Chapter 4
To Whom Do I Write? Chinese EFL Test-Takers' Conceptualisation and Construction of Their Audience in the Aptis Writing Test

Ying Chen and Xiaoxian Guan

Abstract The present study explored whether, and if so how, Chinese test-takers portrayed the audience and catered to the needs of two different audiences specified in the Aptis General Writing Task 4 employing think-aloud protocols (TAPs), questionnaire survey, and semi-structured interviews. Sixty six test-takers participated in the present study, including two employees, at B2 and C level respectively, and two groups of students with different levels of English proficiency, 32 high-proficient (at B2 or C level) and 32 low-proficient (at B1 level). Six test-takers, consisting of two employees and four students, were invited to take part in the think-aloud session. Recordings of test-takers' TAPs and interviews were transcribed, then double coded and analysed on the basis of Berkenkotter's (Coll Compos Commun 32:388–399, 1981) audience awareness coding scheme. Results from both quantitative and qualitative analyses indicated that explicitly specifying audiences in the writing tasks successfully weakened the understood testing context given that test-takers wrote to the audience specified in the task rather than to rater(s) for the mere purpose of getting a high score. But there were marked differences in audience conceptualisation and adaptation strategies between high-proficient and low-proficient writers as well as between the two audience conditions. While shedding light on the construct validity of the Aptis writing component, the findings of this study have important implications for EFL writing teaching, learning and assessment.

Keywords Audience awareness · Audience conceptualisation · Audience-related strategies · Contextualised writing prompts

Y. Chen (✉)
Ocean University of China, Qingdao, China
e-mail: cy@ouc.edu.cn

X. Guan
East China Normal University, Shanghai, China
e-mail: xxguan@english.ecnu.edu.cn

4.1   Context of the Study

In 2018, Council of Europe updated the 2001 scale by adding a number of new descriptors, one of which is the inclusion of formal correspondence (Council of Europe, 2018). The updated scale thus consists of both personal and formal correspondences which address two different registers, informal and formal. The use of writing tasks addressing different registers, however, is not widespread in large-scale standardised tests. One exception is the Aptis General Writing Task 4 designed by the British Council, which assesses test-takers' performance in both informal and formal conditions with the inclusion of register as one important aspect in writing task design and rating rubric (O'Sullivan & Dunlea, 2015). Aptis General, launched in August 2012, is the first variant within the Aptis Testing System which aims to offer test users flexible English language assessment options (Dunlea, 2018; O'Sullivan, 2012). The theoretical model behind Aptis General is the socio-cognitive model advanced by O'Sullivan (2011), O'Sullivan and Weir (2011), and Weir (2005) which pays special attention to the interaction between context and construct, and those contextual parameters such as topic, genre and the intended audience are explicitly specified to provide contextual information to test-takers.

Aptis General Writing, built around a series of theme-related activities, consists of four tasks which range from very basic form filling to quite complex email messages. Task 4 requires test-takers to write two emails in response to a short letter/notice: an informal email to a friend or close family member and a formal email to an unknown reader specified in the prompt. To validate this task, a number of studies, which are freely available on the British Council website (https://www.britishcouncil.org/exam/aptis/research/publications), have been carried out particularly on test development, rater training, refinement of the rating rubric, and cognitive validity. Given that the task requires test-takers to write to two different audiences, a close look at how test-takers conceptualise and construct the two target audiences is necessary. The present study undertakes the project of investigating how Chinese test-takers deal with audience demands by collecting both online think-aloud data and off-line questionnaire and interview data. The findings of this study could add to the theoretical validity of the Aptis General Writing Task 4, and enhance our understanding about the provision of contextual information in writing task design.

4.2   Literature Review

4.2.1   Audience and Audience Awareness: Definitions, Debates, and Disagreements

Researchers in the field of rhetoric and composition have long recognised that audience is imperative for successful communication. Audience, however, is an elusive concept, projecting different images to different writers (Porter, 1996). Broadly speaking, the images of audience can be grouped into two categories

(Park, 1982). One category is the actual people external to a text whom the writer must accommodate and the other is the audience implied in the text, or "audience-addressed" and "audience-invoked" respectively in Ede and Lunsford's (1984) term. Worse still, exact terminology may differ. For example, researchers have differentiated 'reader' from 'audience'. Due to the limited space, the present study will not distinguish 'reader' and 'audience' and will use the widely-accepted term 'audience' here. Those who are interested in the difference between 'reader' and audience' can refer to Park (1982) and Ede and Lunsford (1984). Given the fact that audience is ill-defined, the concept of audience awareness is unavoidably a slippery term, indicating various conceptualisations and interpretations within different rhetorical and compositional epistemologies. Relevant literature indicates that conceptualisations of audience and audience awareness have shifted over time, mainly under rhetorical, cognitive, and sociocultural frameworks (Magnifico, 2010).

Although researchers approach audience and audience awareness from diverse perspectives and with a long tradition of debates and disagreements, these studies do not suggest contradictions or controversies, but rather complement each other in providing a deeper and more comprehensive understanding of the complexities of audience and audience awareness (Kirsch & Roen, 1990). For example, Willey (1990) found that some writers addressed a real audience but invoked a fictional audience at various stages of the writing processes. In their more recent publication, Lunsford and Ede (2009) emphasised that "understanding the complexity of the writing process, audience awareness, and participation calls for more specific grounded, and nuanced analysis than the binary of addressed and invoked audiences can provide" (p. 56). As such, with an attempt to enrich the current scholarship of audience research, the present study explores how Chinese EFL learners conceptualise and construct audience in a standardised writing assessment context by adopting an open and multiple perspective of audience. The working definition of audience is thus established as the person or group to whom writers seek to convey their message through a written text, whether this being the writer him/herself, the abstract or fictional reader imagined by the writer, the people specified in the writing prompts whom writers are asked to write to, the teachers who will read the written texts and give writers feedback, or the raters who will evaluate writers' performance by awarding a score. Audience awareness refers to writers' understanding of the audience's characteristics, expectations and beliefs, and adjusting their message accordingly so as to effectively communicate with the target audience.

