding could be because lower-proficiency listeners were unable to keep up with new input while they were translating or repeating online and therefore failed to develop a concrete mental representation of the text (Vandergrift, 2003).

Conclusion

This study explored the relationships between strategy use, L2 proficiency level, task type, and scores, via cognitive and metacognitive strategy use questionnaires and the TOEIC practice listening test. We found that some strategies—voice and imagery inference and elaboration, approaches, and top-down processing strategies—were used with a similar frequency regardless of learners' proficiency, while others—planning, monitoring and evaluation, linguistic inference and elaboration, and bottom-up processing—were more often used by advanced listeners. Additionally, planning (and inference/elaboration, and top-down processing) strategies were more likely to be used in easier tasks. Finally, the relationship between reported strategy use and test scores was weak, accounting for 7% of the total score variance and 5% to 10% of the score variance for each task type section except the Short Talks.

These findings should not be considered conclusive or comprehensive. First, our data solely relied on the self-reported questionnaires and did not completely reflect all strategies actually utilised by our participants. A triangulation of the current findings is required using stimulated recalls and interviews. Second, we only recruited 170 students at a single institution and thus should be careful about generalising the findings. Finally, the participants in this study only completed the TOEIC practice test, thus they answered merely half of the total number of questions on the full TOEIC listening test. Their reported strategy use may have differed from what they would use in an actual testing situation, due to the increased number of questions and the presumably higher stakes associated with the full TOEIC.

One pedagogical implication of the current study is the need for strategy instruction for less able listeners. Given that some strategies such as planning, monitoring and evaluation, linguistic inference and elaboration, and bottom-up processing were more often used by advanced listeners, learners of lower listening proficiency might benefit from strategy instruction aimed at increasing the use of these strategies. Through increased use of these strategies, lower-proficiency learners may eventually demonstrate better performance by effectively employing these strategies along with other strategies in response to aural input. While strategy instruction has been reported to be effective overall, it has been found to have negligible effects in listening studies (Plonsky, 2011). This may be partly because the listening comprehension processes of successful listeners are complex and interactive depending on task type (Brindley & Slatyer, 2002). As the TOEIC uses a predetermined, fixed range of tasks, it could be possible for test-takers to learn how to apply a focused set of strategies. It would be useful to trace learners' L2 listening strategy development over a period of time to examine the relationship between the duration of strategy training and its effectiveness.

References

- 104 Job Bank. (2012). *104 Job Bank online advertisement*. Retrieved from <http://www.104.com.tw/>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, United Kingdom: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, United Kingdom: Oxford University Press.
- Bacon, S. M. (1992a). Phases of listening to authentic input in Spanish: A descriptive study. *Foreign Language Annals*, 25, 317-334.

- Bacon, S. M. (1992b). The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign language listening. *Modern Language Journal*, 76, 160-178.
- Badger, R., & Yan, X. (2012). To what extent is communicative language teaching a feature of IELTS classes in China? *IELTS Research Reports Volume 13*. Retrieved from <http://www.ielts.org/PDF/Report%204V13.new.pdf>
- Barkaoui, K., Brooks, L., Swain, M., & Lapkin, S. (2013). Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*, 34, 304-324.
- Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, 1, 369-394.
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, 8, 67-91.
- Buck, G. (2001). *Assessing listening*. Cambridge, United Kingdom: Cambridge University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Chamot, A. U., O'Malley, J. M., Kupper, L., & Impink-Hernandez, M. V. (1987). *A study of learning strategies in foreign language instruction: First year report*. Rosslyn, VA: Interstate Research Associates.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Council of Europe. (2014). *Common European framework of reference for languages: Learning, teaching, assessment (CEFR)*. Retrieved from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp
- Cross, J. (2009). Effects of listening strategy instruction on news video text comprehension. *Language Teaching Research*, 13, 151-176.
- Curricula Web. (2015). *Taiwan Universities of Technology*. Retrieved from <http://course-tvc.yuntech.edu.tw/Web/Default.aspx>
- Educational Testing Service Taiwan Representative. (2012). *Taiwan TOEIC statistics report*. Retrieved from http://www.toEIC.com.tw/toEIC_news_page.jsp?type=4
- Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing*, 13, 23-51.
- Green, N. M., & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly*, 29, 261-297.
- Lai, Y. (2009). Language learning strategy use and English proficiency of university freshmen in Taiwan. *TESOL Quarterly*, 43, 255-280.
- Macaro, E., Graham, S., & Vanderplank, R. (2007). A review of listening strategies: Focus on sources of knowledge and on success. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 165-185). New York, NY: Oxford University Press.
- McNamara, T., & Lumley, T. (1997). The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings. *Language Testing*, 14, 140-156.
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge, United Kingdom: Cambridge University Press.
- O'Malley, J., Chamot, A. U., Stewner-Manzanares, G., Küpper, L., & Russo, R. (1985). Learning strategies used by beginning and intermediate ESL students. *Language Learning*, 35, 21-46.

- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. New York, NY: Newbury House.
- Park, G. (1997). Language learning strategies and English proficiency in Korean university students. *Foreign Language Annals*, 30, 211-221.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20, 26-56.
- Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61, 993-1038.
- Prabhu, N. S. (1987). *Second language pedagogy*. Oxford, United Kingdom: Oxford University Press.
- Purpura, J. E. (1997). An analysis of the relationships between test-takers' cognitive and metacognitive strategy use and second language test performance. *Language Testing*, 47, 289-325.
- Purpura, J. E. (1998). Investigating the effects of strategy use and second language test performance with high- and low-ability groups: A structural equation modeling approach. *Language Testing*, 15, 333-379.
- Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge, United Kingdom: University of Cambridge Local Examinations Syndicate and Cambridge University Press.
- Riley, G. L., & Lee, J. F. (1996). A comparison of recall and summary protocols as measure of second language reading comprehension. *Language Testing*, 13, 173-190.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1, 147-170.
- Song, X. (2005). Language learner strategy use and English proficiency on the Michigan English Language Assessment Battery. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, 3, 1-25.
- Song, X., & Cheng, L. (2006). Language learner strategy use and test performance of Chinese learners of English. *Language Assessment Quarterly*, 3(3), 243-266.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.
- Taiwan Executive Yuan. (2006). 行政院修正公務人員英語檢測陞任評分計分標準對照表, 行政院民國 95 年 4 月 4 日院授人力 字第 0950061619 號 [Revised conversion chart of current English proficiency tests in terms of the CEFR levels, Letter 0950061619, issued by Taiwan Executive Yuan]. Retrieved from <http://140.120.1.20/~indodep/chinese/form/E03.pdf>
- Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning*, 53, 463-496.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40, 191-210.
- Vandergrift, L., & Goh, C. (2009). Teaching and testing listening comprehension. In M. Long & C. Doughty (Eds.), *Handbook of second and foreign language teaching* (pp. 395-411). New York, NY: Wiley-Blackwell.
- Wu, J. R. (2008). An investigation of the relationships between strategy use and GEPT test performance. *English Teaching & Learning*, 32(3), 35-69.
- Wu, Y. (1998). What do tests of listening comprehension test? A retrospective study of EFL test-takers performing a multiple-choice task. *Language Testing*, 15, 21-44.

- Yang, H.-C., & Plakans, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly*, 46, 80-103.
- Zhang, W., Liu, M., Zhao, S., & Xie, Q. (2011). English test-taking strategy use and students' test performance. *The Asian EFL Journal*, 13(2), 133-168.
- Zhang, Y. (2006). *New TOEIC*. Taipei, Taiwan: Crane Publishing Company.

Appendix A
Cognitive and Metacognitive Questionnaires

(C refers to cognitive strategy while M refers to metacognitive strategy.)

Directions: Please choose the strategies you have adopted based on the 10 listening questions in the Photographs part you have just finished, and tick them in the appropriate boxes.

1. Before the test, I remind myself not to be nervous. (M)
2. Before the test, I remind myself what I need to pay attention to. (M)
3. Before the test, I listen to the directions carefully. (M)
4. Before the test, I remind myself to think in English while listening. (M)
7. During the test, I pay attention to individual words/phrases. (C)
8. During the test, I pay attention to any words/phrases that are repeated. (C)
9. During the test, I try to understand the main idea rather than the details. (C)
10. During the test, I try to understand the main idea, and then the details. (C)
11. During the test, I utilize the intonation and tone of voice to guess the vocabulary and text. (C)
12. During the test, I utilize the words/phrases that I already know to guess the meaning of any unknown vocabulary or text that I encounter. (C)
13. During the test, I utilize grammatical structures to guess the meaning of the text. (C)
14. During the test, I utilize the tone of voice and take some time to segment the text in order to understand it. (C)
15. During the test, I generally understand the intonation and tone of voice rather than the meanings of the vocabulary and phrases. (C)
16. During the test, I try to understand the listening text by making sense of the details. (C)
17. During the test, I connect any new words with words that I already know. (C)
18. During the test, I utilize my prior knowledge/experience to understand the text. (C)
19. During the test, I create a mental picture of the text in order to comprehend it. (C)
20. During the test, I predict what I will be hearing. (C)
21. When I notice that I am having difficulty during the listening process, I don't give up, but rather I continue listening. (M)
22. During the test, I repeat the words/phrases that I have heard to myself. (C)
23. During the test, I summarize or outline in my own words what I have heard. (C)
24. During the test, I take notes while listening. (C)
25. During the test, I translate what I have heard while listening. (C)
26. When I notice my concentration flagging, I convince myself to focus and continue listening. (M)
27. During the test, I listen to the questions with eyes closed. (M)
28. During the test, I relax to reduce my anxiety so that I can concentrate. (M)
29. After the test, I evaluate how much of the test I understood. (M)
30. After the test, I try to identify the problems/difficulties that I experienced with the listening test. (M)

