

Innovative assessment of aviation English in a virtual world: Windows into cognitive and metacognitive strategies

MOONYOUNG PARK

Chinese University of Hong Kong, Hong Kong SAR, China
(email: moonyoungpark@cuhk.edu.hk)

Abstract

Aviation English proficiency is a core competency in the global air traffic controller profession. There is, however, growing concern about the current ineffective paper-based assessment methods and the severe lack of interactive online testing for such a critical profession, one that should be ideally assessed in an authentic task and situation (Alderson, 2010; Douglas, 2013). The tests, which lack validity and authenticity, cannot capture how knowledge of aviation English communication and strategic competence are used through valid means and, therefore, inevitably fail to predict how test takers actually perform in the target language use situations (Douglas, 2000). In the present study, the researcher examines the potential use of verbal report data produced by test takers interacting in a virtual testing environment. The research seeks to answer two specific questions: (a) What types of strategies are used in the virtual aviation English task performance? and (b) How can the assessed strategies be interpreted in relation to test takers' performance? The analysis of the test takers' verbal reports from stimulated recalls indicates that various cognitive, metacognitive, and communication strategies were used while performing the Virtual Interactive Tasks for Aviation English Assessment, and that there is a positive relationship between the total number of cognitive and metacognitive strategies adopted and the test scores. The findings suggest that the use of an immersive interface and simulated tasks in a virtual world could provide language learners with more authentic opportunities to perform the target tasks and promote strategic, as well as linguistic, competence.

Keywords: aviation English, language testing, Second Life, cognitive strategy, metacognitive strategy, communication strategy

1 Introduction

Aviation English, a key component of air traffic control (ATC) communication, can be defined as a comprehensive but specialized subset of English for specific purposes (ESP) related broadly to aviation and consisting of both plain language and standardized phraseology for radiotelephony communications (International Civil Aviation Organization [ICAO], 2004). Although aviation English between air traffic controllers and pilots is used in highly predictable circumstances following a prescribed sequence (Mell, 1992), miscommunications do occur, especially between non-native English speaking pilots and

air traffic controllers (Cushing, 1994; Tajima, 2004). It has been reported that at least 11% of fatal airplane crashes worldwide were due to miscommunications related to second language use in the period of 1982–1991 (Crystal, 2003). In response to growing concerns about aviation safety, several aviation English tests such as the test of English Language Proficiency for Aeronautical Communication by EUROCONTROL (Alderson, 2009), Test of English Language Level for Controllers and Pilots by TELLCAP (Alderson, 2006), and RMIT English Language Test for Aviation by RMIT University (Zokić, Boras & Lazić, 2012) have been developed based on the ICAO policy of language proficiency requirements (LPRs).

Unfortunately, however, despite the promotion of a standardized quality of the tests, little to no confidence can be placed in the reliability and validity of the several currently available aviation language tests (Alderson, 2006). Although most aviation English testing companies boast their assessments' computer-assisted test service features, the companies do not make full use of the online testing environment. The companies simply present motionless pictures and plain text prompts for playing audio materials and record test takers' voices instead of integrating interactive features in order to assess interactive ATC task performance. Therefore, there is a current pressing need to develop an authentic and interactive aviation English test to observe test takers' aviation English proficiency and capture not only test takers' knowledge of aviation English communication, but also their demonstrated strategic competence in ATC interactions while taking advantage of innovative technology such as virtual worlds (VWs).

To address this discrepancy, the recent efforts of the US Army Research Institute for the Behavioral and Social Sciences (ARI) to develop prototype interactive training applications and assessments in maturing learning technologies, including VWs and collaborative game-based technologies, have been thought provoking and refreshingly positive (Brusso, Wisher, Paddock & Hatfield, 2014). Such endeavors are very much in line with the latest assessment trends in K–12 US classrooms seeking to enhance the feasibility of virtual performance assessments and to evaluate scientific inquiry through authentic interactive testing environments (Clarke-Midura & Dede, 2010). Although VWs were not explicitly created for language education purposes, several language educators and researchers have made use of them as platforms for teaching and learning a variety of disciplinary content, including languages (da Silva, 2012; Sadler, 2012). Instructional benefits of VWs include promoting social interaction (Chuang, Chang & Chen, 2014), creating community (Oliver & Carr, 2009), facilitating collaboration (Peterson, 2012; Vuopala, Hyvönen & Järvelä, 2016), lowering social anxiety (Melchor-Couto, 2017), and enhancing learner motivation and engagement (Verhagen, Feldberg, van den Hooff, Meents & Merikivi, 2012). VWs excel at providing users with highly interactive arenas for dynamic feedback, learner experimentation, and real-time personalized task performance (Abdallah & Mansour, 2015; Lin, Wang, Grant, Chien & Lan, 2014; Swier, 2014; Wang, Calandra, Hibbard & Lefaiver, 2012). Simulating target language use (TLU) situations by providing a set of specific language use tasks that the test taker is likely to encounter outside of the test itself (Bachman & Palmer, 1996; Douglas, 2000) in a VW and providing test takers with authentic and interactive tasks in the VW may encourage test takers to fully demonstrate what they are actually capable of accomplishing in the TLU situations. It is therefore important to explore the affordances of VWs as a potential language assessment platform for observing test takers' strategic competence, as well as linguistic knowledge, during interactions facilitated by authentic target tasks.

This paper will first examine aviation English assessment and the significance of strategic competence and then elaborate on the potential role of VWs for language assessment. What follows is an exploratory study that investigates cognitive, metacognitive, and communication strategies used during interaction in a VW and their relation to test takers' performance.

