Asia Pacific Educ. Rev. (2010) 11:235–242
DOI 10.1007/s12564-010-9072-6

Examining the practice of a reading-to-speak test task: anxiety and experience of EFL students

Heng-Tsung Danny Huang · Shao-Ting Alan Hung

Received: 24 May 2009 / Revised: 25 January 2010 / Accepted: 1 February 2010 / Published online: 2 March 2010

Abstract In the literature, little research has hitherto been conducted to examine the implementation of integrated speaking test tasks. This study, in response, set out to compare the anxiety induced by a reading-to-speak task and the anxiety produced by a speaking-only task and to explore students' experiences of taking the reading-to-speak task. Forty-seven Taiwanese EFL learners completed a reading-to-speak task and a speaking-only task, filled out a state anxiety inventory twice, and responded to an open-ended questionnaire. The paired *t*-test revealed that the reading-to-speak task produced a statistically comparable degree of anxiety as the speaking-only task, possibly owing to the additional anxiety inflicted by its reading component. The qualitative analysis disclosed that students preferred the reading-to-speak task to the speaking-only task, even though it provoked additional anxiety, featured additional challenges, and entailed different response strategies. In light of these findings, implications are proposed for oral assessment research and EFL oral assessment practices.

Keywords English as a Foreign Language (EFL) · Language assessment · Oral proficiency · State anxiety

H.-T. D. Huang (✉)
Foreign Language Education Program, The University of Texas at Austin, 3600 North Hills Dr. #214, Austin, TX 78731, USA
e-mail: danny123@mail.utexas.edu

S.-T. A. Hung
Department of English, National Kaohsiung First University of Science and Technology, 2 Jhuoyue R., Nanzih, 811 Kaohsiung City, Taiwan

Introduction

Integrated tests refer to tests in which "all the tasks are thematically linked and where the input that has been provided forms the basis of the response(s) to be generated by test-takers" (Lewkowicz 1997, p. 121). When compared to their independent counterparts, it is claimed that these tests embrace a higher degree of authenticity (Luoma 2004), equalize test-takers in terms of the information available for them to formulate arguments (Read 1990; Weigle 2004), and initiate positive washback on teaching. However, thus far, most research on integrated tests has centered attention on the writing tasks (e.g., Gebril 2006; Plakans 2008; Watanabe 2001; Weigle 2004). Only a fragment of the research attention has been paid to the practice of integrated speaking test tasks (Iwashita et al. 2008; Lee 2006). This study, in response, examined the test anxiety provoked by integrated speaking test tasks and explored test-takers' experiences during such tasks.

Literature review

Integrated test tasks have been shown to offer three benefits. First, they embrace higher levels of authenticity and predictive validity in academic contexts. As Butler et al. (2000) pointed out, students usually speak after being exposed to textual and/or aural input in the real-life academic context. Integrated test tasks, customarily crafted to include textual or aural input (Lewkowicz 1997), can better mirror the realistic language use tasks in the academic milieu and as such bring with them an air of authenticity (Luoma 2004).

Second, they are believed to promote equity or fairness. In evaluative situations, test-takers usually bring varying degrees of background knowledge to bear on the test tasks,

and those weaker in relevant prior knowledge are thus at a disadvantage when attempting to construct their argument or generate their opinions. As Read (1990) and Weigle (2004) argue, integrated test tasks, by furnishing textual or aural information ahead of time, can effectively mediate the undesirable impact exerted by background knowledge.

Third, they can initiate positive washback effects on teaching, in a way analogous to performance-based assessment (Shohamy 1995). In the face of high-stakes tests such as the TOEFL, instructional practices usually focus on material directly related to the exam at the expense of the development of skills and knowledge not tested on the exam. Such teaching-to-the-test practices would inevitably emphasize drill and rote learning and in turn impair the effectiveness of a curriculum. Integrated tests, taking the form of performance-based assessment, may encourage teachers to adopt "a more holistic approach to instruction" (Miller and Legg 1993, p. 12) where attention may be directed more equally to all skill areas.

