A Case Study of VIP TOEFL Speaking Teachers

Tongxuan Huang
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/creativecomponents

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
Huang, Tongxuan, "A Case Study of VIP TOEFL Speaking Teachers" (2018). Creative Components. 27. https://lib.dr.iastate.edu/creativecomponents/27

This Creative Component is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Creative Components by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
A Case Study of VIP TOEFL Speaking Teacher: Teaching and Perceptions of Changing Speaking Tasks

Tongxuan Huang

Iowa State University
Dedication

For my families, for their support on my decision to study abroad, and I greatly appreciate the guidance and help my major Professor, Dr. Gary J. Ockey, offered me, thank you.

For Nazlinur Gokturk-Tuney and Haeyun Jin, thank you for your advices and suggestions.

Thanks for the advices my committee gave me, Dr. Donna Niday and Dr. James Ranalli.

Thank you, LASO (Language Assessment Student Organization) members for your suggestions on the questionnaire.
Abstract

This study examines several cases of Chinese VIP-TOEFL-prep speaking teachers (how do they teach and what do they use in class). This study also seeks to elicit those teachers’ perceptions of some communicative speaking tasks, which are one-on-one interviews, paired/group discussion, computer as interlocutor. Paired/group discussion and presentation in virtual environment were used in interviews and questionnaires as additional communicative speaking tasks. These additional tasks were selected as examples of computer mediated speaking test tasks. Findings suggest that institution’s policy, student English proficiency level, teachers’ beliefs are contributing factors in deciding whether a class is test-task centered or communicative style. Teachers who have a test-task centered teaching style will continue to teach new test tasks if the test adopts new tasks. Teachers who have a more communicative teaching style either shows concerns about changes or do not want to change their teaching style. This study also confirms previous studies’ findings that testing context in China regards tests as gateway to pass and prefer teaching to test tasks, which would result in negative washback for tests. Based on the opinions of the teachers in this study, the author proposed that it may lead to a more positive washback to raise teachers’ awareness or understanding of test construct.

Keywords: Washback, communicative speaking tasks
The term “washback” has different variations of definition, slightly different in wording. Generally, it is the impact of a test on students, teachers, pedagogy, even society, and other shareholders or other factors related to this one test. What concerns teacher researchers is how a test will affect the way they teach and what do they use to teach, and how these new teaching tools will affect their students. This study aims at a particular test, TOEFL (TOEFL as a high-stakes international test plays a crucial role for students who are from a non-English speaking country to gain admission from colleges in English speaking countries), and a special kind of school, called “cram schools” in China. Nowadays they would prefer the term test prep institutions, which may be an indication of leaders of these institutions trying to change their way of doing things. However, there is little research on the motivations of test-prep-school teachers for what they do in their classes (Yu et al., 2017). This study aims to provide insights into the ways that teachers currently help students prepare for TOEFL iBT in one type of TOEFL prep course and why they choose a certain method, and how they might change their ways of helping students prepare for TOEFL iBT if the test were made more communicative. The next section provides a brief review of washback in the language testing research context and some studies on the washback of TOEFL iBT.

About TOEFL iBT

From an official introduction to TOEFL program: “Since its inception in 1963, the TOEFL test has evolved from a paper-based test, to a computer-based test and, in 2005, to an internet-based test. But more important than changes in the method of test delivery are the changes in the theories of language proficiency that motivated the redesign of the test’s construct and content.” (Enright, n.d.) In a summary of all three stages, the TOEFL program had gone through, Enright (n.d.) stated that the construct of communicative competence guides the design
of the new TOEFL iBT, specifically “communicative language use in academic context”. A series of studies were conducted to support the design and use of this new TOEFL iBT, which now has four language domains (reading, listening, speaking, and writing) in its construct.

**Literature Review**

I begin with a quote from a TOEFL iBT development advisor: “Since TOEFL prep courses will probably always be in existence, they might as well help students become as proficient as possible at something academic and authentic” (Wall & Horák, 2006, p. 14). It might be true that TOEFL prep courses or any other prep courses for other high-stakes tests will always exist. However, it also might be true that not all TOEFL prep courses can offer students a chance to be proficient at the abilities the test is targeting. The issue here is the extent to which the course content and syllabi are well aligned with the constructs of the abilities the test has: are the teachers on the same page with the test developers on the constructs? do the teachers have adequate knowledge of teaching the English language and have appropriate materials for prep courses? what do the teachers think about teaching to the test?

Previous language testing research that tries to answer these questions or similar questions set them under the term "washback", which refers to the influence testing has on teaching and learning (Barnes, 2016; Cheng, Watanabe, & Curtis, 2004). The definition of washback may vary in how it is phrased, or it would be similar to the notion of the usefulness of Bachman and Palmer, which, in this study’s case, they would probably refer to as micro impact, which relates to the effects on classroom instruction (Bachman & Palmer, 1996). The consensus is that washback is a complex phenomenon and is better studied in its context (which could be cultural, institutional etc.) for one given test (Barnes, 2016; Cheng, Watanabe, & Curtis, 2004; Turner, 2009; Wall, 2000).
Watanabe (1997, as cited in Cheng, Watanabe, & Curtis, 2004) attempts to disentangle the complexity of washback by defining several dimensions that represent various aspects of washback: specificity (whether it is general or specific), intensity (whether it determines everything in a classroom or only part of the things in a classroom), length (the duration of the influence of a test), intentionality (whether the influence is what test developers have planned), and value (whether it is positive or negative). What the current study attempts to investigate is a specific washback—what is being taught in classroom, and teachers' attitude to TOEFL iBT and teaching TOEFL iBT and their teaching methodology. Moreover, the Chinese context for any test, language tests included, are “the essential role that examinations have played and continue to play in Chinese society being successful in examinations is the key to succeed in study, work and life.” (Cheng & Curtis, 2010). Thus, the washback of TOEFL in China may be a different case compared to that of other countries.

**Context of Teaching**

Among some teachers and parents in China, there seems always to be a criticism of the way teachers teach English. They contend that Chinese students lack the ability to communicate in English fluently, either in daily conversations nor in an academic environment. In 2000, China Central Television held the first Star of Outlook talent competition to promote English learning nationwide, featuring students from elementary schools to colleges competing in a series of tests, prepared and spontaneous speeches, and conversations. At the time, there was only a small portion of the students, those who either wanted to apply for an English-speaking country’s college or wanted to gain fame by participating in such an event, who cared about practicing speaking. The majority of students did not have the opportunity or (probably) motivation to practice speaking. In 2005, following the introduction of TOEFL iBT with its new speaking
section, test-prep institutions and companies, along with the emerging private high schools, started to build new curricula that follow a four skills paradigm (listening, speaking, reading, and writing) or kept using foreign curricula (A-level courses, K-12 curricula etc.). Thus, two distinct teaching contexts can be identified: (1) regular public school English classes or college English classes, which focus on passing the college entrance exam or the college English test; and (2) test prep institutions and companies, private high schools, and parts of public high schools, which incorporate speaking practice into their curricula.