Appendix B
Pearson Correlations Between Each Strategy and Scores (Photographs Tasks)

	TOEIC ^a	P	ME	LIE	VIIE	A	TDP	BUP
TOEIC ^a	1							
P	.20**	1						
ME	.25**	.53**	1					
LIE	.13	.36**	.44**	1				
VIIE	.08	.35**	.32**	.42**	1			
A	.10	.25**	.40**	.30**	.34**	1		
TDP	.09	.38**	.31**	.38**	.34**	.16*	1	
BUP	.15*	.26**	.39**	.31**	.20**	.38**	.29**	1

Note. $N = 170$. P = Planning, ME = Monitoring and evaluation, LIE = Linguistic inference and elaboration, VIIE = Voice and imagery inference and elaboration, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

^aTOEIC Section 1 score.

* $p < .05$. ** $p < .001$.

Appendix C
Pearson Correlations Between Each Strategy and Scores (Question-Response Tasks)

	TOEIC ^a	P	ME	LIE	VIIE	A	TDP	BUP
TOEIC ^a	1							
P	.23**	1						
ME	.25**	.56**	1					
LIE	.10	.34**	.55**	1				
VIIE	.11	.41**	.37**	.47**	1			
A	-.09	.28**	.36**	.39**	.38**	1		
TDP	.10	.36**	.39**	.38**	.34**	.24**	1	
BUP	.14	.42**	.43**	.40**	.26**	.42**	.34**	1

Note. $N = 170$. P = Planning, ME = Monitoring and evaluation, LIE = Linguistic inference and elaboration, VIIE = Voice and imagery inference and elaboration, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

^aTOEIC Section 2 score.

* $p < .05$. ** $p < .001$.

Appendix D
Pearson Correlations Between Each Strategy and Scores (Conversation Tasks)

	TOEIC ^a	P	ME	LIE	VIIE	A	TDP	BUP
TOEIC ^a	1							
P	.17*	1						
ME	.20**	.63**	1					
LIE	.19*	.50**	.54**	1				
VIIE	.11	.46**	.48**	.50**	1			
A	.12	.36**	.50**	.46**	.41**	1		
TDP	.22**	.55**	.49**	.44**	.40**	.19*	1	
BUP	.22**	.48**	.45**	.44**	.42**	.38**	.39**	1

Note. $N = 170$. P = Planning, ME = Monitoring and evaluation, LIE = Linguistic inference and elaboration, VIIE = Voice and imagery inference and elaboration, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

^aTOEIC Section 3 score.

* $p < .05$. ** $p < .001$.

Appendix E
Pearson Correlations Between Each Strategy and Scores (Short Talks Tasks)

	TOEIC ^a	P	ME	LIE	VIIE	A	TDP	BUP
TOEIC ^a	1							
P	-.01	1						
ME	.00	.62**	1					
LIE	.03	.48**	.50**	1				
VIIE	.03	.44**	.38**	.51**	1			
A	-.04	.37**	.36**	.41**	.40**	1		
TDP	-.11	.40**	.38**	.33**	.39**	.27**	1	
BUP	.02	.45**	.46**	.33**	.33**	.47**	.21**	1

Note. $N = 170$. P = Planning, ME = Monitoring and evaluation, LIE = Linguistic inference and elaboration, VIIE = Voice and imagery inference and elaboration, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

^aTOEIC Section 4 score.

* $p < .05$. ** $p < .001$.

Appendix F
Pearson Correlations Between Each Strategy and the Total Score

	TOEIC ^a	P	ME	LIE	VIIE	A	TDP	BUP
TOEIC ^a	1							
P	.16*	1						
ME	.26**	.53**	1					
LIE	.16*	.36**	.44**	1				
VIIE	.12	.35**	.32**	.42**	1			
A	.09	.25**	.40**	.30**	.34**	1		
TDP	.16*	.38**	.31**	.38**	.34**	.16*	1	
BUP	.19*	.26**	.39**	.31**	.20**	.38**	.29**	1

Note. $N = 170$. P = Planning, ME = Monitoring and evaluation, LIE = Linguistic inference and elaboration, VIIE = Voice and imagery inference and elaboration, TDP = Top-down processing, BUP = Bottom-up processing, A = Approach.

^aTOEIC total score.

* $p < .05$. ** $p < .001$.