Empirical studies on integrated test tasks

A number of studies have hitherto been conducted to shed light on the implementation of integrated writing test tasks, i.e., reading-to-write tasks, in the L2 context. For instance, Watanabe (2001) discovered that reading-to-write prompts could allow teachers to assess students' ability in identifying and integrating important information in the source texts into their own writing. Echoing this positive finding, Weigle (2004) found that an ESL proficiency test with an integrated reading–writing component boasted much higher pass rates and rater agreement and exerted a positive washback on the instruction geared toward the preparation for this test. In the same vein, Gebril (2006) reported that reading-to-write tasks generated scores equally reliable as those produced by independent writing tasks. Furthermore, Plakans (2008) found that reading-to-write tasks stimulated a more interactive process and enjoyed a higher degree of authenticity and student preference. Taken together, the findings in this strand of research suggest that integrated writing test tasks come as a reliable and authentic alternative that could generate favorable impact on both writing performance and writing instruction.

However, few systematic research efforts have thus far been made to investigate integrated speaking test tasks. Among them, Lee (2006) focused attention on the speaking component of the TOEFL-iBT that incorporates independent speaking tasks, reading–listening–speaking tasks, and listening–speaking tasks and found that an increase in the number of tasks influenced the score reliability and that the combination of the subtask scores into composite scores was justifiable, given the high score correlations among them. By the same token, Iwashita et al. (2008) explored the nature of speaking proficiency to bring to light the features that distinguish the performance on the speaking component of the TOEFL-iBT. Results disclosed that linguistic resources, phonology, and fluency all significantly affected the evaluation of the spoken performances.

Test anxiety

A vast body of literature has examined the issue as to how test anxiety relates to performance attainment in evaluative settings. On the whole, although some research indicating counter evidence also exists (e.g., Becker 1982), the collective findings of test anxiety studies portray a moderate, inverse association between test anxiety and performance in a wide range of testing contexts (Zeidner 1998). Stated differently, it was found that the high-test-anxious individuals tended to perform at lower levels when compared to their low-test-anxious counterparts on classroom exams (Alpert and Haber 1960; Morris and Liebert 1970; Salame 1984), placement tests (Madsen 1981), digit-symbol tests (Sarason 1984; Sarason and Stoops 1978), paired-associate learning tasks (Cubberly et al. 1986), intelligence tests (Morris and Liebert 1969), analogical reasoning tasks (Leon and Revelle 1985), among myriad other tests/tasks.

Furthermore, as previous studies converge to underscore the debilitating impact of test anxiety on academic performance in general, researchers have also begun to explore its relationship with performance on foreign languages. For instance, Chastain (1975) investigated the effects of test anxiety, reserved versus outgoing personalities, and creativity on learners' final grades in beginning modern language courses and found an incongruent pattern underlying the relationship between test anxiety and final grades; that is, test anxiety appeared to negatively correlate with French audiolingual class but positively with regular (traditional) French, Spanish, and German classes. Another relevant study was undertaken by Horwitz (1986). Marshaling the validity evidence of the Foreign Language Classroom Anxiety Scale (Horwitz et al. 1986), this study uncovered that test anxiety failed to correlate with the final grades in French classes in any significant fashion. Furthermore, Hembree (1988) employed a meta-analytic technique and found foreign language achievement to surface as one significantly but weakly inverse correlate of test anxiety. Similarly, Oya et al. (2004), examining the impact of personality and anxiety on oral performance, also disclosed that as ESL students' state test anxiety escalated, and their performance on the story-retelling task dropped significantly in clausal construction accuracy. In the same vein, Liu (2007) administered a custom-developed oral English test anxiety scale to Chinese students of English and uncovered that oral English test anxiety was significantly and inversely associated with the oral test performance.

