

# Chapter 9

## Theory-Based Approach to Academic Writing Assessment in Higher Education: A Conceptual Framework for Assessment Design and Development



Cecilia Guanfang Zhao

**Abstract** An examination of the current writing assessment practices indicates that unlike measurement theory, “writing theory has had a minimal influence on writing assessment” (Behizadeh and Engelhard, *Assess Writ* 16(3):189–211, 2011: 189). Despite the widely accepted theoretical conception of writing as a cognitive process and social practice situated in a particular socio-cultural context, the most often employed writing assessment task is still prompt-based impromptu essay writing, especially in large-scale English as a Foreign Language (EFL) assessment contexts. However, assessment specialists have long called into question the usefulness of impromptu essay writing in response to a single prompt. As a response to the above observation of the lack of theoretical support and generalizability of results in our current writing assessment practices, this chapter seeks to propose and outline an alternative writing assessment design informed by and reflecting more faithfully theoretical conceptions of writing and language ability. The chapter starts with a brief review of the existing writing theories and a survey of the current practices of writing assessment on large-scale high-stakes EFL tests, with a particular focus on those within the Chinese context. By juxtaposing theoretical conceptions of the construct and the actual operationalization of this construct on these EFL tests, the chapter highlights several salient issues and argues for an alternative approach to writing assessment, especially for academic purposes in higher education. A conceptual framework for such an assessment is presented to illustrate how writing theories can be used to inform and guide test design and development. The chapter ends with a discussion of the value and practical implications of such assessment design in various educational or assessment settings, together with potential challenges for the developers and users of this alternative assessment approach.

**Keywords** Conceptual framework · Writing assessment · Assessment design · Academic writing

---

C. G. Zhao (✉)  
University of Macau, Taipa, Macau, China  
e-mail: [czhao@um.edu.mo](mailto:czhao@um.edu.mo)

An examination of the current writing assessment practices indicates that unlike measurement theory, “writing theory has had a minimal influence on writing assessment” (Behizadeh & Engelhard, 2011: 189). Despite the widely accepted theoretical conception of writing as a cognitive process and social practice situated in a particular socio-cultural context, the most often employed writing assessment task, other than the earlier discrete-point test items, is still the prompt-based impromptu essay writing, especially in large-scale English as a Foreign Language (EFL) assessment contexts. However, assessment specialists have long called into question the usefulness of impromptu essay writing in response to a single prompt (Cho, 2003; Crusan, 2014). As a response to the above observation of the lack of theoretical support and generalizability of results in our current writing assessment practices, this chapter seeks to propose and outline an alternative writing assessment design informed by and reflecting more faithfully theoretical conceptions of writing and language ability. In the rest of the chapter, I will first offer a brief review of the existing writing theories developed over the past 50 years or so, and then survey the current practices of writing assessment on large-scale high-stakes EFL tests, with a particular focus on those within the Chinese context. I will then synthesize the problems in the current EFL writing assessment practices by comparing and contrasting our theoretical understanding of the construct and the actual operationalization of this construct on these EFL tests, and argue for an alternative approach to writing assessment, especially for academic purposes in higher education. A conceptual framework for such an assessment will then be presented to illustrate how writing theories can be used to inform and guide test design and development. The chapter will end with a discussion of the value and practical implications of such assessment design in various educational or assessment settings, together with potential challenges for the developers and users of this alternative assessment approach.

## 9.1 Theoretical Models of Writing/Language Ability

Existing theoretical models of writing or language ability over the past 50 years have changed from a more static and text-focused view to an increasingly more dynamic and contextualized conception of this construct. For example, writing was once conceptualized largely as mechanical and linguistic accuracy (Hatfield, 1935) and a set of linear processes (Britton et al., 1975; Rohman, 1965), before the influential cognitive process model (Flower & Hayes, 1981; Hayes, 1996; Hayes & Flower, 1980) presented it as a series of non-linear hierarchical mental processes and cognitive activities. Such activities as goal setting, generating, organizing, translating, reviewing, and monitoring, according to Flower and Hayes, could happen at any stage of the composing process, interacting with other factors such as the task environment and individual writers’ motivation, affect, and long-term and working memory capacity. This primary focus on cognition and individual writers was soon criticized for ignoring the sociocultural context in which any act of writing is