2 Literature review

2.1 Aviation English assessment and strategic competence

Language ability is an abstract concept; therefore, it is necessary to define "ability" in explicit terms for language assessment. Such defined ability, inferred from a meaningful interpretation of observed behavior, is called a construct (Bachman & Palmer, 2010). The construct under investigation in this study is air traffic controllers' aviation English ability based on an interactionalist definition (Chapelle, 1998) in which a test taker's performance on a test indicates not only underlying language knowledge and the assessment context, but also strategic competence (Bachman & Palmer, 1996). This may suggest that the test taker utilizes her or his language knowledge in a given assessment context using various strategies. To provide a more articulated concept of strategic competence, the researcher operationalized cognitive, metacognitive, and communication strategies (Chamot & O'Malley, 1987; Dörnyei & Scott, 1997; Purpura, 1999) as follows:

1. Cognitive strategies, which involve the test taker's interaction with the interlocutors and assessment context to be assessed by manipulating it mentally through the processes of identification, retention, storage, comprehension, and retrieval.
2. Metacognitive strategies, which involve executive processes in goal setting, assessment (assessing the situation, self-monitoring, and self-evaluating), and planning (formulating a plan).
3. Communication strategies, which involve the test taker's direct, interactional, and indirect communication-enhancing devices such as paraphrasing, substitution, coining new words, switching to the first language, and asking for clarification.

Although the use of strategic competence implies direct or indirect impact on task performance, unlike communication strategies, both cognitive and metacognitive strategies are not directly observable as they take place as mental processes and are largely ignored in language assessment (Lai, 2011). Thus, only communication strategies, the observable actions that the test taker selects for interacting with other interlocutors, are often included as integral in the assessment criteria under the interaction category adopted in the LPRs' rating criteria by the ICAO (2004).

For the assessment of aviation English, the ICAO has published a set of LPRs and proficiency rating criteria, which consists of six areas of language use: pronunciation, structure, vocabulary, fluency, comprehension, and interaction. Among the six areas of language use, successful interaction is defined by the ICAO as "Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying" (ICAO, 2004: 4-14). Interaction in aviation English is goal oriented and formulaic with the following characteristics: high responsiveness, less valued politeness, restricted turn-taking, and determined topics according to information-transfer requirements

(Van Moere, Suzuki, Downey & Cheng, 2009). In this regard, it is highly important to recognize the role of strategic competence in task performance and to find ways to reflect strategy uses in language assessment.

Despite endeavors to standardize the quality of aviation English assessment, currently released aviation English tests may focus on language ability superficially in the context of aviation English but fail to measure what the test taker actually could accomplish in an authentic TLU situation (Alderson, 2006). It is not enough to merely give test takers topics relevant to the field in which they are studying or working and assess verbal interaction without taking into consideration test takers' cognitive and metacognitive strategy uses. If we want to assess how well test takers can use language for specific purposes, a measure that considers language knowledge and background knowledge, and use of strategic competence related to the TLU situation is required. In other words, the more authentic the testing environment, the more strategic competence the examinees may exploit in the task performance (Douglas, 2000). Accordingly, it is crucial for test developers to provide test takers with authentic and interactive test content and contexts to encourage them to fully demonstrate what they can cognitively process and what they can actually do in real-life ATC communication scenarios.

2.2 Virtual worlds for language assessment

Currently, there are hundreds of VWs that have drawn interest from educators. Table 1 summarizes some popular VW platforms and includes information about their year of release, costs, target audiences, themes, number of members, and installation method.

As highlighted in Sadler (2012), many VWs, including the nine platforms shown in Table 1, share the following common characteristics, such as offering online 3D environments, inclusion of avatars, facilitation of real-time interaction, 24-hour accessibility, provision of social space, heightened numbers of users, and longevity of the environment. Among the features of VWs, interaction is one of the most crucial processes in which language learners can develop both linguistic knowledge and communicative competence (Tseng, Tsai & Chao, 2013). It is thus imperative to provide authentic language learning and interaction opportunities to language learners so that they can practice the target tasks through synchronous or asynchronous interaction (Holliday, 2007; Khalsa, Maloney-Krichmar & Peyton, 2007).

Table 1. Overview of popular VW platforms

	Released	Cost	Audience	Theme	Members	Installation
Active Worlds	1995	Free	Unstated	Social	3+ million	Browser add-on
Club Penguin	2005	Free	6–14	Penguins	200+ million	Browser-based
Habbo	2000	Free	13+	Hotel	273+ million	Browser-based
Second Life	2003	Free	13+	Social	36+ million	Download
SmallWorlds	2009	Free	13+	Social	25+ million	Browser-based
There	1998	\$10/month	18+	Social	1+ million	Download
Twinty	2008	Free	Unstated	Social	Unstated	Download
Whyville	1999	Free	8–14	Educational	7+ million	Browser-based
World of Warcraft	2004	\$14.99/month	Unstated	Role play, battle	5.5+ million	Download

A review of recent empirical research on the use of VWs in language learning and teaching (see Table 2) shows that the two most popular study topics are (1) language gain through negotiation of meaning and/or form in a VW, and (2) usability testing (e.g. learners' perception and affordances of VWs). To explore such research themes, researchers of VWs frequently adopt surveys, pre-/post-tests, and interviews as research instruments. Among the diversity of platforms, Second Life has been one of the most widely explored for educators and researchers. According to a recent survey given to 237 Second Life users (Sadler, 2011), more than 80% of Second Life participants have experienced a great deal of second language practice in their interactions with other users in Second Life. In particular, the *landmark* function in Second Life, similar to an Internet bookmark, enables experienced users, educators, and researchers to link their own virtual space with the landmark and conveniently invite other users. Based on the distinct characteristics of Second Life, it sustains a strong community of educators, numbering in the thousands, including teachers and researchers working with various languages (Sadler, 2012).

In what was considered the early stage of adopting and applying VWs to language learning and teaching, it was adequate to examine users' interactive communication in and perceptions of virtual spaces to identify empirical evidence to support the affordances, capability, and effectiveness of VWs for language learning and teaching. However, greater efforts should now be made to examine non-verbal interaction (e.g. strategic competence, situated cognition) facilitated and promoted by VWs, which this researcher believes is one of the most fundamental advantages of virtual spaces. Additionally, more diverse research areas and research methods with VWs may need to be explored to take best advantage of virtual environments in the realm of second language learning and testing.