Research questions

While research efforts on integrated writing test tasks have demonstrated their superiority over independent writing tasks, corresponding efforts on integrated speaking test tasks have not been made. Furthermore, although test anxiety has thus far been shown to play a significant contributing role in test performance in general and foreign language performance in particular, how it operates to affect performance on integrated speaking test tasks remains unclear. These research gaps combine to prevent the language testing researchers from clearly discerning whether integrated speaking test tasks would induce less anxiety on the part of test-takers when compared to independent speaking test tasks and in turn allow for a more straightforward score interpretation. In other words, is it the case that the scores on an integrated speaking test task can better represent test-takers' actual oral ability because they are less affected by anxiety? Moreover, the absence of relevant research also leaves language instructors who draw on integrated speaking test tasks to undertake their assessment practice with no empirical grounds on which to justify their assessment choice.

Therefore, with the intent to bridge these gaps and to offer empirical evidence to endorse the practice of integrated speaking test tasks, this study thus operationalized the integrated speaking tests with a reading-to-speak task and addressed the following questions:

1. Does the reading-to-speak (RTS) task induce a lower degree of state anxiety than the speaking-only (SO) task?
2. What are students' experiences of taking the RTS task?

Method

Participants

Forty-seven Taiwanese English-as-a-Foreign-Language (EFL) college students from two intact conversation classes were recruited. Consisting of 43 women and 4 men, this group of participants, aged between 19 and 22, possessed an intermediate level of English proficiency at the time of study. Prior to their participation, they had garnered an average of nine years of formal education in English.

Materials

The materials employed in the current study included a personal information sheet, two sets of prompts, a state anxiety inventory, and a questionnaire. Except for the prompts, the other materials were all presented in the students' native language, that is, Mandarin Chinese, with the intent to eschew potential language-induced misunderstandings.

First, the personal information sheet, comprised of five items, gathered information regarding the participants' age, gender, and educational background. To complete this sheet, students either ticked the box that best described them or produced responses in the written form.

Second, two sets of prompts provided the content for the oral tasks. The first set of prompts made possible the SO task and called upon test-takers to capitalize on their general knowledge in formulating responses. This prompt was crafted in compliance with the format that the Test of Written English (TWE) of the TOEFL follows to generate its prompts; that is, it began with several statements sketching out the context for the topic and conclude with a question inviting speakers to elaborate on their personal opinions and ideas pertaining to the topic. The specific prompt was as follows.

> It was reported on May 14th that same-sex couples in California, USA, may be able to obtain marriage licenses in the coming June. Those in support of this decision believe that gay people deserve the same right to marry like everybody else does, while the others claim that marriage should be between one man and one woman. Share your view on this issue. Use specific details or examples to support your opinions.

In terms of the second prompt, it underlay the RTS task and consisted of one reading passage and two questions. The reading passage, adapted from a CNN news report, revolved around how the earthquake in China and its one-child policy combined to exacerbate parents' trauma of losing their beloved children in the earthquake. As for the questions, they shared a close connection to this passage and instructed test-takers to discuss how the one-child policy worsened the parents' pain and what other measures the Chinese government could practice to grapple with its over-population issue. The choice of these two particular topics, namely, same-sex marriages and the earthquake in China, arose mainly from the fact that they constituted two of the most popular subjects at the time of data collection.

Third, the state anxiety inventory (SAI), drawn from the state-trait anxiety inventory (STAI) introduced by Spielberger and his colleagues (Spielberger and Gorsuch 1966; Spielberger et al. 1970), explored the respective anxiety level test-takers suffered while tackling the SO task and the RTS task. This scale is composed of 20 items that bring to light the anxiety reactions that respondents suffer in a given context. Among the 20 items, half of them are worded positively, such as "I felt upset," while the other half are

negatively worded statements, including "I felt happy." To respond to this scale, test-takers recorded an answer for each item on a four-point Likert scale with the options being "not at all," "somewhat," "moderately so," and "very much so." Since this study compared the state anxiety associated with taking the SO task and the RTS task, test-takers answered this scale twice as a way to reveal their states of emotional turmoil induced by these two different types of tasks. Furthermore, with the intent to avoid the potential language-induced misunderstandings, the Chinese version of this scale created and validated by Chung and Long (1984) was adopted in lieu of the original English version. Finally, the Cronbach's alphas for the ratings on these two SAIs reached a magnitude as high as .91 and .89, as such corroborating the scale reliability.