4.2.2 Previous Empirical Studies on Audience Awareness

Recognising the importance of context on communication and writing and also prompted by the increasing demand for authentic tasks in language tests, in late 1970s and early 1980s, an increasing number of scholars and practitioners (e.g., Crowhurst & Piche, 1979; Odell, 1981) strongly recommended providing rich contextual features in writing tasks. Empirical studies on the impact of providing

these features, including specifying the target audiences, have thus been undertaken but yielding conflicting results. Some researchers found that there would be a change in the syntactic complexity when students write for different audiences (Crowhurst & Piche, 1979; Smith & Swan, 1978) and that skilled writers distinguish themselves from their less skilled counterparts in their ability to recognise and address the demands of different audiences (e.g., Ransdell & Levy, 1994; Zainuddin & Moore, 2003). Other studies, however, indicated no significant differences in the holistic scores of different audience conditions (e.g., McAndrew, 1982). Still others, such as Brossell (1983), revealed inconsistent evidence of the effects of contextual features in general and specific audience conditions on test takers' written products. Two reasons can help account for these contradictory results (Chen, 2014). On the one hand, definitions of audience remain elusive, hence researchers contextualise writing tasks in different ways. On the other hand, the majority of research adopted a product paradigm and employed different rating scales to evaluate written performance.

Despite the inconsistencies and even contradictions in empirical research, many scholars and teachers regard the ability to address different audiences appropriately as one importance indicator of the development of writing ability (Camp, 2012; Oppenheimer et al., 2017) and audience awareness is taken as a trait of the scoring scales in many writing programs at U.S. universities (Dryer, 2013).

In contrary to the abundance of research in L1 setting, few research has been undertaken to explore how L2 students conceptualise the target audience and how they deal with audience demands throughout the writing process (Cheng, 2005; Wong, 2005). Still scarce is research focusing on audience in L2 writing assessment prompts (Cho & Choi, 2018). To the authors' knowledge, there are only three published research studies investigating, to some extent, the influence of audience specifications on L2 test-takers' writing performance. The study conducted by Porter and O'Sullivan (1999) explored the effect of the age of the intended addressee on the written performance of Japanese university students and found that 'there is clear evidence to support the assertion that awareness of audience is an important factor affecting the scores achieved in these writing tasks' (p. 71). Chen (2014) developed a four-point holistic scale, ranging from 0 to 3, to quantify Chinese test-takers' sense of audience in texts written on tasks with three different audience conditions and found that some proficient students showed strong awareness of audience when the prompt specified the audience, thus making their essay more impressive. In a much recent study, Cho and Choi (2018) examined the effects of audience specification in a prompt on the quality of summary texts by L2 writers in a standardised testing context in America and found that these writers were able to take the needs of audience into consideration when rhetorical constraints were made clear to them. The researchers concluded that drawing ESL test takers' attention to an audience seemed to evoke their rhetorical awareness, leading them to adapt their writing to the needs of a specified audience. In designing a valid large-scale writing test, therefore, importance should be attached to the specification of an audience and a real need of communication, as Cho and Choi (2018) suggested.

It is worth noting that the few studies which have been undertaken in L2 writing assessment settings all investigate test-takers' audience awareness from the product perspective. The process perspective is therefore needed, which will enable us to get a detailed picture of how L2 writers, faced with cultural, rhetorical and linguistic demands simultaneously, cope with different types of audiences for different purposes in a testing context: to communicate with the audience specified in the writing task and, at the same time, to get a high score. Another point noteworthy of the present study is the use of email writing prompts rather than the widely-researched persuasive writing (Midgette et al., 2008). Here we do not mean persuasive writing is not important, but that by exploring test-takers' performance in other genres, such as email writing in this chapter, we hope to add richer findings to the current literature of the audience awareness research, especially on the intricated relationship between the rhetorical context and the assessment context.

4.3 Methodology

4.3.1 Research Questions

The research questions addressed in this study are as follows:

1. How do EFL test-takers conceptualise the target audiences specified in the Aptis Writing Test? What strategies do they adopt in order to meet the needs of the audience?
2. Are there any differences in the above-mentioned aspects depending on the audience condition (informal vs. formal) and EFL test-takers' writing proficiency?

To answer these research questions, this study adopted a within-subjects and between-subjects mixed design, with the audience condition (informal and formal) as the within-subjects factor and test-takers' writing proficiency the between-subjects factor. Moreover, this study avails itself of triangulation by probing into test-takers' writing process. To be specific, the present research employed think-aloud protocols (TAPs), questionnaires, and interviews to find those moments during the writing process where Chinese test-takers consider the needs of an audience.

4.3.2 Participants

The present study recruited test-takers through posters on university campuses and consultations with human resource department of prospective companies and organisations. Due to logistical constraints, this study selected 64 student test-takers, 21 boys and 45 girls, aged 19–21, who had been learning English for 9–12 years, and two employee test-takers, aged 35 and 36 respectively, both using English as part of their working language. Among the 64 students, 32 are low-proficient and 32 high-

proficient, based on their College English Test (Band 4 and Band 6; CET-4 and CET-6 for short) scores, the most well-known national test of English proficiency of non-English major college students in China, and teacher evaluation. For the low-proficient group, roughly at B1 level, their CET-4 score is below 500 or CET-6 score below 450. The high-proficient group consists of non-English majors, at B2 or C level, whose CET-4 score is above 600 or CET-6 score above 550, and English majors recommended as proficient English learners by their teachers. Two employees include one male, at B2 level, working for a corporate company and one female, at C level, working for a government organisation as a Chinese-English translator.