The use of VWs has the potential to improve language assessment dramatically by validly measuring test takers' sophisticated intellectual and psychosocial performance and compensating for the shortcomings of multiple-choice or simple online tests that contain little to no interactive features. There are two major advantages supported by the literature. First, VWs can increase face validity, or test takers' intuitive judgments about the test (Alderson, Clapham & Wall, 1995). Although frequently dismissed by testers as being unscientific and irrelevant (Stevenson, 1985), face validity can be integral in determining the test's acceptability and reasonableness to those who will be tested and to those who use test results (Messick, 1989). Second, VWs can enhance construct validity. With respect to the construct definition of aviation English, strategic competence serves as a mediator between test takers' language knowledge and context by controlling the interaction between them (Douglas, 2000). By approximating authentic TLU situations in VWs, test takers' use of strategic competence should be promoted to better predict the language learners' future real-life performance in authentic ATC situations.

The goal of the current research project was to explore and identify the potential of a VW (Second Life) for promoting more authentic cognitive and metacognitive strategies, such as those that occur in real-world scenarios, by capturing what specific strategies the test takers actually use during their task performance in Second Life. This investigation sought to answer the following questions: (a) What types of strategic competence are used in the virtual aviation English task performance by the test takers? and (b) How can the assessed strategies be interpreted in relation to test takers' aviation English performance? This would shed light on the potential use of VWs for language assessment as well as windows into strategic competence.

Table 2. Overview of recent empirical research with VWs

Author (year)	Platform	Participants	Theoretical perspectives	Methods	Findings
Melchor-Couto (2017)	Second Life	14 Spanish major undergrads	Negotiation of meaning	Questionnaire	Foreign language anxiety level decreased.
Vuopala <i>et al.</i> (2016)	Second Life	54 higher ed. students	Negotiation of meaning	Analysis of discussion notes and speech turns	Different forms of interaction were found according to virtual environments.
Abdallah and Mansour (2015)	Second Life	10 pre-service English teachers	Situated learning	E-portfolio, post-test, survey	Participants' pragmatic writing and technological self-efficacy were developed.
Can and Simsek (2015)	Second Life	36 pre-service English teachers	User perception	Interview, survey	Interaction in a VW was found very natural and motivating.
Hammick and Lee (2014)	Second Life	58 undergrads	User perception	Post-test, interview	Shy individuals felt less communication apprehension in a VW.
Chuang <i>et al.</i> (2014)	Second Life	52 EFL college learners	Negotiation of meaning	Questionnaire	Learners in a VW used more affective strategies.
Swier (2014)	Minecraft	6 English major undergrads	User perception	Questionnaire	Minecraft was found suitable for communicative tasks.
Lin <i>et al.</i> (2014)	Second Life	4 instructors, 144 Chinese major undergrads	Negotiation of meaning	Journals, observation	Implementing task-based language teaching in a VW promotes more language input and production, and feedback.
Berns, Gonzalez-Pardo and Camacho (2013)	OpenSim	85 undergrads in different majors	Negotiation of meaning and form	Questionnaire	VWs cannot replace learning management system platforms, so both should be used as complementary platforms.
Canto, Jauregi and van den Bergh (2013)	Second Life	36 language major undergrads, 14 pre-service teachers	Negotiation of meaning	Online survey, pre-/post-tests	Virtual telecollaboration has a positive impact on language learning.
Tseng <i>et al.</i> (2013)	Second Classroom	38 pre-service teachers	User perception	Interview	A VW can elicit a certain degree of participation and interaction.
Peterson (2012)	Second Life	8 science major undergrads	Negotiation of meaning and form	Discourse analysis, questionnaire, interview	Second Life provides learner-centered social interaction for target language practice and the development of autonomy.
Wang <i>et al.</i> (2012)	Second Life	20 English major undergrads	Negotiation of meaning and form	Online survey, pre-/post-tests	Participants found Second Life motivating and interactive.

3 Research methods

This study adopts a mixed-method research design (Creswell & Plano Clark, 2007) using quantitative data and qualitative data to answer the research questions: (1) quantitative data from test takers' aviation English test performance scores and coded think-aloud data, and (2) qualitative data from think-aloud discourse data.

3.1 Participants

Participants in this study were five enlisted soldiers (all male) and five non-commissioned officers (NCO; two females, three males) whose job was to conduct ATC in the military air traffic service battalion in an Asian country. Initially, 12 test takers were invited to participate in the stimulated recall data collection; however, due to the quality of audio recording and unexpected absences as a result of emergency duties, 10 test takers' stimulated recall data were analyzed. A brief summary of the participants' demographic information is provided in Table 3.

3.2 Virtual interactive aviation English tasks in Second Life

The Virtual Interactive Tasks for Aviation English Assessment (VITAEA), consisting of pre-existing aviation English tasks developed by the researcher, entail listening to simulated pilots' radiotelephony transmissions involving requesting information, instructions, and clearance, watching animated helicopters landing and departing, reading flight plans and weather information in a digital board in the ATC tower, and then orally responding to the pilots in seven prototype aviation English tasks simulated in Second Life. To explore the target aviation English tasks and task topics in the TLU situation, and to examine rating criteria, the researcher administered a task-based needs analysis survey based on a task-based performance assessment design (Long, 2005) to 87 experienced air traffic controllers from the same target context. Based on this authentic task and task-topic identification, seven prototype aviation English assessment tasks were created (see Table 4).

Table 3. *Descriptive statistics of the participants' demographic information*

Measure	<i>M</i>	<i>SD</i>	Min	Max
Age	26.7	5.44	22	38
– Enlisted soldiers (<i>n</i> = 5)	23.8	1.1	22	25
– NCO (<i>n</i> = 5)	28	6.66	23	38
ATC service period (in months)	29.8	57.48	2	185
– Enlisted soldiers	4	4.47	2	12
– NCO	55.6	75.82	2	185
English test (TOEIC) score	788	149.06	610	970
– Enlisted soldiers	876.67	113.72	750	970
– NCO	655	63.64	610	700
Study abroad period (in months)	27	43.46	0	120
– Enlisted soldiers	64.8	46.16	12	120
– NCO	0	0	0	0

Note. The total number of the participants was 10.