situated, and for its ineffectiveness in preparing students, especially L2 students, for academic writing tasks they would encounter in actual educational settings (Horowitz, 1986; Hyland, 2003; Johns, 1991; Spack, 1988; Swales, 1990). Following this sociocultural turn, researchers and practitioners turned next to genre-based conceptions of writing from various perspectives, including the systemic functional linguistics, English for specific purposes, and rhetorical genre studies approaches (cf. Bawarshi & Reiff, 2010; Hyon, 1996; Johns, 2008). A general consensus is that “genres are both social and cognitive” (Johns, 2008: 239); therefore, the analyses of “context, complex writing processes, and intertextuality” are all critical (Johns, 2011: 64).

In addition to these theoretical models of writing, more general conceptualizations of language ability also abound (e.g., Bachman, 1990; Bachman & Palmer, 1996, 2010; Canale & Swain, 1980). The most influential, particularly for use in test design and development, is probably Bachman and Palmer’s (1996) model of communicative language ability, where language knowledge and strategic competence together define the construct of language ability. The language knowledge dimension is further broken down to organizational and pragmatic knowledge, with the former covering grammatical and textual knowledge, while the latter functional and sociolinguistic knowledge. The strategic competence dimension, on the other hand, covers a series of metacognitive strategies of goal setting, planning, and appraising during actual language use. Bachman and Palmer (2010) believe that an individual’s language knowledge, strategic competence, topical knowledge and affective schemata, and the external factors of language use task and situation, together comprise a conceptual framework for language use.

Beyond general descriptions of writing or language ability, existing literature also offers various conceptions of academic literacy/literacies in particular, although covering the similar perspectives of writing as cognitive processes (academic literacy, Scardamalia & Bereiter, 1991) and social practices (academic literacies, Lee & Street, 1998, 2006). While it is beyond the scope of this chapter to fully unpack the whole field, interested readers may find Bloome et al.’s (2018) review informative. Here, I will briefly review one such model proposed by Snow and Uccelli (2009) when theorizing the challenges of academic language use for native and non-native English speakers alike. In this nested, pragmatics-based model of academic language, Snow and Uccelli (2009) argue that the ultimate purpose of academic literacy practice is to achieve “the two ubiquitous features of communicative tasks—representation of self and of one’s message—under particularly challenging conditions” (p. 122). According to their model, at the fundamental level is one’s ability to “organize discourse,” using discourse markers and reference terms to signal metatextual relationships and conform to conventions of a particular academic (often also technical) discourse community. This level of academic language ability is nested in a higher-order ability to “represent the message,” which involves the proper use of “approved academic genres,” appropriate level of detail and information for the intended audience, and the representation of “abstract, theoretical constructs, complicated interrelationships . . . and other challenging cognitive schemas,” while explicitly acknowledging “sources of information/evidence”

(p. 123). This ability to represent the message is further nested in a yet higher-order ability to “represent the self and the audience,” which entails effective academic voice and identity construction and establishment of a co-membership with an intangible, non-interactive, expert academic audience, through explicit display and extension of one’s knowledge and acknowledgement of “the epistemological status of one’s claims” (p. 123).

Despite the many models of writing/language ability, a general consensus among different schools of thoughts is clear: writing, especially academic writing, is more than producing a linguistically correct text; it is also a cognitive process, and a social interaction between and representation of the author and her audience in a particular communicative situation within in a particular historical and sociocultural context. Based on such an understanding, I now turn to the examination of operationalizations of the construct of (academic) writing ability, as represented by the writing tasks included on various large-scale high-stakes EFL tests in the Chinese context in particular. Juxtaposing the theoretical and the operational definitions of the construct helps reveal the extent to which current testing practices actually align with our theoretical knowledge about writing.

## **9.2 Operationalization of English Writing Ability on Chinese EFL Tests**

Four nation-wide large-scale high-stakes EFL tests are in use for educational evaluation and selection purposes at various stages of postsecondary education in China. These four tests, namely National Matriculation English Test (NMET), College English Test (CET-4 & 6), Test of English for English Majors (TEM-4 & 8), and the national Graduate School Entrance English Examination (GSEEE), affect the entire student population in China. Given the high-stakes nature of these tests, the way they assess English language proficiency, and writing proficiency in particular, will certainly have a huge impact on how EFL writing is conceptualized, taught and learned. A quick survey and analysis of the writing components on these tests, in terms of their design, tasks, and scoring rubrics, would present an operational definition of English academic writing ability within the Chinese EFL setting.