To carry out TAPs, this study selected four students from the two proficiency groups, one boy and one girl from each group, representing a microcosm of the entire student participants in terms of English proficiency. The two low-proficient students, Kun and Chi, majored in Engineering and Arts respectively and at their second year of college by the time of data collection. Their high-proficient counterparts, Hui and Liu, were senior English majors and among top five in their class. The two employees, Wei and Zhang, also participated in the think-aloud session. For all the test-takers, Chinese is their first language.

4.3.3   Writing Prompts

As introduced earlier, Aptis General Writing Task 4 requires test-takers to write two emails to two different audiences (see O'Sullivan & Dunlea, 2015 for the detailed specifications of Task 4). In the present study, test-takers were invited to write to a friend and the customer service team about an online language course. For the first email, the situation is set as follows,

> You have joined an online language course. You've been learning the language for a few months. Your friend is interested in learning the language, too. Write an email to your friend recommending the course and give advice on how to improve it.
>   Write up to 50 words. You have 10 minutes.

And the second prompt goes like this:

> You don't have time to finish your language course. Write an email to the customer service team telling them why you can't continue, and say what you think about the course.
>   Write up to 150 words. You have 20 minutes.

4.3.4   Think-Aloud Protocols and Questionnaires

TAPs can gather on-line data on individuals' underlying cognitive processes, and is particularly useful in tracking the linear unfolding of their thought processes when they perform a task. Although concerns were expressed over the extent to which an

individual's process of thinking aloud actually alters the cognitive processes required to carry out the given task (Stratman & Hamp-Lyons, 1994), it is "the best research tool for teasing out the cognitive processes that reveal themselves in what we call audience awareness" (Berkenkotter, 1981: 389). TAPs are, therefore, adopted to keep track of when and how frequently the considerations about audience entered the test-taker's mind, and to what extent audience-related considerations guided his/her rhetorical, organisational, and stylistic decisions. Drawing on relevant literature, this study worked out a set of instructions to maximise the validity of this tool, including careful selection of those participants who are willing to verblise their thoughts when composing, adequate training with both mathematics problems (Ericsson & Simon, 1993) and sample email writing tasks, use of any preferable language of the participants, and setting no time limit on the whole procedure.

For the analysis of the TAPs, the coding scheme was adapted from the Berkenkotter's (1981) framework and also drawn on Wong (2005). The finalised coding scheme includes four categories. The first category, *analysing/constructing an audience*, consists of four strategies, namely, analysing audience's features (A), identifying self with audience (i.e., role-playing) (SA), identifying audience with self (i.e., projecting) (AS), and creating the rhetorical context (RC). The second category, *goal setting and planning for a specific audience*, refers to generating audience-related goals (AG) or refinements of the plan (R). The third category, *evaluating content and style with regard to anticipated audience response*, is concerned with writers' evaluation of audience's possible response to content (C) or style (ST) of the text after the audience reads it. The final category, *revising for a specific audience*, includes three strategies, i.e., reviewing the text with audience in mind, making sentence- or discourse-level changes (D), or word-level changes (W) in accordance with audience's characteristics or needs.

In addition to TAPs, this study also conducted the questionnaire survey and semi-structured interview to identify whether and to what extent test-takers showed their awareness of the audience when writing two emails. The questionnaire, designed on the basis of the TAPs coding scheme, were piloted and revised several times before being put to use. The final version, worded in Chinese in order to facilitate comprehension, consisted of 16 items and targeted specifically at whether, and if so how, test-takers took into consideration of audiences' needs and expectations. Table 4.3 presented a translated version of the questionnaire used in the study.

The semi-structured interview required test-takers to briefly recall their writing processes, and then elaborate on their considerations of audience specified in the writing prompt including such questions as *(T)o whom do they write in the first and second emails? What are the characteristics of their audiences? Did they notice the differences between these two audiences? If yes, how did they adjust their emails to the different emails?* Meanwhile, test-takers were asked whether or not they had taken raters into consideration when writing two emails. In the present study, interview was implemented in Chinese to facilitate free discussion.

4.3.5 Data Collection and Analysis

Data was collected in two phases. In Phase I, six test-takers were asked to say everything they thought about while they performed the writing task. After the completion of think-aloud, they participated in the questionnaire survey and semi-structured interview. In Phase II, 60 test-takers sat for the normal Aptis writing test in one computer room. Immediately after the Aptis Writing Test, the questionnaire survey was administered and 15 test-takers were randomly sampled to participate in the following interview administered one by one.

In order to enlist participant cooperation, test-takers were first informed of the purpose of this project and their time, support, and contribution were acknowledged and highly valued. To ensure the validity of the TAPs, six test-takers were trained one by one and allowed to think-aloud as long as they would like. Recordings were later transcribed word for word, producing between four and seven pages of transcribed text for each protocol.

Data collected in the present study included six TAPs, 132 emails, 66 questionnaires, and 21 interviews. TAPs and interviews were transcribed and double coded. Both TAPs and interviews were coded on the basis of the Audience Awareness coding scheme mentioned earlier. Given that test-takers were asked to reflect on the differences between the two audiences during the interview session, this study developed one more category when coding the interview data. The inter-coder correlation coefficient for the TAPs data was 0.87 before the resolution, and that for the interview data was 0.84. Disagreements were resolved after discussion. Finally, quantitative data was processed and summarized descriptively with SPSS. Constrained by the small sample, however, inferential statistics could not be performed to examine whether there were significant differences between low- and high-proficient writers.

4.4 Results

4.4.1 Details of Think-Aloud Protocols

Table 4.1 describes the time spent, the number and percentage of Chinese and English words used in the TAPs by six test-takers across two audience conditions, which indicates they varied greatly in every aspect.