Since the college entrance exam and the college English test currently do not have a speaking section, public high school teachers and college English teachers do not have the authority to spend much class time for students to practice speaking. Also, many students do not think much about speaking, as WJS reported that "They (students) care more about CET4 or CET6". It is interesting that the CET official introduction document (Aims and impact of CET, n.d.) states that CET is a formative test. However, many of the college English classes have test centered sessions that may last a whole semester. For English majors in some colleges, test-centered sessions are also part of their courses (Zhang, 2018). There are all sorts of political and/or economic considerations to have so many test centered sessions in public schools. The point here (in a sense of the context of TOEFL in China) is that the education system has a tendency that where there is a test, there is a need for students to pass it for the political and/or economic benefits that passing it can bring about. Teacher Wind would put it in this way “in China, the thing is, most people understand or see the test as a-- well, as a gate you have to get through, right? It's a basic requirement to get into a college in the US or other foreign countries”. Therefore, the classes should focus heavily on test tasks rather than improving the abilities a test is measuring. However, I am not trying to say the abilities students need do not improve.
Certainly, what this tendency results in is that if a test does not assess some abilities, in this case speaking, the classes also omit or do not have enough practice of those abilities. Thus, language classes are not about teaching a language, but are more about passing a language test. To sum up, the education system may be prone to the dark side of washback, which is teaching to test tasks, and ignoring anything else that is not in test tasks. When the tasks that are used on the test are not consistent with the tasks that are expected in real world target use tasks, negative washback results from the tests. Other studies had shed some light on the washback of TOEFL, which will be presented in the following section.

**Previous Research on the Impact of TOEFL**

Alderson and Hamp-Lyons (1996) observed both TOEFL prep classes and non-TOEFL-prep classes in a specialized language institute in the USA, followed by interviews with the two teachers who taught both kinds of classes. One of the claims they wanted to investigate is still among the concerns that people would have about language testing (or testing in general): “Courses that raise TOEFL scores without providing students with the English they will need in language interaction or in the college or university courses they are entering” (Alderson & Hamp-Lyons, 1996, p. 1). They concluded that there was a clear influence of TOEFL on both what and how teachers teach, but it differed from teacher to teacher. From their interviews with teachers, they found a more negative attitude to teaching TOEFL and teachers would claim that teaching to TOEFL was “easy” and it was “hard to make the class fun”. Furthermore, some teachers expressed concerns about not being able to ensure improved scores for their students. Although the TOEFL they studied was quite different from the TOEFL iBT now, it is worth noticing the attitudes and concerns of teachers teaching TOEFL preparation courses.

**Study in Europe**
Wall and Horák (2006, 2008, 2011) conducted a longitudinal study to gain a deeper understanding of the washback TOEFL had on teaching and learning in Central and Eastern Europe. Phase 1 of their study (Wall & Horák, 2006) provided a description of what TOEFL preparation courses were like in Central and Eastern Europe before TOEFL iBT was in practice in 2005. They concluded the following:

“It was recognized that the students had had years of English study before coming to the TOEFL classes, and their purpose was enough to do well on the TOEFL test. The students were clients paying for a service, rather than students in a school system where the curriculum objectives were much greater than the objectives of the final examination and where examination caused “curriculum distortion”. In the case of these institutions, the contents of the TOEFL test were the curriculum, and examination preparation was the students’ only reason for being there.” (Wall & Horák, 2006, p. 119).

Phase 2 (Wall & Horák, 2008) of their study found that test-preparation textbooks were central to the courses they observed. Thus, in Phase 3 (Wall & Horák, 2011) they examined the content and methodology that is aligned with the textbooks used in the courses they observed in Phase 2. Findings in Phase 3 suggested that the course content changed considerably to match the changes brought by the new TOEFL iBT, and new textbooks designed for TOEFL iBT provided a teacher with content and a sense of security in teaching to the test. Based on these findings, I would expect that possible changes to the speaking section would also lead to further changes in the teaching of TOEFL materials.

Study in Vietnam
In an attempt to investigate the influence of TOEFL iBT on what is taught and how it is taught in classrooms in Vietnam, Barnes (2016) observed and interviewed four English language teachers from two language centers in Vietnam and came to the conclusion that teachers highly rely on commercial test-preparation materials and their teaching methods are often overshadowed by the structure of the test prep course. Barnes (2016) argues that there is an influence on the English language classes from TOEFL iBT, but it is mediated by the use of TOEFL iBT textbooks.

These studies depict a sketch of what TOEFL preparation courses are/were like and some insights into what the washback of TOEFL on these courses was. Notably, Barnes and Wall and Horák emphasize the role of the test preparation textbook in influencing washback. Specifically, Wall and Horák state that “communication between the test designers and teachers and students and between testing agencies and authors and publishers who design the textbooks was extremely important” (Wall & Horák, 2006). It would be interesting to investigate if it is the same in Chinese TOEFL preparation courses as the courses in the research mentioned above.

**What is the Intended Impact of TOEFL iBT**

Wall and Horák (2006) summarized the intended impact of TOEFL iBT in general by reading the framework documents of new TOEFL and surveying some of the advisors for the new TOEFL. Most of the impact mentioned in the documents is quite general. (e.g., a positive effect on language instruction and applied linguistics and focus more on communicative language in use in academic contexts) (Wall & Horák, 2006, p. 11). A similar conclusion was drawn from the survey: “the new TOEFL would lead to a more communicative approach to teaching and preparation classes would pay more attention to academic tasks and language, there would be more speaking, there would be integrated skills work, and some aspects would change in the teaching of other skills” (Wall & Horák, 2006, p. 17).
With the above conclusion in mind, I couldn’t help wondering: “Is there more speaking in Chinese classrooms? What are the TOEFL teachers doing in their classes?” In a more recent study, Yu et al. (2017) surveyed the Chinese TOEFL test takers on how they prepare for TOEFL speaking. Their research yielded a quantitative description of what activities Chinese TOEFL test takers do and how long these activities take (Unfortunately, not many interactive activities are there). Yu et al. (2017) suggested that students and teachers have an agreement on the usefulness of each activity and there is a weak correlation between test preparation and test performance. They also suggested that little was known about the motivations of teachers, and why they teach in the way they do now. Thus, this study seeks to address such question by interviewing some teachers from test prep institutions: how they prepare students for taking TOEFL, what teachers think about TOEFL, and if TOEFL were to change its speaking tasks to more interactive ones, how this might affect the types of materials and teaching practices of teachers in these schools.