### ***9.2.1 National Matriculation English Test (NMET)***

According to the official *Guide for NMET* (National Education Examinations Authority, 2019), the writing part of the test intends to measure students’ ability to (1) convey information in a clear and coherent manner, and (2) effectively use the language knowledge they have acquired. Only one writing task is presented on NMET, specifying the basic rhetorical situation and asking students to write a short

text of approximately 100 words to convey specific information provided to them in the prompt in their native language. Prevalent genre types include emails, letters, memos, and announcements, although periodically picture descriptions and expositions can also be found. Four dimensions are included in the holistic scoring rubric, including the coverage of key points listed in the prompt, the diversity and accuracy of lexico-grammatical features, coherence and cohesion, and mechanics. As shown, the rubric ignores the effectiveness and appropriateness of communication despite that the task is framed as a rhetorically situated “authentic” task. Additionally, with the key information and ideas listed in bullet points and provided to the test takers in Chinese, the writing task is reduced to a translation task in essence (Dong et al., 2011), testing students’ ability to use lexico-grammatical features and control basic mechanics. Cai (2002) further points out that such a task design pre-determines not only the content of the writing but also the organizational structure, as most test takers would follow the order of those listed bullet points in their writing. Based on such observations, therefore, this type of writing is also known, among some Chinese scholars, as a “quasi-writing” activity (e.g., Chen, 2017; Lu, 2010).

### **9.2.2 College English Test Band 4 (CET-4) and Band 6 (CET-6)**

As outlined in the official *Guide for CET* (National College English Testing Committee, 2016), the writing part of CET-4 is designed to measure students’ ability to describe and narrate personal experiences, feelings, emotions and events, to describe and explain simple tables, graphs or other graphics, to offer personal opinions on familiar topics, and to handle practical writing. CET-6 builds on CET-4 and requires students to express their opinions on common topics, describe, explain and discuss information presented in tables, graphs, and other graphics. The major dimensions explicitly stated in the *Guide*, defining the construct of writing ability, remain the same across the two levels of the test and include the presentation of ideas, text structure and organization, language use, and use of writing strategies. Unlike NMET, therefore, CET writing tests value author stance and opinions, in addition to organization and language use. Interestingly, they also highlight the proper use of writing strategies that would facilitate the conveyance of ideas and content, although no further explanation is given in either the rubrics or the *Guide* as to what this dimension means and how it would be evaluated.

Both tests require test takers to complete their responses within 30 min and write in response to a single prompt, which oftentimes calls for an expository or argumentative genre, with the length requirement being slightly different (120–180 words for CET-4, and 150–200 words for CET-6). As is the norm in large-scale testing practices, a holistic rubric is adopted for scoring written responses. The rubric, however, only covers the first three dimensions outlined in the official *Guide* for the test, leaving out the assessment of writing strategy use. Moreover,

the descriptors on the rubric are often oversimplified and generic. As an example, the rubric defines the highest level of writing performance as one that is “on topic, with clear ideas, coherent organization, and correct language use” (National College English Testing Committee, 2016: 10). Probably because the rubric is generically constructed, it is applied to the scoring of both CET-4 and CET-6 writing samples. However, the test developers added a note in the *Guide*, stating that although the rubric is shared, CET-4 and CET-6 writing tests are “set at different difficulty levels and with different assessment requirements,” so that “the anchor papers of CET-4 and CET-6 that received the same-level ratings are in fact very much different in quality” (National College English Testing Committee, 2016: 10). With the use of the same scoring rubric and descriptors, it is hard to conceptualize and understand how a level 5 essay on CET-4 should be “very much different in quality” from a level 5 essay on CET-6. The single, prompt-based, often also decontextualized, writing tasks on CET tests also raise questions of its authenticity and interactivity, which in turn threatens validity (e.g., Gu & Yang, 2009; Cai, 2002).