For the time spent, five among six test-takers spent much longer in Email 2 than Email 1, which is understandable considering the required length of Email 2 is three times as long as Email 1. There is one exception, though. Hui spent less time in Email 2 than in Email 1. A closer look at his protocols revealed that he spent more than 15 min to create a full picture for an online language course. Later he realised he "thought too much about the online language course", and had to cut it down because of the length limit. Another English major, Liu, was caught in a similar

Table 4.1 Details of TAPs

Candi.	Gender	Email 1				Email 2				Total	
		Time	Chinese	English	Word	Time	Chinese	English	Word	Time	Word
Kun	M	22:17	557 (51%)	525 (49%)	1082	55:20	1734 (64%)	979 (36%)	2713	77:37	3795
Chi	F	11:58	182 (34%)	352 (66%)	534	76:20	2030 (54%)	1738 (46%)	3768	88:18	4302
Hui	M	27:04	4 (0%)	1863 (100%)	1867	23:06	347 (20%)	1415 (80%)	1762	50:10	3629
Liu	F	23:45	600 (31%)	1348 (69%)	1948	38:10	29 (1%)	2872 (99%)	2901	61:55	4849
Wei	M	22:30	2565 (91%)	239 (9%)	2804	33:32	2713 (85%)	465 (15%)	3178	56:02	5982
Zhang	F	16:07	604 (52%)	553 (48%)	1157	57:48	2476 (59%)	1698 (41%)	4174	73:55	5331

situation. She pondered over what type of language course it was and later had to delete several sentences because she wrote too much.

As to the language used, four test-takers (Kun, Chi, Wei, Zhang) code-switched during the TAPs and used both Chinese and English when brainstorming, although the percentages differed. Wei predominantly used Chinese, whereas other three participants presented a roughly even distribution. Different from these four test-takers, Hui in Email 1 and Liu in Email 2 used English in their protocols except when retrieving specific words (such as 'account' and 'irresistible') from their long-term memory. Another tendency is that compared with Email 1, test-takers used more Chinese in Email 2, with two exceptions (Liu and Wei). The possible reason might be that Email 2 is relatively more complicated and cognitively demanding.

In terms of time length and the number of words produced in the protocols, the two low-proficient test-takers, Kun and Chi, spent the most time but generated almost the least amount of protocols. Hui and Liu spent almost the least time, while the two employees, Wei and Zhang, generated the most protocols. In addition, female test-takers generally spent more time and produced more protocols than their male counterparts.

4.4.2 Audience-Related Strategies Across the Two Audience Conditions

A glance at Table 4.2 indicated test-takers did show some awareness of audience. In both emails, test-takers analysed the audience's characteristics and needs mainly with the tactics of considering the audience's name and creating the rhetorical context for the audience, and also evaluated the content and style of their emails with regard to anticipated audience response.

In terms of test-takers' use of audience-awareness protocols across the two emails, some salient patterns pop out. In Email 1 test-takers most frequently analysed their audience, whereas in Email 2 they most frequently evaluated the content and style with regard to audience needs and expectations. Also in Email 2, test-takers tended to generate more audience-related goals than in Email 1. These differences might reside in the different complexity of the two emails as well as in the different cognitive load required from the test-takers. But for both emails, test-takers seldom generated sub-goals or refinements of the plan, which means test-takers rarely changed or improved their plans once made.

Among the four categories of audience-related protocols, 'revising for a specific audience' was least used in both emails although sufficient time was allowed. It might be due to Chinese test-takers' habitual practice of one-shot draft in tests and their assignments. This result is cross-checked by the questionnaire and interview data.

A relatively larger pool of the questionnaire data (Table 4.3) indicated a similar pattern in the use of audience-related strategies as the TAPs displayed. Generally

Table 4.2 Response frequencies for audience-related strategies in TAPs (N=6)

Cand.	Email 1										Email 2										Total
	Analys.				Goal.		Evaluat.		Revis.		Analys.				Goal.		Evaluat.		Revis.		
	A	SA	AS	RC	AG	R	C	ST	D	W	A	SA	AS	RC	AG	R	C	ST	D	W	
Kun	1	/	/	1	/	/	/	/	/	/	/	/	/	2	/	/	/	1	/	/	5
Chi	1	/	/	1	/	/	/	/	/	/	1	/	/	1	1	/	1	2	/	/	8
Hui	1	2	/	3	/	/	2	/	/	/	/	/	/	1	/	/	2	/	/	/	11
Liu	1	/	/	3	/	/	1	4	1	/	2	/	/	2	1	1	2	2	/	/	20
Wei	1	/	/	1	/	/	/	1	/	/	/	/	/	2	1	/	1	/	/	/	7
Zhang	1	2	/	4	1	/	/	1	/	/	/	/	/	2	3	/	8	3	/	2	27
Total	6	4	0	13	1	0	3	6	1	0	3	0	0	10	6	1	14	8	0	2	78
Subtotal				23		1		9		1				13		7		22		2	

Table 4.3 Response frequencies for audience-related strategies in questionnaires (N = 66)

Statements	Completely agree	Agree	Disagree	Completely disagree
01. Consider what to write to meet the task requirements.	54	12	/	/
02. Consider the purpose of the email.	54	10	2	/
03. Consider whom to write.	46	15	5	/
04. Regard the rater as audience.	5	10	26	25
05. Analyse the features of audience.	17	33	14	2
06. Consider how to meet audience's expectations.	13	24	26	3
07. Consider audience's possible responses.	7	20	22	17
08. Consider the register.	38	21	6	1
09. Always take audience into consideration while writing.	14	34	16	2
10. First write out my thoughts and then revise.	8	27	25	6
11. Check whether the writing purpose has been accomplished.	38	25	3	/
12. Review/revise the email from audience's perspective.	3	17	28	18
13. Evaluate the content.	2	12	29	23
14. Evaluate the register.	10	22	24	10
15. Make sentence- or discourse-level changes.	2	11	29	24
16. Make word-level changes.	25	32	9	/

speaking, almost all the test-takers tended to consider what to write to meet the task requirement, the purpose, and the audience of the email. Few regarded the rater as their audience. Most test-takers analysed the features (e.g., characteristics and needs) of their audience, considered how to meet their audience's expectations, and decided on the register to be used. But less than one thirds of test-takers took into consideration the audience's possible responses after reading their emails. While writing, most test-takers took their audience into consideration, although the degree varied from person to person. Only one third of test-takers reported they first wrote out their thoughts and then revised their emails. After writing, almost all the test-takers checked whether or not the purpose had been accomplished. But less than one third reviewed or revised their emails from the audience's perspective. Even less evaluated the content or the register. Almost all the test-takers made word-level changes rather than sentence- or discourse-level changes.