The teachers Yu et al. (2017) had are from test preparation schools. However, they did not specify which type of test prep courses those teachers were teaching at the time. There are two major types of test-prep courses in China: one usually has more than 20 or 30 students, the other has only one (that’s one of the reasons why it’s called VIP course). This study primarily focusses on the teachers from VIP course. In addition, the students mention in this paper are mostly high school students (either public or private), some middle school student (sometimes an institution or company has primary school students).

Thus, the two research questions I have in this study:

1) How do VIP TOEFL teachers teach test preparation speaking classes?
2) How might including some interactive speaking task types affect teachers’ instruction and attitudes toward teaching TOEFL?

Methods

Participants

There were 10 participants in total. They are working in Beijing (n=3), Qingdao (n=2), Chongqing (n=3), Puyang (n=1), and in Canada (n=1, she is an independent teacher and teaches her classes online). Seven of them are TOEFL-VIP-speaking teachers. (VIP-TOEFL-courses are one-one-one English courses provided by test prep institutions/companies to Chinese students. Many of the courses follow the four-skill paradigm, i.e. reading, listening, speaking and writing.) Two of them are VIP-speaking-teachers (VIP-speaking-courses are general English one-on-one courses aimed at improving students’ abilities of communicating in English for reasons other than academic communication or passing English tests). One of them is a college English teacher. They are either former colleagues of the researcher or were recruited by the former colleagues of the researcher. A pseudonym was assigned to each participant at the time of the interview or upon the completion of the questionnaire.

Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Teaching English</th>
<th>Teaching TOEFL</th>
<th>Gender</th>
<th>Other than iBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amai</td>
<td>4</td>
<td>2</td>
<td>F</td>
<td>N</td>
</tr>
<tr>
<td>Cardinal</td>
<td>10</td>
<td>8</td>
<td>F</td>
<td>N</td>
</tr>
<tr>
<td>Ivy</td>
<td>4</td>
<td>4</td>
<td>F</td>
<td>N</td>
</tr>
<tr>
<td>Ly</td>
<td>5</td>
<td>5</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>Wind</td>
<td>6</td>
<td>5</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>Zls</td>
<td>5</td>
<td>5</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>Hemi</td>
<td>7</td>
<td>7</td>
<td>F</td>
<td>N</td>
</tr>
<tr>
<td>Average</td>
<td>5.8</td>
<td>5.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1 shows how many years the TOEFL-speaking teachers have been teaching English and teaching TOEFL speaking, ranging from two years to eight years for TOEFL speaking. Since the TOEFL iBT was introduced in 2005, none of them are familiar with previous versions of the TOEFL test. Three of them started as a TOEFL prep teacher and three of them started as general English teachers or teaching assistants. This sample of TOEFL prep teachers has a balanced gender distribution.

Table 2

<table>
<thead>
<tr>
<th>Name</th>
<th>Teaching English</th>
<th>Teaching speaking</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hanny</td>
<td>6</td>
<td>6</td>
<td>F</td>
</tr>
<tr>
<td>Yuki</td>
<td>2</td>
<td>2</td>
<td>F</td>
</tr>
<tr>
<td>Wjl</td>
<td>12</td>
<td>4</td>
<td>F</td>
</tr>
</tbody>
</table>

Table 2 shows the length of being an English teacher for three non-prep teachers. It is worth noticing that Wjl has been an English teacher for twelve years and has just been teaching speaking for four years. It is because of a change in the English teaching syllabus in her institution. This change was made to follow the Ministry of Education's suggestion on English education for post-secondary institutions, which recommends "promoting students' abilities to communicate in English in academic settings and abilities to conduct daily conversations." It should also be mentioned here that Hanny is currently one of the directors of a group of non-test-prep speaking teachers. Her company is making an effort to combine non-test-prep teaching with test-oriented teaching. There will be more details in the discussion section.

Those who are not TOEFL speaking teachers are included to provide some details of the overarching teaching context for English teachers in China.
An Independent TOEFL Teacher (The Case of Cardinal) Cardinal graduated from Peking University, works on her own, and does not answer to any test prep institutions or companies. As shown in Table 1, she has been teaching TOEFL preparation classes for eight years now. She is confident and knowledgeable. "I scored 30 in the reading, listening and writing sections, which makes it easier for me to address students’ concerns.” Interestingly, she did not mention how she did on the speaking section.

Her classes are online one-on-one classes. She did not give too much detail on what she did with each student. But, from her brief description of her class, I was able to learn that her class starts with a test, in which she assesses students' English proficiency level. Then, they set a schedule together, in which the students are required to do the following assignments and send them to her: word dictation, watching TED talks and summarizing or retelling, picking a TV show to watch or be assigned one. Her one-on-one class was described as follows: "I usually try to strike a balance between improving their scores and improving their general English. I also come up with solutions to cultivate their interests and develop a habit (or the skills) of learning on their own after my classes.” She also started study groups online with her students under one rule, “English only”. One of the things she believes in is immersion learning, and she is doing what she can as an online teacher to provide more opportunities for her students to use English outside of class.

In her case, her knowledge about English, both teaching and learning, her being confident, and her long-time experience as an English teacher make her classes the most communicative one among the classes teachers in this study have. However, she does not have in-class group/paired discussions; thus, she would only change her teaching for group/paired discussion tasks, not for any other tasks proposed in this study.
Teaching experience and test-taking experience  Five of the teachers had been teaching TOEFL preparation courses for no more than five years. Two of them are more experienced (around 8 years of teaching TOEFL preparation courses). It is intuitive to assume that with more experience in teaching, teachers would be more likely to search for a way to improve students’ speaking abilities than simply teaching to the speaking tasks. But, Amai has started trying to help students to improve their speaking, as previously mentioned and she has been a TOEFL teacher for only two years. Yet, she did say, “Probably four years (being an English preparation teacher), I guess. But actually, I started teaching, I worked part-time before graduation.” Thus, she might have had an idea about what influenced students’ language abilities a while ago. Nonetheless, in the cases of this study, more experienced teachers (such as Hemi and Cardinal) shared their ideas about what is affecting students’ speaking abilities and how they were dealing with some of these factors more than the less experienced teachers.

For example:

“It is my responsibility to make them realize that getting satisfactory scores on standardized tests should not be their ultimate goal” (Cardinal).

“I tried with some students, first I ask them to write down their ideas in Chinese with details, then I would want them to translate those ideas into English. I would pay attention the word choice, grammar” (Hemi).

“I want them to read more articles and learn how those writers report their findings. Then, I would explain the differences I know between written and spoken English, they should be paying attention to” (Hemi).

Instruments
The main instrument for data collection was a set of interview questions adapted from Wall, & Horák, (2006) and some hypothetical questions about how teachers might adapt their materials and instruction if TOEFL adopted certain interactive speaking tasks.