### **9.2.3 Test for English Majors Band 4 (TEM-4) and Band 8 (TEM-8)**

According to Jin and Fan’s (2011) test review, TEM is an achievement test that intends to assess whether undergraduates majoring in English have achieved the required English proficiency at the end of their 4th semester and 8th semester during their undergraduate studies, hence TEM-4 and TEM-8, respectively. According to the official *Guide for TEM-4* (Pan, 2016), the writing section is designed to measure students’ basic competence in “written expression” through a performance task that requires students to write in response to a given prompt, graphic, or short reading excerpt. Students are expected to write approximately 200 words within 45 min in such genres as exposition, argumentation or narration. Written responses are evaluated in terms of content relevancy and adequacy, organization and coherence, and language accuracy and appropriateness. Similarly, the writing section on TEM-8 also measures writing ability through a performance task, although the official *Guide for TEM-8* states explicitly that TEM-8 only adopts an integrated reading and writing task (Deng, 2017). Students need to process *two* short reading excerpts and write approximately 300 words within 45 min on TEM-8; other than that, all the conditions and evaluative criteria stay the same as those for TEM-4. It should be noted that this integrated reading-to-write task type was only recently introduced onto TEM tests in 2016, after 25 years of impromptu opinion writing test.

### 9.2.4 *Graduate School Entrance English Examination (GSEEE)*

According to the official *Guide for GSEEE* (National Education Examinations Authority, 2018), the writing section comprises two tasks, one requiring students to complete a short 100-word practical writing in the genre of either letter, memo, abstract, or report (Task A), while the other requires students to write a conventional narrative, descriptive, expository, or argumentative essay of 160–200 words, based on a given prompt, picture, graph, or outline (Task B). The evaluative criteria cover the following four dimensions, as explicitly stated in the *Guide*: (1) correct use of grammar, spelling and punctuation, and appropriate use of vocabulary; (2) adherence to genre conventions; (3) appropriate organization that brings out clarity and coherence; (4) appropriate register in relation to the specified purpose and audience of the writing, if given. Written responses are rated holistically on a scale of 0–5. Despite the listed evaluative criteria in the *Guide*, however, the actual rubric seems to prioritize task completion (i.e., coverage of required content and points), lexicogrammatical accuracy, cohesive device use, proper register and format, as well as length of response as key evaluative criteria. Overall, therefore, the evaluation of writing ability, or that of text quality, still seems to focus on lexico-grammatical accuracy, due to either the neglect or the vague description of other dimensions.

## 9.3 Problem Statement and the Need for Alternative Approaches to Academic Writing Assessment

A brief review of the writing sections on these large-scale high-stakes national EFL tests points to the fact that they all assess writing based on a written product. In contrast, the writing theories developed in the past few decades have highlighted that “writing is text, is composing, and is social construction” (Cumming, 1998: 61), and that “effective writing integrates the product with the process within a specific context” (Hildyard, 1992: 1528). The EFL testing practices reviewed above, therefore, show a significant under-representation of the construct of writing ability.

Furthermore, an examination of the specific writing tasks reveals an overreliance on the use of decontextualized generic “essay” writing tasks. The endorsement of this impromptu opinion-based essay writing as more or less the only task type on these large-scale high-stakes EFL tests is particularly problematic. Such a task type fails to see writing as a social action and interaction situated in a particular rhetorical and sociocultural context, leading to not only construct underrepresentation but also a lack of authenticity in task design, which, from a testing perspective, could threaten the validity and usefulness of such an assessment approach (cf. Moore & Morton, 2005). Specifically, authentic academic writing tasks at the postsecondary level often involve in-depth and critical processing and use of sources, and evidence-based or data-driven argument construction and presentation, rather than a simple opinion

statement or personal response to an everyday topic. Sadly, however, these types of topics accounted for over 64% of all the CET writing topics, based on Gu & Yang's (2009) study of CET writing tasks over a period of two decades (1989–2008).

In addition to such personal topics adopted by most of the EFL tests in China, the prompts would often list all the key information that test developers expect students to cover in their responses. Such a design reduces a writing task to either a translation task (in the case of NMET where such bullet points are listed in Chinese) or a task that does not involve much thinking or planning or organization of content. In fact, Gu and Yang's (2009) study showed that 97.5% of all the CET writing tasks would fall into this category that they termed "outline-provided" type of writing, which is also the most prevalent type on GSEEE. In such cases, writing is underrepresented as linguistic accuracy and rigid formality only (Chen, 2017), as reflected in the rubrics themselves.