The in-depth interview data offered us the opportunity to capture rich data about test-takers' use of audience-related strategies, revealing that candidates' analysis of the audience varied in the two audience conditions. In Email 1, 13 out of 15 interviewees reported a comparatively rich analysis of the features of their friend, and constructed this friend as an imaginary close friend on the basis of the writer's real friend. As for the other two interviewees, one regarded himself as the friend and the

Table 4.4 Response frequencies for low and high-proficient writers' audience-related strategies in questionnaires (N = *12*)

Statements	Low-proficient		High-proficient	
	AGREE	DISAGREE	AGREE	DISAGREE
01. Consider what to write to meet the requirements.	12	/	12	/
02. Consider the purpose of the email.	11	1	12	/
03. Consider whom to write.	12	/	12	/
04. Regard the rater as audience.	4	8	/	12
05. Analyse the features of audience.	2	10	12	/
06. Consider how to meet audience's expectations.	/	12	10	2
07. Consider audience's possible response.	/	12	9	3
08. Consider the register.	5	7	12	/
09. Always take audience into consideration while writing.	2	10	12	/
10. First write out my thoughts and then revise.	/	12	/	12
11. Check whether the writing purpose has been accomplished.	/	12	12	/
12. Review/revise the email from audience's perspective.	/	12	6	6
13. Evaluate the content.	/	12	9	3
14. Evaluate the register.	/	12	10	2
15. Make sentence- or discourse-level changes.	/	12	8	4
16. Make word-level changes.	4	8	10	2

Note: 'AGREE' here refers to both 'Completely agree' and 'Agree' in the questionnaire, by combining the two options together. The same is true for "DISAGREE"

other considered the rater as the friend. But when writing Email 2, most interviewees took the customer service team for granted but seldom analysed their features.

4.4.3 Effects of Writing Proficiency on Audience-Related Strategies

To probe into whether or not low-proficient and high-proficient writers analysed and constructed audience differently, the present study further investigated 12 low-proficient writers whose score, ranging from 1.5 to 2.5, ranked the lowest among 66 test-takers, and their 12 counterparts whose score, ranging from 5 to 5.5, ranked the highest. Table 4.4 lists perceived audience-related activities by the low-proficient and high-proficient writers. A glance at Table 4.4 reveals that low-proficient and high-proficient writers analysed and/or constructed audience in

similar but also dramatically different ways. Almost all the low-proficient and high-proficient writers considered what (content), who (audience) and why (purpose) before writing, despite one low-proficient writer reported he did not consider the purpose of the email. On the other hand, neither group admitted to first writing out their thoughts and then revising according to audience's characteristics or needs.

Aside from these similarities, low- and high-proficient writers exhibited distinctive differences in their audience analysis and construction strategies. To be specific, low-proficient writers did not analyse audience's features or characteristics but just wrote out their own thoughts. After finishing their draft, low-proficient writers did not check whether the writing purpose had been accomplished. Nor did they review or revise the email from the audience's perspective. Although one third of low-proficient writers reported that they made word-level changes in accordance with the audience's needs and characteristics, their revision was mainly concerned with spelling. They did not make any major changes at the syntactic or discourse level. As a matter of fact, low-proficient writers either struggled to generate contents or tried hard to translate their thoughts into English, hardly thinking beyond the content of their emails. For instance, when Kun spoke aloud his writing processes, he experienced great trouble in generating contents especially in the second email by laboriously searching for "what to write next". There was no surprise then when low-proficient writers paid scarce attention to the audience specified in the writing prompts.

Different from low-proficient writers, high-proficient writers were at great ease when translating their thoughts into English and many even planned their writing in English. All of them analysed the audience's features, and the overwhelming majority endeavoured to meet audience's expectations and evaluate the content with regard to anticipated audience response. When working on the second email, some high-proficient test-takers took into consideration the customer service team's response and expectations in order to make their writing more convincing. For example, when Zhang brainstormed the excuses for not having time to finish the language course, she evaluated the persuasiveness of her excuses.

> Why can't I finish this course? I am working in one government institution. I usually have a lot of work by the end of the year and have to stay up late. So I do not have time to learn online. This reason seems convincing enough.

Once the emails were drafted, high-proficient writers considered audience's possible response and checked whether or not the writing purpose had been accomplished. One half reviewed or revised the email from audience's perspective and the majority made word-level changes in accordance with audiences' characteristics and needs. Some even made sentence- or discourse-level changes such as resequencing the order of the content especially when test-takers compiled their reasons for lacking the time to finish their language course in the second email, deleting some content which test-takers regarded as irrelevant or inappropriate, etc.

For example, the TAPs data indicated that different from low-proficient writers, Zhang, a high-proficient writer, exhibited the greatest frequency and widest distribution of audience-related strategies. One remarkable feature of Zhang's audience-

related protocols is that she frequently evaluated audience response to her content, which accounts for almost one third of the total instances.

The above differences between low-proficient and high-proficient writers in their audience analysis and construction strategies can also be echoed by the evidence gained from the in-depth interview data, as illustrated by the following excerpts.