Table 4. *Final blueprint for the Virtual Interactive Tasks for Aviation English Assessment*

Final blueprint – Virtual Interactive Tasks for Aviation English Assessment	
Stimulus	Items per task situation
1. Colloquial communication	4 questions – service period; perception about ATC job; local food; places to eat out
2. Changed flight plan	3 expected responses including mixed language skills – listening, writing, and speaking
3. Terminal information	1 expected response including mixed language skills – listening, speaking, and reading
4. Departure procedure	6 expected responses including mixed language skills – listening, speaking, and reading
5. Transition procedure	4 expected responses including mixed language skills – listening and speaking
6. Arrival procedure	5 expected responses including mixed language skills – listening and speaking
7. Arrival and departure procedure	10 expected responses including listening, speaking, and reading

Note. The total test time was 20 minutes.

The evaluation of test takers' aviation English performance in Second Life was founded on two different rating criteria: (1) an ICAO-implemented language-centered rating rubric, and (2) a task-centered rating rubric focusing on the task accomplishment. The rationale for having two different rating criteria was that even though ICAO's analytical assessment criteria were well defined and appropriate for assessing complex and multidimensional aviation English proficiency, such a language-centered assessment rubric would still be unable to directly determine successful accomplishment of the particular target tasks. To make better use of this innovative assessment in Second Life as a vehicle for determining task accomplishment and the elicitation of language performance, respectively, the criteria for task-based aviation English performance assessment aimed to not only inform inferences about test takers' aviation English proficiency (ability), but also provide test users with the experience of successfully completed authentic tasks.

In the Manual on the Implementation of ICAO Language Proficiency Requirements (ICAO, 2004), six levels of operational proficiency, ranging from pre-elementary (Operational Level 1) through expert (Operational Level 6), are provided. Six dimensions of proficiency are evaluated:

- Pronunciation – pronunciation, stress, rhythm, and intonation;
- Structure – grammar, sentence patterns, global meaning errors, and local errors;
- Vocabulary – style, tone, lexical choices, which correspond to context and status, idiomatic expressions, articulation of subtle differences or distinction in expression, and meaning;
- Fluency – naturalness of speech production, absence of inappropriate hesitations, fillers, and pauses that may interfere with comprehension;
- Comprehension – clear and accurate information transfer that results in understanding;
- Interactions – sensitivity to verbal and non-verbal cues and appropriate response to them.

A primary concern of the task-centered rating rubric is test takers' accomplishment of target tasks in TLU situations. Raters were asked to identify the level of task

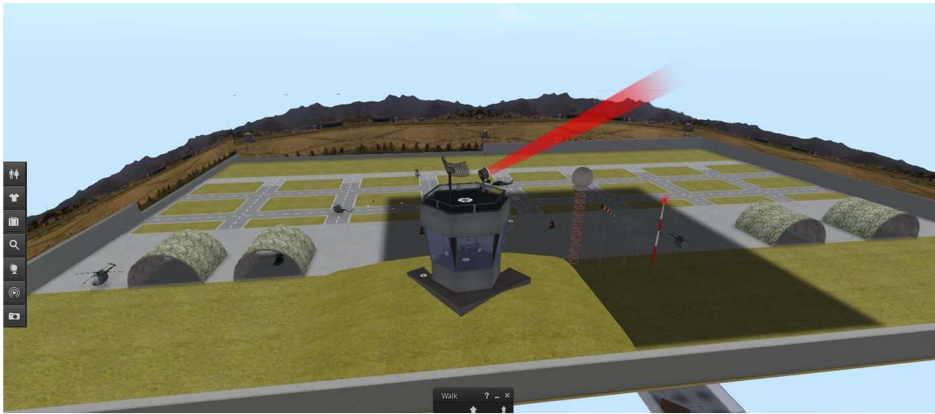


Figure 1. Overview of the virtual interactive aviation English test environment simulated in Second Life



Figure 2. Actual and emulated image of the front view of the ATC tower

accomplishment as (1) excellent, (2) acceptable, or (3) unacceptable. Task-centered rating criteria in each level correspond to one another. The prototype of the task-centered rating rubric was revised according to domain experts' consensus and feedback before piloting.

Figure 1 shows the Second Life simulated airbase environment, created based on authentic satellite images, in which test takers engaged in aviation English communication. When test takers log into the ATC tower environment in the testing situation, they can interact with simulated pilots and helicopters and access ATC information, such as weather information, airbase information, notice to airmen (NOTAM), and flight plans. Figure 2 presents the image displaying a screen-capture image of the ATC tower in the Second Life testing environment.

3.3 Data collection and analysis

The VITAEA were administered with 10 air traffic controllers one at a time in the airbase over the period of a month. Three laptop computers were used for the test administration: the first was one of the best high-performance PC gaming laptops on the market, and had been set up with a screen-capturing program called Camtasia to record test takers' performance during the test; the second was used for the simulation test administration (e.g. operating helicopters, playing audio files interactively) in Second Life by the researcher; the third was connected to the same Second Life server to provide a high-quality screen-capture angle for recording video during the task performance. The time duration of the test taking itself ranged between 16 and 20 minutes per person. When combined with the follow-up stimulated recall, the entire process lasted approximately 50 minutes per test taker.

For the test rating, two expert air traffic controllers with more than 15 years ATC experience each were recruited to holistically rate the test takers' task performance. For analytical assessment using ICAO's LPRs' rating criteria, three applied linguists in the US (one native speaker of English from the US and two international but native like) were also recruited to rate the same task performance audio files.

In addition to test takers' performance scores, stimulated recall data from the 10 test takers were collected to answer the research question on test takers' strategy use during the target task performance in Second Life. From these verbal reports, recorded using a digital MP3 recorder, inferences could be made about what and how strategies were planned, used, and assessed during the virtual interactive task performance in Second Life. Immediately after administering the aviation English tasks in Second Life, the researcher trained the participants in how to verbalize their cognitive/metacognitive processes during their task performance. In each of the test takers' radiotelephony transmissions, participants were encouraged to think out loud whenever they clicked on navigation buttons, moved the mouse cursor, or read particular information.