To be fair, however, recent years have witnessed certain changes in the writing task design on some of these EFL tests. As mentioned earlier, the TEM test battery recently introduced the reading-to-write task. Students are now required to process reading materials, although still rather limited in length and complexity, before they are required to produce a written text. This is certainly better aligned with authentic academic writing tasks, at least for the English majors who are expected to complete their coursework and degree thesis in English (such considerations may also explain why out of the four large-scale nation-wide EFL tests, only TEM seems to have implemented such a new task design). While this reform represents a step forward in the EFL test developers' conception of writing ability, a scrutiny of the prompt itself and the scoring rubric for this new task type still reveals a surface-level application of source-based writing. For example, the new TEM writing tasks only ask students to first summarize the main points in the reading passage(s), and then express their opinions on a related topic. What we could infer from such prompts is that source materials are used only for some generic summary tasks, independent of the subsequent writing task. Although the prompt also includes a line saying that students "can support [themselves] with information from the excerpt(s)," the use of and interaction with source materials are not explicitly required, hence unlikely to be valued. Indeed, if we turn to the actual scoring rubric, it becomes clear that none of such aspects of writing ability as knowledge construction, social interaction, and representation of self and audience is being considered or assessed. As an example, the descriptors used to evaluate TEM-8 writing samples define the highest level of student responses as those that showcase "effective communication with accuracy," which is further defined as fully addressing the writing task (i.e., contain both summary and opinion) with "logical organizational structure, ... clearly stated main ideas, and sufficient supporting details," and with "almost no errors of vocabulary, spelling, punctuation or syntax," while "[using] the language appropriately" (Deng, 2017: 30–33). Apparently, the adequate and critical use of source materials for academic interaction and communication is not included as part of the evaluative criteria. In fact, the descriptors are almost the same as those used in the rubrics for any conventional impromptu essay writing tasks. The only required use of source materials also stops at what Bereiter and Scardamalia (1987) would call the



“knowledge telling” level, neglecting that writing, particularly academic writing in higher education, is often for the purpose of knowledge transformation and construction.

Based on the above analysis of the overall EFL writing assessment design, the specifics of the writing tasks and prompts, as well as the scoring rubrics, it is not difficult to note that indeed “writing theory has had a minimal influence on writing assessment” (Behizadeh & Engelhard, 2011: 189). The developments in our theoretical understandings of the construct of writing and language ability are inadequately reflected in writing assessment practices, particularly and especially in the EFL context. Furthermore, as Hamp-Lyons (2016c) pointed out, writing assessment for academic purposes in higher education (HE) in particular has significantly lagged behind our “knowledge in what the language(s) of higher education look and sound like and how they ‘work’ linguistically, socially, culturally and interculturally” (p. 17). It is obvious that alternative means and forms of academic writing assessment are needed to more faithfully reflect the authentic writing tasks people encounter in HE and to better capture both the breadth and depth of this construct of academic writing ability. Only with the use of such alternative assessments, especially on the large-scale high-stakes tests, will we be able to introduce a more positive washback and ultimately help EFL students to develop the much-needed writing competence to support and facilitate successful academic communications and knowledge making in HE. The next section will hence present an alternative approach to academic writing assessment design and illustrate how writing-theory-informed design of academic writing tasks may be used to better capture the breadth and depth of the construct of academic writing at the tertiary level.

#### **9.4 A Theory-Based Approach to Academic Writing Assessment Design**

In order to address the aforementioned issues of construct underrepresentation, lack of authenticity, as well as the minimal influence of writing theory on our current writing assessment practices, writing assessments should evaluate not only the written product (writing as text), but also the writing process (writing as cognitive activity) and the social construction and interaction as mediated by the text (writing as social act). Of course, some attempts have already been made in the field to cover the breadth of the construct. The earliest and most commonly referenced attempt is the development and use of analytic rubrics, or what Hamp-Lyons (2016a, b) would call multiple-trait rubrics, when scoring students’ written products. By incorporating more dimensions and more detailed descriptors into the rubric, it is hoped that, in addition to the conventional trichotomy of content, organization, and language & mechanics, those often neglected components can also be evaluated, including for example, audience awareness, authorial voice, register and genre knowledge, pragmatic competence, communicative effect, citation and reference format, as well as