> I just wrote to a friend, quite close friend. I didn't write her name, either. Actually, I didn't analyse the specific features of the audience. (Test-taker 17, low-proficient)

> After I finished my writing, I usually check whether or not the writing purpose has been fulfilled. In this case, for the first email, I'll ask myself "did I recommend the course to my friend?". And for the second email, I'll look at whether the audience can understand what I've written and how will they react after reading my email. (Test-taker 26, high-proficient)

To further investigate how the low-proficient writers and the high-proficient writers dealt with both the communicative context presented in the writing task and the testing context embedded in this study, we looked specifically at test-takers' responses in Statements 04 and 05 and also explored the reasons behind. Surprisingly, two thirds of the low-proficient writers and all the high-proficient claimed they did not regard the rater as audience even though their performance would be rated by professional raters. The reason is that the writing task used in the present study clearly set the communicative context for the writers, that is, the test-taker joined an online language course and needed to write one email to their friend recommending the course and write another email to the customer service team to discontinue this course. Thanks to the authenticity of the context, the test-takers easily entered into and submerged themselves in this context and communicated with the two audiences like they did in real life. For example, Test-taker 17, high-proficient, stated during the interview, "No, I did not take the rater into consideration. In fact, I totally forgot this was a test!" Meanwhile, the high-proficient writers tended to analyse the features of audience such as "my friend Sara likes English very much and she dreams to improve it" (Test-taker 49, high-proficient).

The majority of the low-proficient writers also put themselves in the communicative context specified in the writing task. However, four among 12 low-proficient writers still wrote to the rater and ten did not analyse the features of the audience. There are two reasons which help account for this result. First, low-proficient writers seldom use English to communicate with their friends. Their English writing is mostly the argumentative writing with the only purpose of fulfilling the homework or testing requirements, as the questionnaire data showed, where audience is always the teacher or the rater. Second, low-proficient writers are characterised by "knowledge-telling" (Bereiter & Scardamalia, 1987) and could not decenter from his or her own perceptions of reality to consider the needs of the audience. As a matter of fact, some text-takers projected themselves as the audience. For example, Test-taker 37 explicitly reported "I imagined myself as my friend". Therefore, even the low-proficient writers were asked to write to their friend or the custom service

team, they failed to analyse the features of the audience and just wrote what they thought.

It should be noted that although the test-takers claimed they did not write to the rater, most of them did take the rater into consideration and tended to use safe expressions and structures as the interview data indicated. For instance, Test-taker 57, low-proficient, remarked "after writing I checked whether there were typos or grammatical mistakes" and Test-taker 42, high-proficient, stated "logic and accuracy were my top priorities".

4.5 Discussion

4.5.1 EFL Test-Takers' Audience Awareness Across the Audience Conditions

Different from Chen (2014), most test-takers in the present study paid great attention to the audience specified in the Aptis Writing Test. Before writing, they tended to think about both the purpose and the audience of their writing, and also created a rhetorical context for their writing. Meanwhile, they tend to analyse the audience's features and adapt their writing to different audience conditions, as both the questionnaire data and the displayed. When writing to their friend (the informal audience), test-takers more frequently analysed the characteristics of the audience compared with their writing to the customer service team (the formal audience). This might be because an identifiable audience, such as a friend, makes it easier to keep that audience in mind (Rubin & O'Looney, 1990). A remote and distant audience, the customer service team in this case, however, is just a far-away existence and makes the dialogue between the writer and the audience less accessible (Cohen & Riel, 1989). So many test-takers resorted to the traditional way of writing to express their own views or opinions regardless of the needs and/or expectations of the customer service team.

Another finding is that few test-takers regarded raters or teachers as the audience. Specifying the audience in the Aptis Writing Test, therefore, does not create a dilemma for Chinese test-takers, despite the fact that the writing test itself entails a kind of context, in which test-takers are aware that they are asked to write a text to be evaluated later by a certain rater in a certain way. Nevertheless, the influence of the testing context cannot be overlooked. Although most test-takers did not write to markers or teachers, they indeed tended to use 'safe' phrases or structures for the sake of accuracy and make word-level rather than sentence- or discourse-level changes even with the convenience of the computer facility. This holds true for both informal and formal audience conditions. To strengthen the positive influence of the communicative context presented in the writing task, test developers need to emphasise the communicative adequacy in their rating rubric.

4.5.2 Effects of Test-Takers' Writing Proficiency on Their Audience Awareness

In line with the previous research (Black, 1989; Kirsch, 1991; Wong, 2005), the present study also found that audience awareness is a robust indicator of writing proficiency and differentiates between high and low achievers. High-proficient writers employed relatively rich and varied strategies to conceptualise their audience and cater to their needs, such as analysing the needs of the audience and evaluating content and style with regard to anticipated audience responses. Their low-proficient counterparts, however, focused almost only on the content of their emails and rarely considered the needs and expectations of the audience. Such finding is understandable and reasonable because low-proficient writers face more struggle when transforming the inner language to written text (Gregg et al., 1996).

Although test-takers in the present study took into consideration of the needs and expectations of the audiences, low- and high-proficient writers conceptualised and constructed their audiences in markedly different ways. Generally speaking, low-proficient writers had great trouble in completing the task and rarely thought beyond the content of their emails. Compared with low-proficient test-takers' topic-bound protocols, their high-proficient counterparts adopted rich and varied audience conceptualisation and adaptation strategies. Among these strategies, the appropriate use of the evaluation strategy could help distinguish experienced writers from poor writers.

Another finding common in the testing setting is that for both audience conditions, test-takers seldom generated sub-goals or refinements of the plan, which means they rarely changed or improved their plans once made. This might be due to the often-used one-shot writing test which deprives test-takers of the opportunity to revise their essays due to test time limit and the pen-and-paper test format. Test-takers are thus accustomed to the first draft practice for the sake of cleanness of their written scripts. Even when writing in computer-delivered mode in the present study, they still hardly make major changes beyond the sentence level. Such practice goes contrary to some research suggesting writers should get their ideas down first before they can be expected to revise toward audience needs and expectations (Frank, 1992; Roen & Willey, 1988; Rubin & O'Looney, 1990) and advocating for considering the audience needs during the process of revision (Fitzgerald & Stamm, 1990; Midgette et al., 2008). Further research is, therefore, needed to explore whether or not test-takers who are familiar with the process writing approach, if encouraged to, revise their written scripts in accordance with the different audiences.