Initially accomplished by Purpura (1999) and Dörnyei and Scott (1997), the transcribed protocols were analyzed using grounded theory analysis techniques (Corbin & Strauss, 2008) using open coding to identify an array of strategic actions, and then axial coding to group those actions into a certain number of core strategy types for completing the aviation English tasks in Second Life. For communication strategy use analysis, a coding scheme was adopted from Dörnyei and Scott (1997). The verbal reports focused on the types of cognitive, metacognitive, and communication strategies used in the VITAEA and their relationship with task performance. To assess to what degree a set of test takers' transcripts were consistently coded by the two different coders, intercoder agreement for an entire set of data was calculated with the finding of acceptable intercoder reliability (87.4% agreement) (Paltridge & Phakiti, 2015).

Regarding the expected relationships between the participants' test scores and their strategy use, it was hypothesized that the test scores and the number of strategies the test takers adopted would be positively correlated. To operationalize this, the total number of cognitive and metacognitive strategy use, the number of cognitive and metacognitive strategy types, the number of communication strategy use, task-centered rating scores, and language-centered rating scores were calculated and analyzed. However, due to the small sample size, descriptive statistics and data visualizations are used to display the findings.

4 Results and discussion

In this section, the researcher presents the results of the study sequentially based on the research questions.

To find the answer to the first research question, verbal reports from stimulated recalls with the 10 air traffic controllers were transcribed and analyzed. The verbal reports focused on the types of cognitive, metacognitive, and communication strategies used in the VITAEA and their relationship to task performance.

Instances of each type of cognitive and metacognitive strategy identified from the stimulated recall data analysis are provided below:

- Clarifying/verifying (CLAR)

I had to ask the pilot for repetition of his last transmission as I was not familiar with the pilot's way of reading numbers. (Test taker #1)

- Assessing the situation before events < ASSIT >

I changed the screen view of Second Life to the right side to follow up the aircraft when I received the call. As I could see the aircraft is leaving the control zone from the east bound, I approved frequency change. (Test taker #1)

- Applying rules (APR)

As the pilot said the aircraft is at two miles east of the airbase and requested landing, I approved the landing. (Test taker #4)

- Linking with prior knowledge (LPK)

This request of amending a flight plan was very similar to IFR plan change report, so I wanted to let the pilot know that I am ready to copy the changed flight plan. (Test taker #2)

- Monitoring < MON >

I moved the mouse and screen in Second Life as I wanted to see the moving aircraft clearer. (Test taker #4)

- Inferencing (INF)

I was confused, but tried to guess. (Test taker #8)

- Self-evaluation < SE >

On the runway the pilot just said "ready for take-off". I realized that I should have approved take-off in advance. (Test taker #7)

Findings from the stimulated recall data analysis reveal a variety of cognitive and metacognitive strategies that were adopted during the test task performance in Table 5. *Applying rules* accounts for the largest portion out of the total number of strategy uses (44.09%) followed by *Linking with prior knowledge* (16.13%). Together, these two strategies constitute more than 60% of the entire reported strategy use, which corresponds to the findings from the actual aviation English communication contexts in the current study. As aviation

English consists of standard phraseology, a simplified, situational, and highly structured version of English used in a very restricted context, test takers may need to utilize these linguistic rules, as well as learned situational rules, in order to perform the aviation English communication tasks. This contextual feature may account for the high frequency of use of the two strategies. Additionally, for those who have been involved with ATC relatively longer than other test takers, there was a tendency to apply more experience and/or knowledge to the problem-solving process.

In addition to test takers' cognitive and metacognitive strategy use, their communication strategies, adopted during their test task performance, were coded by two coders (the researcher and an applied linguist) and analyzed. Table 6 displays identified communication strategies adopted by the test takers during the VITAEA task performance.

Table 5. *Identified strategy use during task performance in VITAEA*

Type of strategy	Sources (Test takers)	Cases	Proportion %
Cognitive strategy use	Applying rules (APR)	10	164 44.09
	Linking with prior knowledge (LPK)	10	60 16.13
	Inferencing (INF)	10	40 10.75
	Clarifying (CLAR)	9	20 5.38
Metacognitive strategy use	Monitoring during performance < MON >	9	35 9.41
	Self-evaluation after performance < SE >	9	31 8.33
	Assessing the situation before performance < ASSIT >	9	22 5.91

Table 6. *Identified communication strategy use during task performance in VITAEA*

Type of strategy	Sources (Test takers who used the strategy)	Cases	Proportion %
Communication strategy use	Direct appeal for help	6	10 18.46%
	Use of filler	7	10 15.38%
	Omission	7	9 13.85%
	Asking for repetition	4	8 12.31%
	Mumbling	5	6 9.23%
	Indirect appeal for help	3	6 9.23%
	Self-repair	5	5 7.69%
	Response confirmation	1	3 4.62%
	Response rephrase	2	2 3.08%
	Express non-understanding	1	1 1.54%
	Message replacement	1	1 1.54%
	Response expand	1	1 1.54%
	Response repetition	1	1 1.54%

As seen in Table 6, a total of 13 types of communication strategies were identified. Among the various communication strategies used, four of the most frequently used are *Direct appeal for help* (i.e. by using test takers' first language during the target task performance) (18.46%), followed by *Use of filler* (e.g. *Uhm ...*) (15.38%), *Omission* (i.e. no response) (13.85%), and *Asking for repetition* (e.g. *Can you say it again; one more time; say again*) (12.31%). These four strategies account for 60% of the total communication strategy use by the test takers.

As Douglas (2001) highlighted, communication strategies were indeed employed even when there were no obvious difficulties during the test task performance. Collected data also show that test takers often use repetition of the pilot's response to acknowledge the successful transmission of the delivered information, such as read back. Yet, given the limited number of participants and scarcity of data, there was a tendency for most experienced air traffic controllers to not apply communication strategies as frequently as novice controllers. Compared to most experienced air traffic controllers, novice controllers tended to utilize more communication strategies as they encountered more communication breakdowns or comprehension issues than the experienced controllers. To achieve communication goals in each air traffic control procedure and to compensate for limited communication, novice air traffic controllers in the current study adopted various communication strategies throughout the task performance.

In answering the second research question exploring the relationship between test takers' strategy uses and their VITAEA scores, the *N* size was too small to use correlations. However, positive relationships between the number of different strategies used and test takers' performance (test scores) were identified. The results are presented in Table 7 and Figure 3.