The interview questions (see Appendix A) focus on: (1) teaching method (activities, materials, assessment of students' abilities), and (2) teachers’ opinions on six speaking tasks (what they think about a task and how they would prepare their students for such a task), which are either being used by some high stake tests or are just conceptual tasks. These task types were selected to prompt teachers' thoughts about the current speaking tasks and on how they would adapt their teaching to prepare students for new speaking tasks.

The six speaking tasks (in Appendix B), which teachers were asked to consider, are a one-on-one live interview, one-on-one online interview, group/paired discussion, group/paired discussion in a virtual environment delivered with avatars, presentation using avatars, and artificial intelligence/computer as an interlocutor.

One-on-one (online) oral interview. An examiner asks a student a set of predetermined questions with possible follow questions depending on the students’ responses.

Group/paired discussion. Two or more students are given a topic and amount of time to discuss a topic. Test administrators do not normally engage in the discussion (Ockey, 2003).

Virtual environments (VEs). VEs allow test takers to communicate with other test takers in real time through voice activated avatars. When a student speaks at a remote location, others can see the avatar on their computer screen and hear the actual voice of the other speaker (Ockey, Gu, & Kheener, 2017).
Computers as speaking partners (AI). A computer serves as a discussion partner for a test taker. The AI asks the student questions and “attempts” to discuss a topic with the test taker. State of the art computer systems can only have about four appropriate turns with a test taker (Ockey, 2017)

A short video clip was sent to each participant to illustrate the idea of a computer as interlocutor. The video contained a map direction task in which the computer would capture keywords in test takers' words and make a new prompt like "did you say go north?").

The tasks were chosen for the following reasons: (1) they are not currently a part of TOEFL iBT; therefore they could prompt teachers to think about changing their teaching if there was any change in the TOEFL iBT, (2) these tasks aim to assess interactional competence, which Fulcher (2003) mentions "is often defined in terms of how speakers structure speech, its sequential organisation and turn-taking rules, sometimes including strategies." (p. 44) Ockey (in press) points out that interactive tasks "aim to assess a test takers' ability to interact with other interlocutors, which would imply both production and comprehension.", (3) the asynchronous feature of the current TOEFL iBT speaking tasks do not effectively assess interactional competence, in part, because interactional competence is not even considered in the current rating scale (Ockey et al., 2015), (4) the avatars tasks and computer-as-interlocutor tasks are part of this writer's notion of what is computer-mediated speaking tasks and could conceivably be used to assess interactive speaking ability in remote contexts. One example of computer as an interlocutor is the Computerized Oral English Test of the National Matriculation English Test in Guangdong, China (speaking section is for students who want to major in English or trade). In part two (role play), a computer plays a role of “interacting” with test takers (Cheng & Curtis, 2010, p. 234-247). Although in this example, computers, based on their description, are not yet
functioning as a human-like interlocutor, it is important that test developers are working on incorprating technologies into speaking tests.

**Procedures**

Interviews were conducted through online instant messaging applications, Tencent QQ (an instant messaging software that can be used on PCs and cellphones) and WeChat (an instant messaging application on cellphone) if a teacher said they have time for an online interview. All interviews were audio-recorded and transcribed. If a teacher said they did not have time for an interview, the same list of interview questions used in the online interviews was emailed to them as an open-ended questionnaire. Teachers were asked to complete the questionnaires and return them by email.

**Coding Nodes (in NVivo)**

All interview transcription and questionnaires were coded manually in NVivo according to the following nodes. The nodes are (in an alphabetic order): Assessment (includes a child node: beliefs about TOEFL), classroom description (includes child nodes: activities, aim, material, teaching philosophy), non-test prep class (includes the same child nodes with classroom description), speaking task types (includes AI interlocuter, group or paired discussion, one on one interview, virtual environment), student, test-taking experience, and test-taking technique. Classroom description is intended to address RQ 1 and Speaking task types is meant to address RQ 2. The other nodes were categorized to illustrate the teachers' motives behind the choices they made in their classes.

**Results and Discussion**

**Classroom Description**
This node is defined as any description about activities in class and homework (activity), the aim of the class (aim), materials used in a class (material), and the beliefs a teacher has about how to teach speaking and what to teach in speaking classes (teaching philosophy). Teaching philosophy is included because, for example, sometimes teachers are both describing how they spend class time and showing why they choose to spend class time in the way they do. An example of a classroom description node is as follows:

“I usually try to strike a balance between improving their scores and improving their general English.” (Cardinal)

As Cardinal says, she spends time on both improving students’ speaking abilities and on TOEFL speaking tasks. Therefore, teaching philosophy is included in classroom description rather than in a stand-alone category for teachers’ motivation.

**Speaking Task Types**

In the interview questions, there are six questions about speaking tasks, which are live one-on-one interview, online one-on-one interview, group/paired discussion, group/paired discussion using avatars, presentation using avatars, and computer as an interlocutor. However, most of the teachers do not think there are many differences between a live one-on-one interview and an online interview, and they do not mention differences between a group/paired discussion using avatars and presentation using avatars, which will be interpreted in the discussion section. Therefore, speaking task types has four child nodes: AI interlocuter, group or paired discussion, one-on-one interview, and virtual environment. They include descriptions of teachers’ opinion on the six speaking tasks and what they would do to prepare their students for the tasks.

For example,
“Well, for the VR I think it, the good thing about it is, it helps students to relax as
compared to a real-life presentation-- and for preparations, I think the first step is to let them get
used to operating the VR machines or the glasses and the controllers.” (Wind)

“I think the main points are to show what it is like, to make students experience it
first. Only in such an environment, students know how to perform themselves and display
themselves. What followed will be how to respond to the examiners' questions. And the rest is
similar to the preparation to the interview with the live examiners.” (Wjl)

Test-prep Institutions/Companies

In this study, the information about private high schools is from teachers in test prep
institutions and companies, who learned the information from their students. Therefore, there is
no separate section for private schools. For the test prep institutions and companies, teaching
speaking is sometimes test-centered, or in a communicative way, or somewhere in between. It
depends on an institution or company's policy on teaching, students' English proficiency level,
and teachers' teaching philosophy to decide whether test tasks or communicative teaching weighs
more. Five of the six TOEFL teachers said that their classes are to help their students pass
TOEFL.