4.5.3 Contextulisation of EFL Writing Learning, Teaching, and Assessment Prompts

Writing is a communicative act. Just as Hamp-Lyons and Kroll (1997: 8) put it, writing is "an act that takes place within a context, that accomplishes a particular purpose, and that is appropriately shaped for its intended audience". Writing prompts should, therefore, better reflect such social nature. But the importance of the context is not well implemented in many large-scale writing assessments as well as classroom instructions, both in L1 (Behizadeh & Pang, 2016; Olinghouse et al., 2012) and L2 (Al-Mohammadi & Derbel, 2015; Chen, 2014) settings. This de-contexulisation of writing prompts confine writers to a cognitive and social vacuum and thus greatly diminishes audience awareness (Cohen & Riel, 1989). To make sure test-takers better display their writing performance, it is advisable and imperative, therefore, to contextulise writing prompts in both writing instruction and assessment.

Aside from the provision of contextualised writing prompts, another caveat both scholars and practitioners should bear in mind is that test-takers' performances should be evaluated in accordance with the desirable criteria in the context specified in the writing prompts. Otherwise, the whole endeavor might be jeopardized. For instance, working with 360 undergraduate education majors at two universities in America, Brossell (1983) tested the hypothesis that writing tasks with full contextual features would elicit higher quality essays than other less specified ones. Instead of evaluating the communicative effect from the target audience's perspective, the researcher informed student writers that their performance would be judged by trained raters, which might reinforce the typical assessment situation and thus weaken the effects of the contextualised writing task, thereby blurring the results. Currently, audience awareness is beginning to be used as an important trait on scoring rubrics for native (Oppenheimer et al., 2017) as well as nonnative writers (Cho & Choi, 2018). Although unestablished in terms of both validity and reliability in the implementation stage, this trait, undoubtedly worthwhile, calls for more in-depth research in the near future so as to broaden the construct of language proficiency and provide a valuable source of hypotheses for understanding an interactionalist approach to construct definition (Jin, 2017).

4.6 Conclusion

Both the online think-aloud data and off-line questionnaire and interview data collected in this study showed that awareness of the audience's needs and expectations is an important factor affecting test-takers' decisions on how and what to present when they compose their writing. Although whether, and if so how, using different audience-related strategies affects the quality of writing is still an open question, the findings of the present study can tentatively lend support to the

necessity of the inclusion of both informal and formal registers in the writing test given that test-takers adopted different audience conceptualisation and construction strategies when writing to different audiences. Meanwhile, specifying the contextual features engaged by the test-takers successfully weakens the understood testing context given that test-takers write to the audience specified in the writing prompts rather than write to raters for the mere purpose of getting a high score. The contextualisation of writing prompts is, therefore, of significant importance in task design which can have the potential to promote positive washback in teaching and learning.

Although the present study employed multiple research methods to triangulate the data, certain limitations were inevitable from the outset. Interpreting the results, therefore, must acknowledge the following limitations. First and foremost, except for two employees, 64 student test-takers were recruited from one university due to practical limitations of time and manpower. Moreover, many critical factors which might influence the process or product of EFL writing were not included, such as candidates' motivation for writing, L1 competence, etc. A large sample of candidates from a diverse of L1 backgrounds is needed to further explore the strategies used by test-takers to adapt to different audiences specified in the writing prompts.

Second, TAPs were used to investigate test-takers' audience conceptualisation and adaptation strategies. Although TAPs are very informative, they could not be exempt from the main criticisms, i.e., veridicality and reactivity (Ericsson & Simon, 1993; Stratman & Hamp-Lyons, 1994). Future research can adopt other methods with the help of modern technology, such as key board tracking and eye-tracking to better track test-takers' writing processes, and event-related potentials (ERPs) to measure how test-takers' brains response differently when writing to two different audiences.

References

Al-Mohammadi, S. A., & Derbel, E. (2015). To whom do we write? Audience in EFL composition classes. In R. Al-Mahrooqi (Ed.), *Methodologies for effective writing instruction in EFL and ESL classrooms* (pp. 197–208). IGI Global.

Behizadeh, N., & Pang, M. E. (2016). Awaiting a new wave: The status of state writing assessment in the United States. *Assessing Writing, 29*, 25–41.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Erlbaum.

Berkenkotter, C. (1981). Understanding a writer's awareness of audience. *College Composition and Communication, 32*, 388–399.

Black, K. (1989). Audience analysis and persuasive writing at the college level. *Research in the Teaching of English, 23*(3), 231–253.

Brossell, G. (1983). Rhetorical specification in essay examination topics. *College English, 45*(2), 165–173.

Camp, H. (2012). The psychology of writing development – And its implications for assessment. *Assessing Writing, 17*, 92–105.

Chen, Y. (2014). *An investigation into the context validity of EFL writing tests*. China Renmin University Press.

Cheng, F. (2005). Audience strategies used by EFL college writers. *Journal of Pan-Pacific Association of Applied Linguistics (PAAL), 9*(2), 209–225.

Cho, Y., & Choi, I. (2018). Writing from sources: Does audience matter? *Assessing Writing, 37*, 25–38.

Cohen, M., & Riel, M. (1989). The effect of distant audiences on students' writing. *American Educational Research Journal, 26*(2), 143–159.

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*. Available online: https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Crowhurst, M., & Piche, G. L. (1979). Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. *Research in the Teaching of English, 13*(2), 101–109.

Dryer, D. B. (2013). Scaling writing ability: A corpus-driven inquiry. *Written Communication, 30*(1), 3–35.

Dunlea, J. (2018). Aptis. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–7). Wiley.