Despite the limitation of the small *N* size and the variables' (test scores and frequency of strategy use/type) unit, as can be seen in Figure 3, overall it appears that a positive relationship exists between test scores and strategy use. According to Figure 3(a) and (b), there appears to be a positive relationship between the total number of strategies used and the two scores on the VITAEA.

Table 7. Summary of the nine test takers' strategy uses and test scores

Examinee#	Strategy total	Strategy type	Com-Strategy	TaskCS	LangCS
1	9	2	7	0.27	2.11
2	43	6	0	1.69	4.61
3	36	7	7	0.54	2.44
4	17	4	6	0.21	2.83
5	42	8	8	0.79	5.11
6	42	10	0	1.10	4.44
7	33	7	0	0.54	4.83
8	0	0	0	0.52	4.11
9	31	8	4	0.56	3.11
10	26	4	5	0.33	3.61

Note. Strategy total = total number of cognitive/metacognitive strategy uses; Strategy type = cognitive/metacognitive strategy types; Com-Strategy = communication strategy use; TaskCS = task-centered scores of VITAEA; LangCS = language-centered scores of VITAEA.

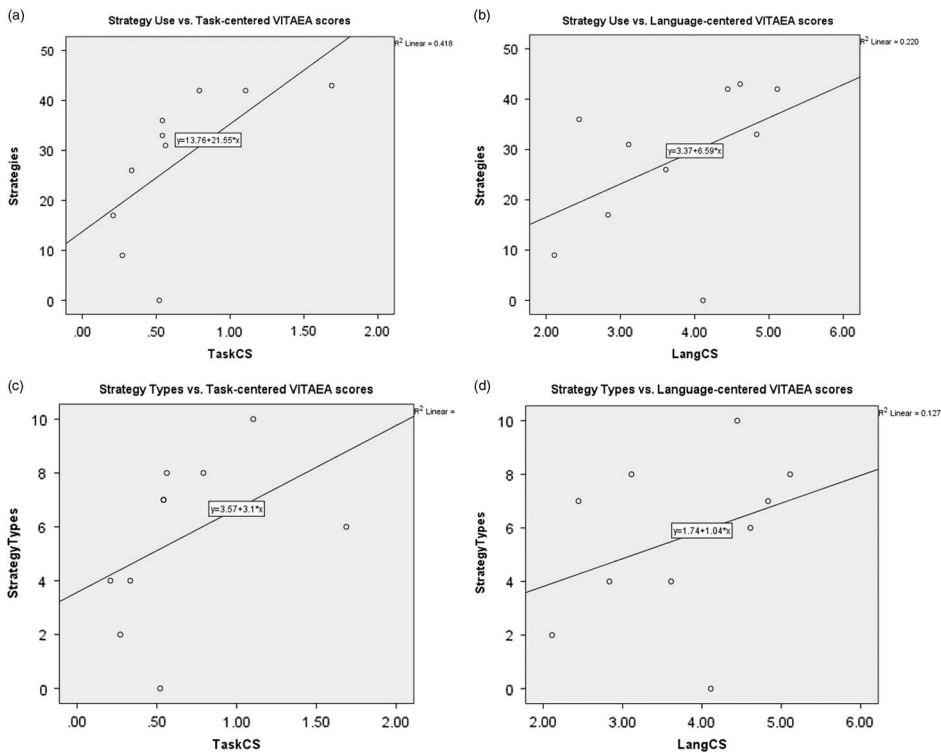


Figure 3. Relationships between VITAEA scores and strategy uses

This may imply that test takers with higher test scores on the VITAEA tended to use a greater number of cognitive and metacognitive strategies during the test task performance.

More significantly, as is apparent in Figure 3(a), a positive relationship between task-centered VITAEA scores and strategy use may also indicate that those test takers who were given a high score on the ATC task utilized a greater number of cognitive and metacognitive strategies during the task performance. This may imply that the use of interactive assessment tasks in a virtual world could facilitate situated cognition and promote test takers' use of various cognitive and metacognitive strategies during the task performance. Furthermore, a positive relationship between the total number of strategy types test takers utilized and task-centered scores on the VITAEA also suggests there is a positive relationship between the two variables. This could also suggest that language performance in an authentic virtual environment could promote the usage of strategic competence in not only total frequency of the use, but also the various types of strategies.

Overall, the findings regarding strategy use provide some backing for the assumption that strategies required by tasks are construct relevant based on the positive relationship between the VITAEA test scores and test takers' strategy use. Moreover, the results are in accordance with the interactionist construct definition (Chapelle, 1998) in which strategic competence directs and assesses world knowledge and language knowledge during task performance. This critical finding of a positive relationship between cognitive/metacognitive strategy use and task-centered scores on the VITAEA provides some empirical evidence that

corresponds to Douglas's emphasis on the test developers' responsibility for "providing sufficient contextual information to enable the test takers to establish the context, to know where they are, and engage an appropriate discourse domain" (2000: 76) in the authentic virtual environment by actively utilizing strategic competence.

In the case of cognitive and metacognitive strategy use, data were extracted through a retrospective recall right after the test to identify test takers' cognitive processes. By contrast, communication strategy use data were gathered from test takers' actual test task performance by conducting a coded analysis of the test takers' oral responses during the test performance to examine what amount and types of communication strategies the test takers adopted while they were trying to accomplish the given tasks. A closer look at the coded communication strategy data reveals that, in many cases, test takers utilized a variety of communication strategies (e.g. direct appeal for help; asking for repetition) when they were faced with difficulties during the task performance; yet there were also some cases in which test takers adopted communication strategies when there had been no obvious difficulties in task completion.

5 Conclusions

The design and implementation of VITAEA for this study was motivated by the need to develop an aviation English test in a virtual world that was well grounded in principles from second language acquisition and language testing. The VITAEA incorporates a sequence of authentic aviation English tasks in Second Life to boost interactive completion of target tasks and designed to promote test takers' strategy uses. The findings reveal that virtual interactive tasks in a virtual world facilitated the employment of a variety of strategy types and their frequency to accomplish target tasks. Furthermore, a positive relationship between participants' test scores and the number of adopted strategies suggests that strategies engaged in during task completion in Second Life are construct relevant and in accordance with theoretical expectations (Chapelle, 1998, 2005; Douglas, 2000).