“Well, I hate to say it but yeah, it's to help them improve their score.” (Amai)

"Well, to be honest, I think well, the purpose of my class is to help them work on the
best of things for passing TOEFL because I think most of the students come to me for TOEFL
training.” (Ly)

"But I think, in my class, I would teach them both (passing and improve abilities) and
actually I would even focus more on the actual ability, how to improve their vocabulary, how to-
- and then, listening skills too -- how to understand the conversations and lectures better.”
(Wind)

“I think it depends. The students have more time to prepare, they have like half a year to prepare for this test. Then I'll give him or her a very thorough plan for improving his or her language ability, as well as his skills and strategies for taking the test. But, some students they came to our training center, they all have like one or two weeks before the test. All they want us to do is to have them summarize the skills that they need.” (Zls)

“I think. For first (work on the best things for passing TOEFL) and second (to improve the students’ general English) and third (to prepare students for working in an academic environment), they could improve in certain degrees.” (Ivy)

Ivy mentioned that she believes although her institution wants their students to pass TOEFL (that is getting a score higher than the minimum required score for any foreign college), she also wants her students to be able to speak English fluently. Teacher Wind believes that "improving your English ability actually equals to passing the test. Because I think TOEFL test is really well designed”. The other TOEFL teacher (Cardinal) is the only one who does not regard passing TOEFL as the aim of her classes. She has a communicative approach to teaching, and she develops a system to keep her students motivated to complete assignments. Her case will be discussed in the following section.

It would seem that besides independent teachers, the test prep institutions and companies in this study prefer test centered classes, which is not surprising considering they need to make a profit to satisfy their stakeholders. However, the teachers have concerns about actually improving students’ speaking abilities if they have a chance to do so (for example, if there were
enough time before a student takes the TOEFL test). Moreover, the beliefs teachers have about the TOEFL test also influence their opinion on the aim of their classes.

To address the concerns some teachers have about not improving students’ language abilities, Hanny mentioned that her institution is trying to combine general English speaking classes, which are for anyone who wants to speak English fluently either in daily conversations or in academic environments, with test-oriented speaking classes. She said: "teachers from both classes (test-oriented and non-test-prep) scheduled meetings to talk about different ways of teaching speaking and to present mock teaching sessions". Her team, herself included, and her supervisors have been preparing their speaking teachers by learning TKT (teaching knowledge test) test courses and passing TKT. They also have a teacher training course that uses materials from CELTA (Certificate in Teaching English to Speakers of Other Languages) courses. Her comments on some of the mock teaching sessions presented by test prep speaking teachers are as follows: "I can see that some of them were trying to use some teaching methods I've talked about (for example, controlled practice, TTT-test teach test, etc.). But, their activities are not, what we call well-designed, no offense, but that's what I saw.". So far, this is just "an initial stage". There hasn’t been any final decision on how to combine two kinds of speaking classes, yet.

The next two sections will address the two research questions proposed in this study. The first part is a description of the VIP-TOEFL-speaking class in a test prep institution/company and from an independent teacher. The second part is about what the six teachers think about the six speaking tasks in this study and how they would prepare their students for those tasks if TOEFL would use them to assess communicative competence.

**What is the VIP-TOEFL-Speaking Class Like?**
Before stepping into individual cases, some of the common themes are summarized below.

**Materials** All of the TOEFL speakers said they would use Official Guide (OG) and TOEFL practice online (TPO) for beginner and intermediate student. Some students have taken TOEFL several times and are quite familiar with OG and TPO. LY uses his collection of TOEFL speaking topics. Because some of LY’s students would report to him what topic was in a test they had just taken, he has a chance to compile his own collection of TOEFL speaking topics. Others mentioned they would turn to test prep materials published by other companies, like Pearson, New Oriental, or any mock tests they have. If a student is good at speaking or has done all the practices in OG or TPO, as in teacher Cardinal and teacher Hemi’s case discussed below, any topics in life could be prepared by the teacher to use in class. Textbooks or test-prep materials are not the only source for teachers to prepare a class.

Several applications on cellphones are used to help students memorize a list of high-frequency words in previously collected TOEFL tests. Most of these applications used word lists which were selected based on frequency searches in TOEFL texts.

TED talks are used both by the test prep teachers and non-test-prep teachers as practice outside of class for students to summarize and retell. Cardinal, who teaches TOEFL prep course online, uses Tencent QQ as the main medium for course instruction. She also mentioned using clips of TV shows in English as part of the course materials or practice materials.

**Activities** The activities mentioned by the teachers in this study are role play, presentation, monologue (or, in one of the teachers' words "talking to yourself on a certain topic") and note taking practice.
For students who are not familiar with TOEFL or are not proficient with English, there are not many activities in their one-on-one classes, which are more like a lecture on TOEFL speaking tasks and ways to prepare for those speaking tasks.

**Students** The students in VIP classes are from public high schools and private high schools. Their English abilities vary, in terms of TOEFL speaking scores, they range from 5 to 27 on TOEFL iBT speaking’s 30-point scale. However, it is worth noticing that some high-ranking high schools in China have a new department, in which they focus on helping the students there to get accepted by a foreign university (often the students there do not take the Chinese college entrance exam). To achieve that aim, these departments hire native speaker teachers and have adopted several foreign syllabi (A-level courses, part of some K-12 curricula etc.), which private high schools have been doing for some time. Students in such departments or in private high schools have more opportunities to use English both in class and outside class than a student in a regular public high school.

**The VIP-TOEFL-speaking class**

The name of this type of class suggests that it can help students to get a desirable speaking score. Parents who want to send their children to study abroad want nothing else but a passing TOEFL score (passing the “gateway”) when their children finish a VIP course. Wind best captures this expectation with the comment:

“because we are, you know, the institution has to make money and we are responsible. The students come to us because they want to get higher TOEFL score and in order to get into a college so, we cannot tell them like, after ten or twelve classes, we cannot tell
them, ‘Well, your ability has improved but sadly it did not show on the score’ No, it's not acceptable. So, you know, the parents will ask for a refund if it happened.” (Wind)

In addition, as discussed above, the English proficiency level varies among students. The teachers choose Chinese over English as the instruction medium for the sake of comprehension. Amai said: “But the thing is, sometimes students can get really confused when you use too much English, and yes, sometimes they would just ask you, ‘Could you please speak Chinese because we don't understand you and it's not clear?’” It is to serve the expectation of passing TOEFL that the class has a test-centered approach and is mostly done in Chinese.

VIP classes also mean that it is one-on-one for the teacher and student. The standard class time is two hours per session. It would be challenging for the teacher to come up with topics to practice if a student has taken TOEFL several times and completed all the mock tests they could find. Hemi told an interesting story about one student who had been practicing speaking with her for over one year, and they had gone through every TOEFL speaking topic they had. Thus, they went on talking about music, movie etc. She jokes about it: “In a way, we were preparing for the independent tasks, you know.”

To draw a picture of a two-hour-session, it starts with short talks, which could be in English depending on the students' level, then it moves on to one of the speaking tasks in TOEFL, which would be discussed in Chinese most of the time. After the discussion, the student would have time to give an answer to the speaking task and the teacher would give feedback on the student's pronunciation, word choices, organization etc. When they reach an agreement on the feedback, they continue their speaking practice. For students who have limited time before they take the next test, teachers would spend more time on test-taking strategies, such as speech templates, time management, how to give an answer without fully understanding listening or
reading and so on. For students with a low English proficiency level, the class would be more of a lecture.