Fourth, a questionnaire made up of 10 open-ended items invited the test-takers to expound upon their experiences of completing the RTS task (see "Appendix 1"). To fill out this questionnaire, test-takers simply narrated in written form their thoughts and opinions on such issues as the anxiety they suffered, the difficulties they ran up against, the strategies they called into action, and so forth. Furthermore, in attempts to allow test-takers to fully express themselves, they were encouraged to respond to this questionnaire in their first language, that is, Mandarin Chinese.

Oral tasks

One SO task and one RTS task served as the principal test tasks for the current study. For the SO task, test-takers verbalized their opinions without the benefit of any input provided ahead of time. For the RTS task, they first read a passage and then voiced their thoughts in answer to two related questions. In terms of time allowance, for the SO task, test-takers were given 5 min to complete this task, that is, 3 min for preparation and 2 min for responding, whereas for the RTS task, they had 5 min for reading and preparation and another 5 min for recording their thoughts, totaling 10 min. With respect to task sequencing, students were instructed to first undertake the SO task and then perform the RTS task. The choice of this particular sequencing arose from two major considerations: administration practicality and task familiarity. First, computer laboratory schedules and the class schedules dictated that the participating students take both tasks as a whole group on the same day, effectively precluding the possibility of counterbalancing the tasks. Second, given that the majority of students in the present study had had antecedent experiences taking the SO task, it was thought that beginning with this task might work better to reduce the potential method effect (i.e., the influence brought about by the machine-mediated oral test format) than starting with the comparatively unfamiliar RTS task.

Procedure

The administration of this study spanned ~1 h. To begin, the researchers carefully explained the procedures to the students and answered all questions they had, followed by the students filling out the personal information sheet. Next, the students performed both the SO task and the RTS task abiding by the time allowance. Subsequently, they completed the STI twice to indicate the respective levels of anxiety that they endured in grappling with the two types of oral tasks. Finally, the data collection concluded with the students responding to the open-ended questionnaire to share their experiences in taking the RTS task.

Data analysis

The data gleaned in this study were reported or analyzed in the following manners. First, students' demographic and educational backgrounds gathered from the personal information sheet were simply reported to sketch out a general picture of the group. Second, students' self-ratings on the state anxiety scale were analyzed by means of several consecutive steps. To begin, the researchers divided the 20 items on each SAI into two groups on the basis of their wording orientations, namely positively worded or negatively worded. Next, the responses to each item were coded in compliance with the coding scheme shown in Table 1. As demonstrated by this table, for the positively worded items, the researchers replaced the four response categories with four ascending numerical values, i.e., one for "not at all," two for "somewhat," three for "moderately so," and four for "very much so." As per the negatively worded items, the coding was reversed by assigning four numbers in descending order. That is, for the negatively worded items, four stood for "not at all," three for "somewhat," two for "moderately so," and one for "very much so." Subsequent to this coding practice, the researchers then summed the numbers on the two SAIs into two composite scores to represent their state anxiety levels induced by the two oral assessment tasks. Finally, following this summing procedure, the researchers called into action a paired *t*-test to statistically compare these two

Table 1 Coding scheme for SAIs

Group	Category			
	Not at all	Somewhat	Moderately so	Very much so
Positively worded items	1	2	3	4
Negatively worded items	4	3	2	1

scores as an attempt to shed light on the question as to whether the students experienced a similar degree of state anxiety in tackling the SO task as in coping with the RTS task.

Third, the questionnaire responses were qualitatively analyzed with reference to four a priori categories—anxiety, difficulties, strategies, and preferences. The analysis procedure began with the researchers carefully reading, re-reading, and then coding the responses into the aforementioned four categories. Second, they inspected all the entries in each category in search for emerging themes that best captured students' experiences of taking the RTS task. Third, as these representative themes emerged, the researchers then revisited the responses and examined if the analyses they had done so far would lead them to see the coding in a different light and thus call for changes. Finally, the researchers generated interpretations for the emergent themes. Throughout the entire process, the researchers checked for conflicting evidence by constantly referring back to the questionnaire responses and making adjustments to the interpretations accordingly, as an attempt to lend better credibility to the findings.