The current study has several limitations. The first limitation concerns the small sample size used for statistical analysis. Due to limited access to the entire population of military air traffic controllers, only 10 participants' data could be used for a non-parametric statistical analysis. Although the results of the study cannot be generalized at this time, it would have been beneficial to include more participants in the data collection by conducting various surveys, interviews, and tests. Second, despite the researcher's efforts to create an authentic environment, there is much room for improvement in establishing life-like simulations in the TLU situations and test tasks.

Findings from the current study still hold implications for the field of language assessment in language for specific purposes and computer-assisted language learning. Despite the great potential for virtual environments, the field of instructed second language acquisition has mainly constructed an experimental arena for language learning or immersion based on a Vygotskian approach (Schwienhorst, 2002); so, why could these environments not be used for language assessment? The researcher believes that the use of virtual environments could dramatically improve language assessment, especially in language for specific purposes assessment, by allowing the observation of test takers' use of situated cognition (cognitive and metacognitive strategies) in addition to the collection of their verbal responses. Although the current study could not distinguish what specific portion of test takers' strategy use accounts for their task accomplishment, a strong positive

relationship between test takers' use of cognitive/metacognitive strategies and their task accomplishment scores was identified. This meaningful finding could not have been obtained if the current study had utilized a paper-based test or a computer-based test with motionless pictures and hypertext only on the screen. In this regard, an immersive interface and simulated real TLU situations in virtual environments could provide test takers with more authentic opportunities to accomplish the target tasks. For future research, it is the researcher's hope to utilize virtual environments for language for specific purposes assessment to motivate and encourage task-based researchers, language testers, and computer-assisted language learning practitioners who are especially interested in virtual worlds to collaborate in designing and developing more authentic language learning and testing arenas for language learners.

Acknowledgements

I wish to thank the anonymous reviewers and editors for their insightful and constructive comments. My sincere appreciation is extended to the recruited military air traffic controllers for expressing their enthusiasm for and willingness to participate in this research.

References

- Abdallah, M. M. S. and Mansour, M. M. (2015) Virtual task-based situated language-learning with *Second Life*: Developing EFL pragmatic writing and technological self-efficacy. *Arab World English Journal*, **2**: 150–182.
- Alderson, J. C. (2006) *Final report on a survey of aviation English tests*. Lancaster: Lancaster University and the European Organisation for the Safety of Air Navigation (Eurocontrol).
- Alderson, J. C. (2009) Air safety, language assessment policy, and policy implementation: The case of aviation English. *Annual Review of Applied Linguistics*, **29**: 168–187. <https://doi.org/10.1017/S0267190509090138>
- Alderson, J. C. (2010) A survey of aviation English tests. *Language Testing*, **27**(1): 51–72. <https://doi.org/10.1177/0265532209347196>
- Alderson, J. C., Clapham, C. and Wall, D. (1995) *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L. F. and Palmer, A. S. (1996) *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bachman, L. F. and Palmer, A. S. (2010) *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Berns, A., Gonzalez-Pardo, A. and Camacho, D. (2013) Game-like language learning in 3-D virtual environments. *Computers & Education*, **60**(1): 210–220. <https://doi.org/10.1016/j.compedu.2012.07.001>
- Brusso, R. C., Wisher, R. A., Paddock, A. and Hatfield, J. (2014) *Best practices and provisional guidelines for integrating mobile, virtual, and videogame-based training and assessments*. Fairfax: ICF International. <https://doi.org/10.21236/ADA596660>
- Can, T. and Simsek, I. (2015) The use of 3D virtual learning environments in training foreign language pre-service teachers. *Turkish Online Journal of Distance Education*, **16**(4): 114–124.
- Canto, S., Jauregi, K. and van den Bergh, H. (2013) Integrating cross-cultural interaction through video-communication and virtual worlds in foreign language teaching programs: Is there an added value? *ReCALL*, **25**(1): 105–121. <https://doi.org/10.1017/S0958344012000274>
- Chamot, A. U. and O'Malley, J. M. (1987) The cognitive academic language learning approach: A bridge to the mainstream. *TESOL Quarterly*, **21**(2): 227–249. <https://doi.org/10.2307/3586733>