Some descriptions provided by the teachers include the following:

“So, I would teach them how to give like a versatile answer, something like that, that even if they don't fully understand, they kind of still give an answer.” (Amai)

“But most of the time, we speak a lot. Actually, my students speak a lot and I'm listening to a lot... I will point out the place or the words and they really need to correct or revise.” (Ivy)

“Well, actually in my case, I mostly talk--I mean speak Chinese because most of my students are not very good at English.” (Ly)

“When they came they're already like, twenty-six so, all you have to do is find their little problems and help them fix it.” (Wind)

**Thinking in English**

Amai mentioned that one of the difficulties her students have is the L1 influence: “when someone who is not good at speaking when they speak, they actually think in Chinese and then, translate it into English, and that caused a lot of mistakes”. She tried to help her students by first teaching them set phrases and a phrasal verb. Her idea of thinking in English instead of in Chinese is similar to a non-test prep teacher’s (Yuki) beliefs about monologue (“That is because I do believe that preparing monologue is a good way to practice English even if you are talking to yourself. It can help you to think in English”). These two teachers are looking at students’ difficulties from a different perspective, what they believe to be the factors that they thought are
important for students to improve their speaking abilities. Hemi and Cardinal have more critical thoughts on how to improve students' speaking abilities: "I think many of my students do not read or listen enough, so they don't have enough words to express ideas in English. Or they don't know how to express an idea appropriately in English, let alone when they have limited time to think about a topic" (Hemi). What I interpreted from their comments is that some teachers are searching for more than tasks to teach in a class. Cardinal's case below would be a good example of such a search.

**Speaking Tasks and Construct Awareness**

**One-on-one interview**

Many teachers in this study would take the questions on one-on-one interview as a comparison between TOEFL and IELTS speaking tasks.

“Okay, if TOEFL uses one-on-one examiner, so can I understand the question as what IELTS students do to prepare for their IELTS speaking exam” (Wind).

“So, it's quite hard. So that's why I think it's-- TOEFL tasks is more difficult than IELTS. Because students who take IELTS, they would speak in front of the real examiner” (Ivy).

Some teachers have a clearer understanding of different task types.

“Okay, so TOEFL is so different from IELTS. It is not only about a machine, a computer and a live examiner; it’s also about the test forms; there are integrated test, and independent test, which kind of like an IELTS task that the students would be given a short question and then they, all they had to do is answer the question. But as, you may also know, for integrated test the students had to listen to a conversation or lecture, read a short passage or notice from school and
they have to, to combine this information and to rephrase and retell this information, so it's
totally different” (Wind).

In sum, many of them prefer a one-on-one interview, which they believe, for some
students, interview or talking to another person makes the students less nervous and may give
students more clue about whether the examiner understands them, or whether they are on the
right track.

“Talking with a live examiner is better than talking to a computer” (Cardinal)

When they thought about preparing their students for such tasks, most of them would turn
to IELTS interview questions and use role-play or mock interviews. Again, Cardinal says: “I
would still teach in the same way because my current method also prepares students for
situations like that. I focus on improving pronunciation, and making students feel comfortable
about thinking and communicating in English instead of just memorizing contents and making
recordings.”

**Group/paired discussion**

This is a rare activity in VIP-TOEFL-speaking classes unless a student's English is fluent
enough to have a discussion with his/her teacher. Therefore, some of the teachers said they
would change if this task were in the TOEFL speaking section:

“'I think that's a great idea, especially with larger classes because, in speaking
classes, the students are supposed to speak more but with a big class, it's impossible for every
one of them, for the teacher to give every one of them to speak. So, group discussion is a better
way” (Amai).
“I would organize several students to discuss the topics they encounter in reading/listening materials. For instance, I would ask some students to watch TED talks, documentaries, and TV series, and write/talk about them” (Cardinal),

Others have some concerns about having a group/paired discussion:

“Well, first of all, I don't think it's a good idea because if it's just work in groups they may be influenced by their partners. That sometimes it is unfair for some good students. Because if they cannot understand what their partner is saying, then they cannot communicate well” (Zls).

“If there's this live or pair discussion task well to have my students prepare for it, I would actually have a role play with them. I will play- probably work as their partners so we can have a discussion. And, well, I think--but I think there is some potential problem because some students may not be very confident to speak up and I think my role is to motivate them to join in the discussion to convey their ideas as much as they could” (Ly).

“They won't get better while talking to each other but only get worse because they might just gather mistakes the others are making and they couldn't judge if it was right or wrong so usually I--well, usually if, if they're not requirements I would not let my students do group discussions or it depends if, if I could tell like 2 or 3 of the students in the class there are, they had the ability, they're able to use the language fluently I will let them do the discussions” (Wind).

The rest of the teachers reported they are aware of what this task can do, but it depends on the students’ level mostly and the institution’s policy.

AI/Computer as an Interlocutor
An All of the participants understand the concept of computer interlocutor, which, in a perfect situation, should be functioning like/as a human interlocutor. However, it is still a long way into the future before we could have a perfect situation for computers to be an interlocutor. Ivy said: "I think it's quite weird. Because a kind of machine just repeat, can you repeat this again? Can you repeat this again? Don't you think it's really weird? I don't think this kind of thing will happen in the real life. Maybe if we match this kind of situation in real life, I may take a picture and then send a picture to another person and say, okay do you mean this?"

What she meant was that computers do not have the ability to maintain a conversation like humans do. Thus, we should not use it as part of an important test. If we were to use a perfect version of such a computer interlocutor, the teachers would not have any concern and would not change their teaching. Because they said they have addressed how to make conversations in their classes. It is just with different students, the proportion of teaching making conversations varies.

“So, if there are more complicated versions I have to know what situations might be given, so I can teach the students how to solve them” (Wind).

“I think if you'll use an AI to communicate with test takers you need to make sure that the AI can function as a real person in a test. When that happens, it won't make any differences to teach, just because the partner is an AI but is not a real person.” (Zls).

**Virtual Environments and Avatars**

Because it is in a game-like environment, the first thing teachers mentioned was that it would be interesting for students or, at least, students who are interested in games. Then, they would report more concerns, like they may not have enough equipment for students to get
familiar with the virtual environment: some important information might be lost if using avatars (eye contact, facial expressions, etc.). Last, some of them said that whatever a task may be (a presentation, group discussion in a virtual environment), some of the English language abilities remain the same.

“Oh, that's cool. That's really cool. Because if it's changed that way, I think like lots of students may probably accept this idea or embrace this change because it feels like, well, they're in games maybe. It's not like a traditional test or anything because it's more like a game so I think it will be welcomed by lots of students” (Ly).