Results and discussions

This section presents and discusses the results of the statistical analysis performed on students' ratings on the SAIs and the qualitative examination conducted on their responses on the open-ended questionnaire. For the statistical analysis, the minimal level of significance was set at .05.

Research question one

The first research question inquired whether the RTS task would induce a lower degree of anxiety than the SO task. To respond to this question, a paired *t*-test was performed to compare the anxiety levels that students experienced in the act of completing the two different oral test tasks. Table 2 lays out the results.

Table 2 Paired *t*-test comparing the mean differences between students' levels of state anxiety in taking the RTS task and the SO task

State anxiety	*N*	*M*	SD	Skewness	Kurtosis
RTS task	47	43.62	14.58	.08	−1.08
SO task	47	44.91	11.24	−.14	−.86

Comparison	Paired differences				
	M	SD	SE	*t*(46)	Sig.
RTS task versus SO task	−1.30	22.95	3.35	−.39	.70

As evinced by this table, both of the state anxiety ratings were characterized by a normal distribution, given that the absolute values of their skewness and kurtosis statistics all fell below the value of two (Garson 2009). Moreover, the state anxiety rating, on average, emerged to be higher for the SO task ($M = 44.91$, SD = 11.24) than for the RTS task ($M = 43.62$, SD = 14.58). Referring to this finding and the maximum rating of the SAI, viz., 80 points, the students in the present study appeared to experience merely a moderate degree of anxiety when grappling with both tasks.

As Table 2 further makes clear, the difference between the state anxiety ratings for the two tasks failed to reach the minimal level of statistical significance ($t(46) = -.39$, $p = 0.70$). This non-significant finding suggested a negative answer to the first research question; that is, students did not suffer a significantly lower level of anxiety when dealing with the RTS task than when coping with the SO task.

This finding ran counter to the assumption that the RTS task constitutes a less anxiety-provoking oral assessment measure. By way of interpretation, this statistically comparable level of anxiety might have stemmed from the additional anxiety inflicted by the RTS task. As a majority of students noted in their responses to the questionnaire, although the reading passage of the RTS task did help reduce their anxiety by offering ideas they could appropriate in generating their oral responses, this very passage simultaneously induced further anxiety by provoking such undesirable emotions as fear for misunderstanding and worry over being unable to finish reading the passage within the given time allowance. In other words, the RTS task decreased the anxiety related to the lack of relevant background knowledge but concurrently increased the anxiety arising from the pressure of needing to deal with the provided input, eventually appearing as equally anxiety-producing as the SO task.

Research question two

The second research question explored students' experiences of completing the RTS task. Toward this end, the researchers conducted a qualitative examination of the students' written responses to the open-ended questionnaire and as such derived four themes that functioned to reflect students' experiences of performing the RTS task, each representing one of the four a prior categories (anxiety, difficulties, strategies, and preferences). What follows brings into focus each of these themes.

The RTS task induced additional anxiety

A number of students indicated experiencing additional anxiety when taking the RTS task due mainly to the reading passage that preceded the response task. In their eyes, the reading passage, on the one hand, functioned to

contextualize the test and supply/activate pertinent knowledge and as such allayed the anxiety springing from the lack of ideas on which to elaborate. However, on the other hand, this passage imposed further cognitive loads by demanding comprehension of the textual ideas within the time limits and thus led to the formation of additional anxiety. As one of the students put it,

> When taking the RTS task, I needed to first read the passage and then come up with answers. The reading part made me really anxious because I was afraid that I might not understand the passage or could not finish reading it in time.

The other reason that surfaced rather frequently as a source of the additional anxiety pertained to the constraint on creative responses. A host of students believed that the RTS task delimited the path they could take in answer to the questions and in turn stifled their creativity. For instance, one student mentioned that "the content of the passage disallowed me to express my own thinking."