Ede, L., & Lunsford, A. (1984). Audience address/audience invoked: the role of audience in composition theory and pedagogy. *College Composition and Communication, 35*, 155–171.

Ericsson, K. A., & Simon, H. A. 1993. *Protocol analysis: Verbal report as data* (Revised ed.). The Massachusetts Institute of Technology.

Fitzgerald, J., & Stamm, C. (1990). Effects of group conferences on first graders' revision in writing. *Written communication, 7*(1), 96–135.

Frank, L. A. (1992). Writing to be read: Young writers' ability to demonstrate audience awareness when evaluated by their readers. *Research in the Teaching of English, 26*(3), 277–298.

Gregg, N., Sigalas, S. A., Hoy, C., Wisenbaker, J., & McKinley, C. (1996). Sense of audience and the adult writer: A study across competence levels. *Reading and Writing, 8*(1), 121–137.

Hamp-Lyons, L., & Kroll, B. (1997). *TOEFL 2000 – Writing: Composition, community, and assessment* (TOEFL monograph series report no. 5). Educational Testing Service.

Jin, Y. (2017). Construct and content in context: Implications for language learning, teaching and assessment in China. *Language Testing in Asia, 7*(1), 12.

Kirsch, G. (1991). Writing up and down the social ladder: A study of experienced writers composing for contrasting audiences. *Research in the Teaching of English, 25*(1), 33–53.

Kirsch, G., & Roen, D. (1990). Introduction: Theories and research on audience in written communication. In G. Kirsh & D. H. Roen (Eds.), *A sense of audience in written communication* (pp. 25–39). Sage.

Lunsford, A. A., & Ede, L. (2009). Among the audience: On audience in an age of new literacies. In M. E. Weiser, B. Fehler, & A. M. González (Eds.), *Engaging audience: Writing in an age of new literacies* (pp. 42–69). National Council of Teachers.

Magnifico, A. M. (2010). Writing for whom? Cognition, motivation, and a writer's audience. *Educational Psychologist, 45*(3), 167–184.

McAndrew, D. A. (1982). *The effects of an assigned rhetorical context on the syntax and holistic quality of the writing of first year college students*. Unpublished dissertation at State University of New York.

Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and the audience awareness goals for revision on the persuasive essays of fifth and eighth grade students. *Reading and Writing, 21*, 131–151.

O'Sullivan, B. (2011). Language testing. In J. Simpson (Ed.), *Routledge handbook of applied linguistics*. Routledge.

O'Sullivan, B. (2012). *Aptis test development approach* (Aptis technical report (ATR-1)). British Council.

O'Sullivan, B., & Dunlea, J. (2015). *Aptis General technical manual version 1.0* (Technical Report TR/2015/005). British Council.

O'Sullivan, B., & Weir, C. (2011). Language testing and validation. In B. O'Sullivan (Ed.), *Language testing: Theory and practice*. Palgrave.

Odell, L. (1981). Defining and assessing competence in writing. In C. R. Cooper (Ed.), *The nature and measurement of competency in English* (pp. 95–138). National Council of Teachers of English.

Olinghouse, N. G., Zheng, J., & Morlock, L. (2012). State writing assessment: Inclusion of motivational factors in writing tasks. *Reading & Writing Quarterly, 28*(1), 97–119.

Oppenheimer, D., Zaromb, F., Pomerantz, J. R., Williams, J. C., & Park, Y. S. (2017). Improvement of writing skills during college: A multi-year cross-sectional and longitudinal study of undergraduate writing performance. *Assessing Writing, 32*, 12–27.

Park, D. B. (1982). The meanings of "audience". *College English, 44*(3), 247–257.

Porter, J. E. (1996). Audience. In T. Enos (Ed.), *Encyclopedia of rhetoric and composition: Communication from ancient times to the information age* (pp. 42–49). Garland.

Porter, D., & O'Sullivan, B. (1999). The effect of audience age on measured written performance. *System, 27*(1), 65–77.

Ransdell, S. E., & Levy, C. M. (1994). Writing as process and product: The impact of tool, genre, audience knowledge, and writer expertise. *Computers in Human Behavior, 10*(4), 511–527.

Roen, D. H., & Willey, R. J. (1988). The effects of audience awareness on drafting and revising. *Research in the Teaching of English, 22*(1), 75–88.

Rubin, D. L., & O'Looney, J. (1990). Facilitation of audience awareness: Revision processes of basic writers. In G. Kirsh & D. H. Roen (Eds.), *A sense of audience in written communication* (pp. 280–292). Sage.

Smith, W. L., & Swan, M. B. (1978). Adjusting syntactic structures to varied levels of audience. *Journal of Experimental Education, 46*, 29–34.

Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols: Issues for research. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89–111). Sage.

Weir, C. J. (2005). *Language Testing and Validation: An evidence-based approach*. Palgrave.

Willey, R. J. (1990). Pre-classical roots of the addressed/invoked dichotomy of audience. In G. Kirsh & D. H. Roen (Eds.), *A sense of audience in written communication* (pp. 25–39). Sage.

Wong, A. T. Y. (2005). Writers' mental representations of the intended audience and of the rhetorical purpose for writing and the strategies that they employed when they compose. *System, 33*, 29–47.

Zainuddin, H., & Moore, R. A. (2003). Audience awareness in L1 and L2 composing of bilingual writers. *TESL-EJ, 7*(1), A-2.

Ying Chen is Associate Professor in Applied Linguistics at Ocean University of China. She obtained her PhD from Shanghai Jiao Tong University, China. Her research interests are in writing assessment, vocabulary acquisition, pragmatics assessment, and second language development. She has led several funded projects on second language assessment.

Xiaoxian Guan, senior lecturer at East China Normal University, has acquired her doctorate degree in Linguistics and Applied Linguistics from Shanghai Jiao Tong University. Her research interests include language testing and assessment. She is currently in charge of the school-based English proficiency test of her university.