“Yeah, that I think they're really helpful and interesting. Attractive, yeah” (Ivy).

“I think it is partly interesting. But if you cannot do it, then just try to make sure that it is fair for everybody” (ZIs).

Wind expresses one strong concern about virtual environments:

“I think it's important to see his face to look him, to look him eye to eye and you can get a lot of information from their facial expressions and the tones they're speaking and how they react, how they look, I think that's quite important so, I would say I would save my opinions about this VR thing. I would prefer a person to person or since I'm, maybe because I'm outgoing, I would prefer a man-to-man conversation rather than VR or internet interviewing. Yes, and also for teaching, I think, to sit in a classroom is way better than using Facetime or VR.”

Conclusion
Some teachers teach to the test either because their institutions require them to do so (the overarching Chinese test context is that a test is a “gateway” to better opportunities at work or study), or because they believe that the test is well designed and teaching to the test is good enough. Therefore, for them, if the test were to change its speaking tasks, they would start teaching to the new tasks. Some teachers would not change their teaching, regardless of what the task might be, for reasons like: (1) they want to improve students' speaking abilities and they don’t think teaching to the test will help to do that, and (2) despite the task types, the language abilities students need to pass the test or get a high score are the same. This split situation is similar to what Cheng (2005) found in her study of HKCEE (Hong Kong Certificate of Education Examination), in which some teachers were skeptical about changes in HKCEE and some welcomed the changes.

Figure 1

If we consider the relationship between teaching, test task, and test construct (as shown in Figure 1), there could be an explanation for this split situation. Traditionally, test developers focus more on the alignment of test tasks and test constructs. Teachers focus more on what tasks are used in a test. When test developers investigate what the impact of their test is, they usually look at teaching, which, in many cases, puts a huge amount of weight on test tasks. In turn, test developers would categorize such teaching to test tasks as negative washback, because teaching
to test tasks would omit some language abilities that were in the test construct but not well represented by test tasks. Hence, teaching to tests tasks could be bad language teaching, which could be viewed as negative washback. Ideally, if a test’s tasks are in perfect alignment with its construct, teaching to test tasks shouldn’t be a negative washback because students could improve their abilities by practicing with these test tasks. This alignment requires test developers’ great effort to achieve such an ideal situation. Before that can happen, it would be possible to counter the effect that is brought by misalignment between test tasks and test constructs by having teachers better understand test constructs. By raising teachers’ awareness or understanding of test constructs, that is the ability the test aims to measure, more teachers might be better equipped to check if their teaching methods can improve students’ language abilities. They might also be able to determine if a test task is well aligned with a test construct and what can be done in a classroom setting to use this test task to improve students’ language abilities than to just get students familiar with this test task.

Some teachers in this study have begun to search for the abilities to pass TOEFL. Their decisions are based on either their experiences with many students or on their intuitions about learning a language. These experiences and intuition may not lead to teaching that is well aligned with test constructs. Instead, they may focus more on testing strategies, such as how to pass multiple-choice tests. Therefore, their teaching may still not match help students develop the language skills that test developers had targeted. By teaching teachers about a test’s construct, or at least raise their awareness of test constructs, teachers may be able to focus on the question: "What is the best way to teach/improve the abilities this test is measuring?" Moreover, if there were to be any change in test constructs, teachers would not be affected by changes in test tasks if their teaching is already aligned with the test construct. (In the case of Cardinal, she believes
her teaching is communicative and the tasks proposed in this study are communicative; thus, she did not report the need to change most parts of her class). It would be most desirable that teaching, test tasks, and test constructs are in well alignment, as Bachman and Palmer (1996) said: “the characteristics of the test and test tasks correspond more closely to the characteristics of the instructional program.” (p. 33)

The overall context for TOEFL in China is that TOEFL is a ticket to studying abroad. Passing it means getting the ticket. Therefore, passing it dictates that most of the TOEFL-oriented classes in China are heavily test task centered. Some teachers from test-prep institutions and some independent teachers view passing TOEFL and improving English abilities as equal. It suggests that raising teachers’ awareness of test constructs may be a way to make test preparation more effective for learning a second language.

**Limitations**

To start with, this is just a small-scale case study. A different sample from an institution that has a different teaching philosophy may yield contrasting results to the results of this study. For some teachers, they only received the interview questions as questionnaires. The interview questions are mostly open-ended question (not even yes-no questions), making it hard for the participants to have the patience to answer every question in detail. For example, some teacher just answered: "I will not change anything" to the question: "How would you prepare your students for such a task?". Future research may have a larger sample size and consider a questionnaire that may prompt more details from participants.

Despite these limitations, this study suggests that test-prep speaking teachers choose their teaching style and material based on their institutions/companies’ policies, their students’ level,
and what they believe to be the best way of learning English. Thus, teachers perceive changes in test tasks differently, from welcoming it to not caring about task types. Some teachers said they would change their teaching if some of the interactional tasks were used in TOEFL, while others said they would not change anything. It might better prepare test-prep teachers for language teaching than test coaching to raise test-prep teachers’ awareness or understanding of test constructs.

Summary

This study seeks to provide some information about the motivations of VIP TOEFL speaking teachers. The cases investigated here suggest that VIP TOEFL speaking teachers do not have much freedom over what they do in their classes. The primary objective of their course is to help students get a higher TOEFL score. Teachers with less teaching experience tend to teach to the test tasks, while more experienced VIP teachers are not satisfied with their choices of teaching materials and methods. More experienced VIP teachers are searching for ways to help students to improve English language abilities, such as creating an immersive English environment for their students, using authentic English texts/videos/audios in their classes, and so on.

This study suggests that in test-prep teacher training, it would be beneficial for both teachers and students to teach teachers knowledge of test construct. Because, for new teachers it would raise/strengthen their awareness of side effects of teaching to test tasks and for experienced teachers it would save them time to search for the abilities a test is measuring. More experienced teachers would be able to identify what underlying ability might be missing from a certain type of test task and consider whether it would be worth it or not to help their students
with improving the missing ability. It would provide teachers more assurances when they have to choose whether to stick to teaching to test tasks or to focus on language abilities.

That being said, further studies with a larger sample are needed to reach a generalizable conclusion on teachers’ motives. It is also necessary that future researchers summarize the methods used in a VIP speaking course and motives or rationale behind each method. So that it would be clearer if teaching is aligned with current test construct or future would be test constructs.