The RTS task incurred additional challenges

Most students agreed that the SO task emerged as the more difficult task because of the absence of input and the necessity of relying completely on themselves for ideas. However, some pointed out that the RTS task, featured by two major challenges, could also be rather difficult. These challenges included the possible conflict between students' prior knowledge and the ideas presented in the passage and the time required for preparation. In terms of the ideational conflict, a few students found the message conveyed by the passage differed from their previous relevant knowledge on the subject and as such felt disoriented in formulating their answers. Turning to the time requirement issue, students expressed that the RTS task called upon them to perform multiple tasks, i.e., reading and taking notes, and thus demanded comparatively more preparation time than the SO task. The following comment highlighted these two challenges.

> When taking the RTS task, it took me longer to get ready to speak because I had to do several things at one time, such as reading, taking notes, and thinking about what I wanted to say… Some ideas in the passage were different from what I know about that topic, so I was confused but decided to go with the passage because I did not want to lose points.

The RTS task entailed specific response strategies

Students reported that when participating in the SO task, they put into practice three principal response strategies. The first pertained to jotting down keywords and the second related to outlining their answers on the basis of these keywords. Third, they silently rehearsed their answers by referring to their generated outline. However, when it comes to the RTS task, they indicated that they did not merely call into action the response strategies they employed for the SO task, but mobilized three additional strategies specific to this task. First, they reverted to the questions as a guide to gain a better understanding of the reading passage. Second, they appropriated the ideas presented in the reading passage as a way to strengthen their response. Third, they noted down the vocabulary items and phrases that could lend themselves to expressing their intended messages. As one student elaborated,

> When reading the passage, I underlined the words and phrases that I might use to talk about my ideas. I felt that if I used these words and phrases, my answer would be more connected to the topic and more convincing.

The RTS task met with immense preference

Although the RTS task came with additional anxiety and challenges and required different response strategies, an overwhelming number of students still expressed preference for this task when prompted to choose between the two types of test tasks. As observed from students' responses, this preference, for the most part, sprang from two primary reasons. First, students favored the RTS task because it offered input that not only activated their pertinent background knowledge but featured ideas and vocabulary items that could be readily incorporated into their own oral response. Second, they voted for the RTS task on account that its reading component served as an inspirational guide and hinted at the direction they could take in producing their answers. For example, one student pointed out that the reading passage allowed him/her to quickly think of and outline what to say and how to say it.

Conclusion and implications

This study set out to investigate the anxiety induced by the RTS task and the SO task and to explore students' experiences of taking the RTS task. Following the statistical analysis and the qualitative examination, three major findings sprang forth. First, the RTS task produced a statistically comparable degree of anxiety to the SO task, possibly owing to the additional anxiety inflicted by its reading component. Second, some students considered the RTS task difficult because it posed additional challenges concerning ideational conflicts and preparation time. Third, the RTS task stimulated the use of unique response

strategies, including previewing the question and appropriating the ideas, vocabulary, and phrases presented in the passage. Lastly, a tremendous preference for the RTS task was shown on the part of the students because its reading element activated/furnished background knowledge and acted to inspire and guide the oral production task.

In light of these findings, two implications are proposed for oral assessment research and EFL oral assessment practices. For oral assessment research, the RTS task, vis-à-vis the SO task, does not produce scores that lend themselves more to straightforward interpretations. As anxiety has been shown to impose a negative impact on test performance and would as such muddy score interpretations, if an oral test task was less affected by anxiety, the interpretations of the scores it generates would then be more straightforward. That is, the test users could interpret the score as more accurately reflecting the test-taker's ability. However, now that the RTS task has been demonstrated to induce a similar level of anxiety as the SO task, it implies that the RTS task does not permit straightforward score interpretations any more than the SO task does.

For EFL oral assessment practices, it is advised that EFL instructors capitalize on the RTS task to perform their oral evaluation in an attempt to derive more reliable estimates of students' oral proficiency. As shown by previous research (e.g., Madsen et al. 1991), test-takers' attitudes toward a test (i.e., test affect) influences how involved they are in taking that test. If the involvement is low, the test score only captures a partial, if not inaccurate, picture of the test-takers' ability. Put differently, if test-takers embrace a positive attitude toward a certain test, they take that test more seriously, which in turn renders their performance on that test a more reliable representation of their ability. Since most students responded to the RTS task in a favorable manner, the use of this task to engage oral assessment might thus put the testers in a better position to more reliably reveal students' oral proficiency. Therefore, although this integrated speaking test task does not produce scores less affected by anxiety, it should still take precedence since it finds more favor in students' eyes.