- Chapelle, C. A. (1998) Construct definition and validity inquiry in SLA research. In Bachman, L. F. & Cohen, A. D. (eds.), *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press, 32–70.
- Chapelle, C. A. (2005) Interactionist SLA theory in CALL research. In Egbert, J. L. & Petrie, G. M. (eds.), *CALL research perspectives*. Mahwah, NJ: Lawrence Erlbaum, 53–64.
- Chuang, M.-T., Chang, B. and Chen, H.-C. (2014) Investigating reported social and affective strategy use by EFL learners in virtual and real learning environments. *International Journal of Applied Linguistics & English Literature*, 3(2): 27–33. <https://doi.org/10.7575/aiac.ijalel.v.3n.2p.27>
- Clarke-Midura, J. and Dede, C. (2010) Assessment, technology, and change. *Journal of Research on Technology in Education*, 42(3): 309–328. <https://doi.org/10.1080/15391523.2010.10782553>
- Corbin, J. and Strauss, A. (2008) *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks: Sage. <https://doi.org/10.4135/9781452230153>
- Creswell, J. W. and Plano Clark, V. L. (2007) *Designing and conducting mixed methods research*. Thousand Oaks: Sage.
- Crystal, D. (2003) *English as a global language* (2nd ed.). New York: Cambridge University Press.
- Cushing, S. (1994) *Fatal words: Communication clashes and aircraft crashes*. Chicago: University of Chicago Press.
- da Silva, K. T. F. (2012) *Preparing language teachers to teach in virtual worlds: Analyzing their content, technological, and pedagogical needs*. Iowa State University, unpublished PhD.
- Dörnyei, Z. and Scott, M. L. (1997) Communication strategies in a second language: Definitions and taxonomies. *Language Learning*, 47(1): 173–210. <https://doi.org/10.1111/0023-8333.51997005>
- Douglas, D. (2000) *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Douglas, D. (2001) Language for specific purposes assessment criteria: Where do they come from? *Language Testing*, 18(2): 171–185. <https://doi.org/10.1191/026553201678777068>
- Douglas, D. (2013) ESP and assessment. In Paltridge, B. & Starfield, S. (eds.), *The handbook of English for specific purposes*. Malden: John Wiley & Sons, 367–383.
- Hammick, J. K. and Lee, M. J. (2014) Do shy people feel less communication apprehension online? The effects of virtual reality on the relationship between personality characteristics and communication outcomes. *Computers in Human Behavior*, 33: 302–310. <https://doi.org/10.1016/j.chb.2013.01.046>
- Holliday, L. (2007) Theory and research: Interaction and awareness. In Egbert, J. & Hanson-Smith, E. (eds.), *CALL environments: Research, practice, and critical issues* (2nd ed.). Alexandria, VA: Teachers of English to Speakers of Other Languages, 163–170.
- International Civil Aviation Organization (ICAO) (2004) *Manual on the implementation of ICAO language proficiency requirements* (ICAO Doc 9835). Chicago: Author.
- Khalsa, D. K., Maloney-Krichmar, D. and Peyton, J. K. (2007) Theory and research: Interaction via computers. In Egbert, J. & Hanson-Smith, E. (eds.), *CALL environments: Research, practice, and critical* (2nd ed.). Alexandria, VA: Teachers of English to Speakers of Other Languages, 19–31.
- Lai, E. R. (2011) *Metacognition: A literature review: Research report*. Upper Saddle River: Pearson Education.
- Lin, T.-J., Wang, S.-Y., Grant, S., Chien, C.-L. and Lan, Y.-J. (2014) Task-based teaching approaches of Chinese as a foreign language in Second Life through teachers' perspectives. *Procedia Technology*, 13: 16–22. <https://doi.org/10.1016/j.protcy.2014.02.004>
- Long, M. H. (2005) Methodological issues in learner needs analysis. In Long, M. H. (ed.), *Second language needs analysis*. Cambridge: Cambridge University Press, 19–76. <https://doi.org/10.1017/CBO9780511667299.002>
- Melchor-Couto, S. (2017) Foreign language anxiety levels in Second Life oral interaction. *ReCALL*, 29(1): 99–119. <https://doi.org/10.1017/S0958344016000185>
- Mell, J. (1992) *Study of verbal communication between pilot and air traffic controller in standard and non-standard situations*. Ecole Nationale de l'Aviation Civile, unpublished PhD.

- Messick, S. (1989) Validity. In Linn, R. L. (ed.), *Educational measurement* (3rd ed.). New York: Macmillan, 13–103.
- Oliver, M. and Carr, D. (2009) Learning in virtual worlds: Using communities of practice to explain how people learn from play. *British Journal of Educational Technology*, **40**(3): 444–457. <https://doi.org/10.1111/j.1467-8535.2009.00948.x>
- Paltridge, B. and Phakiti, A. (2015) *Research methods in applied linguistics: A practical resource*. London: Bloomsbury.
- Peterson, M. (2012) EFL learner collaborative interaction in Second Life. *ReCALL*, **24**(1): 20–39. <https://doi.org/10.1017/S0958344011000279>
- Purpura, J. E. (1999) *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge: Cambridge University Press.
- Sadler, R. (2011) *Virtual Worlds for language learning: From theory to practice*. Bern: Peter Lang.
- Sadler, R. (2012) Virtual worlds: An overview and pedagogical examination. *Bellaterra Journal of Teaching & Learning Language and Literature*, **5**(1): 1–22.
- Schwienhorst, K. (2002) Why virtual, why environments? Implementing virtual reality concepts in computer-assisted language learning. *Simulation & Gaming*, **33**(2): 196–209. <https://doi.org/10.1177/1046878102332008>
- Stevenson, D. K. (1985) Authenticity, validity and a tea party. *Language Testing*, **2**(1): 41–47. <https://doi.org/10.1177/026553228500200105>
- Swier, R. (2014) Tasks for easily modifiable virtual environments. *The JALT CALL Journal*, **10**(3): 203–219.
- Tajima, A. (2004) Fatal miscommunication: English in aviation safety. *World Englishes*, **23**(3): 451–470. <https://doi.org/10.1111/j.0883-2919.2004.00368.x>
- Tseng, J.-J., Tsai, Y.-H. and Chao, R.-C. (2013) Enhancing L2 interaction in avatar-based virtual worlds: Student teachers' perceptions. *Australasian Journal of Educational Technology*, **29**(3): 357–371. <https://doi.org/10.14742/ajet.283>
- Van Moere, A., Suzuki, M., Downey, R. and Cheng, J. (2009) Implementing ICAO language proficiency requirements in the Versant Aviation English Test. *Australian Review of Applied Linguistics*, **32**(3): 27.1–27.17.
- Verhagen, T., Feldberg, F., van den Hooff, B., Meents, S. and Merikivi, J. (2012) Understanding users' motivations to engage in virtual worlds: A multipurpose model and empirical testing. *Computers in Human Behavior*, **28**(2): 484–495. <https://doi.org/10.1016/j.chb.2011.10.020>
- Vuopala, E., Hyvönen, P. and Järvelä, S. (2016) Interaction forms in successful collaborative learning in virtual learning environments. *Active Learning in Higher Education*, **17**(1): 25–38. <https://doi.org/10.1177/1469787415616730>
- Wang, C. X., Calandra, B., Hibbard, S. T. and Lefaiver, M. L. M. (2012) Learning effects of an experimental EFL program in Second Life. *Educational Technology Research and Development*, **60**(5): 943–961. <https://doi.org/10.1007/s11423-012-9259-0>
- Zokić, M., Boras, D. and Lazić, N. (2012) Say again. *International Journal of Education and Information Technologies*, **6**(4): 363–370.

About the author

Moonyoung Park, PhD, is an assistant professor in the Department of Curriculum and Instruction at the Chinese University of Hong Kong. His areas of specialization are aviation English assessment, language teacher education, computer-assisted language learning and assessment, and curriculum and instructional design.