In spite of its limitations, this study provides insights into changes teachers might make into their practices if TOEFL iBT uses more interactive speaking assessments. Less experienced teachers would use more interactive tasks, imitating the tasks themselves. On the other hand, more experienced teachers would not change their teaching practice very much because they believe that their approach already teaches students how to interact.
A CASE STUDY OF VIP TOEFL SPEAKING TEACHERS  37

Reference


Appendix A

Interview questions

Research identification number:

Male/Female:

School:

Date of interview:

Teaching experience

1. How many years of experience as a teacher of English do you have?

2. How long have you been teaching TOEFL preparation classes?

3. Have you ever taught preparation classes for TOEFL CBT or paper-based TOEFL? (not the iBT TOEFL)
   If so, what you think about the differences between these different versions of TOEFL?

4. Have you taken the TOEFL exam yourself?
   If so, do you think this experience influences how you teach test preparation classes?

The TOEFL preparation classes

5. What do you see as the primary aim of the class?
   a. work on the best things for passing TOEFL
   b. to improve the students’ general English
   c. to prepare students for working in an academic environment.

6. Please explain why you selected one of the answers to question 7.
Students

7. How do you assess your students’ English abilities before they can join your TOEFL iBT preparation class?

8. In terms of TOEFL scores, what level are your students at roughly? (please provide a general range of TOEFL scores for your students)

Methodology

9. Which language (English or Chinese) do you use most in class?

10. On what do you base your decision about which language (English or Chinese) to use?

11. What kinds of books, or any other resources do you use for the speaking classes? Why?

12. What kinds of activities do you usually do for speaking class?

Assessment

13. How do you assess your students’ speaking abilities? Please explain what kinds of tests and tasks you use. For example, do you use questions from TOEFL or start a conversation with your students?

Test taking techniques

14. Do you cover speaking test taking techniques in your classes? If so, can you tell me about how you use these techniques.

15. As far as you know, do students do any study outside of class to help prepare for the TOEFL speaking section? If so, what do they do?

16. How do you encourage them to study speaking outside class? (e.g., by giving tips or ideas for what to do)
Now I would like to show you six speaking tasks that could be used as assessment for speaking abilities. After each task we would like to hear your opinion about the task and what you would do to prepare your students for such a task. These speaking tasks may have been used for tests in the past or could be used for future testing purposes. If you have any questions about the tasks, feel free to ask me at any time.

17. If the TOEFL iBT used a one-on-one oral interview with a live examiner, what would you do to prepare your students for this task? (In terms of the way of teaching, teaching material, teaching activities, tests used in class, and the amount of time spent on speaking). If you would not change anything, please let me know.

18. If the TOEFL used a one-on-one oral interview through Facetime or Skype, what would you do to prepare your students for this task? (In terms of the way of teaching, teaching material, teaching activities, tests used in class, and the amount of time spent on speaking.). If you would not change anything, please let me know.

19. If the TOEFL used a live paired/group discussion task on an academic topic, what would you do to prepare your students for this task? (In terms of the way of teaching, teaching material, teaching activities, tests used in class, and the amount of time spent on speaking.). If you would not change anything, please let me know.

20. If the TOEFL used a paired/group discussion task using avatars in a virtual classroom, what would you do to prepare your students for this task? (In terms of the way of
teaching, teaching material, teaching activities, tests used in class, and the amount of time spent on speaking.). If you would not change anything, please let me know.

21. If the TOEFL used a presentation task using avatars in a virtual classroom, where examiners are present through avatars, what would you do to prepare your students for this task? (In terms of the way of teaching, teaching material, teaching activities, tests used in class, and the amount of time spent on speaking.). If you would not change anything, please let me know.

22. If the TOEFL used a map direction task, in which a computer is your interlocuter (it would ask questions/offer prompts like: “would you repeat that?” “You said go south”), what would you do to prepare your students for this task? (In terms of the way of teaching, teaching material, teaching activities, tests used in class, and the amount of time spent on speaking.). If you would not change anything, please let me know.

And one final question:

23. Do you have any more comments or test-taking stories about TOEFL you would like to share with us?
Appendix B

Descriptions or examples provided to interviewees

One-on-one oral interview

No specific example were given to the interviewees for none of them asked for clarification or explanation of what is one-on-one interview.

Paired discussion

Example: (provided to all interviewees)


Now, you are going to listen to two speakers talking about pros and cons about social networking.

After listening to the female speaker, I would like (Student A) to tell us what the opinion of the speaker is about social networking, and why the speaker has this opinion.

Then, after listening to the male speaker, I would like (Student B) to tell us what the opinion of the speaker about social networking, and why this speaker has this opinion.

Please retell the talk in your own words. Each of you will have 1 minute for retelling the speaker’s point of view. You will listen to the recording only once, and you cannot take notes. You need to listen carefully to the both speakers because you will discuss both speakers’ opinions later.

Now, please listen carefully to the female speaker.

[I spend two to three hours per day using social networking sites. Not many things in the world are
free, but social networking sites are. These sites have given me a chance to keep in touch with family and friends and helped me professionally. My research productivity has increased since I started using Academia, which is a social platform for academics. Many scholars, including me, post their research papers on this site, so that anyone can access their research. This has also helped me to build social networks with others who have research interests similar to my own.]

Now Student A, you have 1 minute to retell the female speaker’s point of view.

[Mina: Retelling the female speaker’s talk for 1 minute]

Now, please listen carefully to the male speaker.

[I never use social networking sites. I read a report that says as the number of Facebook friends a person has increased, the stress the person gets when they use the website also increases. Also, not surprisingly, people tend to spend far less time talking with their friends and families in face-to-face conversations because of the increased online interactions. Maybe most importantly, social networking sites are a huge waste of time and money. I can tell you that my boys are getting poor grades in school because they are always facebooking friends when they should be doing homework.]

Now (Student B), you have 1 minute to retell the male speaker’s point of view.

[Hiroshi: Retelling the female speaker’s talk for 1 minute]

Now please defend the speaker’s position that you just retold, and try to convince your partner of the speaker’s views on the controversy. Even if you do not agree with the speaker, you need to support the speaker’s position.
Please try to have short exchanges and share time equally with your partner. You have 4 minutes for this discussion. Either of you can begin.

[Student A and Student B talking to each other: 4 Minutes of Discussion]

**Avatars**

A video clip of an online Virtual Reality game called VR Chat was sent to interviewees to help them understand the idea of avatar and virtual environment. Video (only 1 minute of the whole video were used) is from [https://pan.baidu.com/s/1-dz03wjz5mrhNd7M4T0BPg](https://pan.baidu.com/s/1-dz03wjz5mrhNd7M4T0BPg)

**AI as interlocutor: a map direction task**

A video clip of a map direction task (from [https://www.youtube.com/watch?v=MzL-B9pVbOE](https://www.youtube.com/watch?v=MzL-B9pVbOE)) were used to illustrate the idea of using computers as interlocutor.

A few points the interviewees should know:

1. What we want you to focus on is the conversation the two speakers had, not the direction task
2. Pay attention to the “male” speaker in the video
3. We could use computers to “talk with” test takers. For now, technology is not ideal.