Future research

Granted that the present study emerged as but one research attempt, it follows that more empirical endeavors are warranted if the practice of the RTS task is to be promoted with greater confidence. Seven viable research avenues for future exploration are delineated as follows. First and foremost, a larger number of participants could be included in order to maximize the generalizability of the statistical findings to the targeted population. Second of all, in lieu of the machine-recorded format, a live interview modality could be harnessed to better simulate the real-life communication situations and as such increase the authenticity of the RTS task. Thirdly, the performance on the RTS could be compared and contrasted with that on the SO task in an effort both to reveal the task effect and to unveil the different discourse features to which the two types of test tasks attach importance. Fourth, the sequence of administering the SO task and the RTS task could be counterbalanced so as to rule out the potential order effect. Fifth, as previous research has indicated that men and women tended to exhibit different levels of anxiety in the foreign language learning context (e.g., Andrade and Williams 2009; Williams and Andrade 2008), a roughly equal number of male and female students could be recruited to circumvent the possible gender effect. Sixth, as cultural membership has been shown to constitute a source of variation in the intensity of anxiety experienced and the responses to anxiety on the part of college students (e.g., Ishii et al. 1978; Matsumoto et al. 1988), a study comparing Asian learners and non-Asian learners in terms of their anxiety reactions to the two types of oral tasks could be conducted to cross-validate the results derived in the current study. Last but not least, interviews could be employed to obtain more in-depth qualitative information regarding students' experiences of and perceptions toward the practice of the RTS task and to generate triangulation data that could better function to explain and/or consolidate the quantitative results.

Appendix 1: The questionnaire probing the experience of taking the RTS task

1. Which task made you more anxious, the SO task (i.e., the task *without* a reading passage) or the RTS task (i.e., the task *with* a reading passage)? And what made you anxious when taking the RTS task?
2. Which of the two tasks did you find more difficult? And what difficulties did you encounter in taking the RTS task?
3. What strategies did you use to prepare for the RTS task?
4. In your opinion, what are the advantages of the RTS task?
5. In your opinion, what are the disadvantages of the RTS task?
6. What do you think about the reading passage provided by the RTS task? For instance, how appropriate was its difficulty level? How useful was the information if provided? How adequate was the English presentation of this passage? Etc.
7. What do you think about the question of the RTS task? For instance, how relevant was it to the reading

passage? How much did it allow for creativity? How much background knowledge did it require? Etc.
8. Do you think the RTS task can better reveal students' oral proficiency level? Why or Why not?
9. Which of the two tasks do you prefer and why?
10. Do you have any other thoughts about the RTS task?

References

Alpert, R., & Haber, R. N. (1960). Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology, 61*, 207–215.

Andrade, M., & Williams, K. (2009). Foreign language learning anxiety in Japanese EFL university classes: Physical, emotional, expressive, and verbal reactions. *Sophia Junior College Faculty Journal, 29*, 1–24.

Becker, P. (1982). Towards a process analysis of test anxiety: Some theoretical and methodological observations. In H. M. van der Ploeg & C. D. Spielberger (Eds.), *Advances in test anxiety research* (Vol. 1, pp. 11–17). The Netherlands: Swets & Zeitlinger.

Butler, F. A., Eignor, D., Jones, S., McNamara, T., & Suomi, B. K. (2000). *TOEFL 2000 speaking framework: A working paper (TOEFL Monograph No. 20)*. Princeton, NJ: Educational Testing Service.

Chastain, K. (1975). Affective and ability factors in second language acquisition. *Language Learning, 25*, 153–161.

鍾思嘉、龍長風 [Chung, S. C., & Lung, C. F.]. (1984). 修訂情境與特質量表之研究 [Modifying the state-trait anxiety inventory]. *測驗年刊 [Psychological Testing], 31*, 27–36.

Cubberly, W., Weinstein, C. E., & Cubberly, R. D. (1986). The interactive effects of cognitive learning strategy training and test anxiety on paired-associate learning. *Journal of Educational Research, 79*(3), 163–168.

Garson, G. D. (2009). *Statnotes: Topics in multivariate analysis.* Retrieved 12/21/2009 from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Gebril, A. (2006). *Independent and integrated academic writing tasks: A study in generalizability and test method.* Unpublished doctoral dissertation, The University of Iowa, Iowa City.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47–77.

Horwitz, E. K. (1986). Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *TESOL Quarterly, 20*, 559–564.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70*, 125–132.

Ishii, S., Cambra, R., & Kloph, D. (1978). Communication anxiety: A comparison of Japanese and American college students. *Otsuma Review, 10*, 86–94.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24–49.

Lee, Y. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing, 23*(2), 131–166.

Leon, M. R., & Revelle, W. (1985). Effects of anxiety on analogical reasoning: A test of three theoretical models. *Journal of Personality and Social Psychology, 49*, 1302–1315.

Lewkowicz, J. A. (1997). The integrated testing of a second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Language testing and assessment 7* (pp. 121–130). Dordrecht, The Netherlands: Kluwer.

Liu, M. (2007). Language anxiety in EFL testing situations. *ITL International Journal of Applied Linguistics, 153*, 53–75.

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Madsen, H. S. (1981). Determining the debilitative impact of test anxiety. *Language Learning, 32*(1), 133–143.

Madsen, H. S., Brown, B. L., & Jones, R. L. (1991). Evaluating student attitudes toward second-language tests. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 65–86). Englewood Cliffs, NJ: Prentice Hall.

Matsumoto, D., Kudoh, T., Scherer, K., & Willbott, H. (1988). Antecedents of and reactions to emotions in the United States and Japan. *Journal of Cross-Cultural Psychology, 19*(3), 267–286.

Miller, M., & Legg, S. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement:Issues and Practice, 12*, 9–15.

Morris, L. W., & Liebert, R. M. (1969). Effects of anxiety on timed and untimed intelligence tests: Another look. *Journal of Consulting and Clinical Psychology, 33*, 240–244.

Morris, L. W., & Liebert, R. M. (1970). Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance. *Journal of Consulting and Clinical Psychology, 35*, 332–337.

Oya, T., Manalo, E., & Greenwood, J. (2004). The influence of personality and anxiety on the oral performance of Japanese speakers of English. *Applied Cognitive Psychology, 18*, 841–855.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing, 13*, 111–129.

Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes, 9*, 109–121.

Salame, R. F. (1984). Test anxiety: Its determinants, manifestations, and consequences. In H. M. van der Ploeg, R. Schwarzer, & C. D. Spielberger (Eds.), *Advances in test anxiety research* (Vol. 3, pp. 83–119). The Netherlands: Swets & Zeitlinger.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46*(4), 929–938.

Sarason, I. G., & Stoops, R. (1978). Test anxiety and the passage of time. *Journal of Consulting and Clinical Psychology, 46*, 102–109.

Shohamy, E. (1995). Performance assessment in language testing. *Annual Review of Applied Linguistics, 15*, 188–211.

Spielberger, C. D., & Gorsuch, R. L. (1966). The development of the state-trait anxiety inventory. In C. D. Spielberger, & R. L. Gorusch. *Mediating processes in verbal conditioning*. Final report to the National Institutes of Health, U.S. Public Health Service on Grant MH-7229, MH-7446, and HD-947.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. D. (1970). *STAI: Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.

Watanabe, Y. (2001). *Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions*. Unpublished doctoral dissertation, University of Hawaii, Manoa.

Weigle, S. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing writing, 9*, 28–47.

Williams, K., & Andrade, M. (2008). Foreign language learning anxiety in Japanese EFL university classes: Causes, coping, and locus of control. *Electronic Journal of Foreign Language Teaching, 5*(2), 181–191.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum Press.