paraphrasing, summarizing and synthesis skills for integrated academic writing tasks in particular (e.g., Banerjee et al., 2015; Chan et al., 2015; Knoch, 2009). While analytic or multiple-trait scoring certainly contributes to a more systematic evaluation of the various dimensions that together define writing ability, the use of such rubrics is, nevertheless, limited in that not all aspects of the writing ability may be explicitly manifested in the written product and readily translated into a dimension on a rubric. Most obvious of all is the aspect of cognitive process and metacognitive strategy use involved in the completion of a writing task. The end-product may not be able to provide enough evidence for raters to reliably evaluate students' competence in these areas. This probably also explains why the official CET *Guide* includes the use of writing strategies as one of the four key dimensions of writing ability, but leaves it completely unattended to in the actual scoring rubric.

Perhaps to address some of these unresolved issues, Beck et al. (2015) recently proposed that we should go "beyond the rubric" in our evaluation of students' writing, and to use "think-aloud as a diagnostic assessment tool" to help us gain insights into the composing process so as to identify students' "strengths and challenges as writers, beyond what is discernible from evaluating their writing alone" (p. 670). While think-alouds can tap into the implicit composing processes, hence adding that part of the construct back into our assessment of writing ability, the applicability of such an assessment approach is probably limited to classroom use only, due to practicality considerations. Hence, new means and forms of academic writing assessment are needed. One alternative, I believe, is to streamline the composing process to the extent possible, eliciting the cognitive activities and social interactions in particular in our task design. Based primarily on Flower and Hayes's (1981) cognitive process model of writing, Bachman and Palmer's (1996, 2010) conception of communicative language ability, and Snow and Uccelli's (2009) nested model of academic language ability, I will illustrate how these theories may guide us in our design and development of a cognitive-process-based academic writing assessment for use with students in higher education.

As Flower and Hayes's (1981) model highlights task environment as an important dimension in any act of writing, writing assessment should also seek to specify for the test takers the topic and communicative purpose of writing, the rhetorical context, as well as the intended audience. In terms of topic selection, large-scale language tests often spare no effort to make sure that test takers write on a familiar topic. This probably explains why most of the writing tasks on the Chinese EFL tests surveyed above are about some aspects of students' everyday life. While minimizing the influence of topical knowledge on language performance is desirable for the purpose of assessing test takers' language knowledge, it is not when the purpose of assessment is to measure one's writing ability, especially for academic purposes in higher education. After all, a major function of language use in higher education is precisely for learning. We use language to learn about and communicate new information and ideas, new knowledge and discoveries. Consequently, writing tasks would only be more authentic, and fair too, if students can write to learn about a relatively new topic. In fact, empirical studies indeed revealed that almost all the university writing tasks "involved a research component of some kind, requiring

the use of either primary or secondary sources or a combination of the two,” as opposed to the writing tasks on language tests that focus primarily on “[writing from] prior knowledge” (Moore & Morton, 2005: 52). Similarly, Deane and his colleagues also pointed out that “writing in a school context is almost always engaged with, and directed toward, texts that students read, whether to get information, consider multiple perspectives on an issue, or develop deeper understandings of subject matter” (Deane et al., 2008: 78). An integrated writing assessment design is therefore more representative of and congruent with actual writing practices in authentic educational settings. The key is to provide the right type of input that would offer and stimulate ideas, and at the same time be comprehensible to students at a particular level of language proficiency and cognitive maturity.

To be more authentic, such input could, and probably should, go beyond one or two short excerpt(s) to include multiple sources and materials. If technologically feasible, such input could be made accessible to students through hyperlinks that would lead to further processing of additional materials of different degrees of relevance. Students’ ability to select relevant input for use in their writing may well be part of their academic writing ability, in that it would provide evidence into the information processing, critical thinking, and evaluation skills involved in actual academic writing. Meanwhile, the use of such materials and information also reflects another key dimension in Flower & Hayes’s (1981) model, wherein the writer’s long-term and short-term memory would interact and influence the composing process.

Once this task environment (i.e., rhetorical situation and topical knowledge) is specified in the prompt and input, the rest of the writing assessment can simulate the general process of composing and be organized into roughly three stages or sections, reflecting the three key components in Flower and Hayes’s (1981) model: planning, translating, and reviewing (see Fig. 9.1). Of course, these cognitive processes are nonlinear and recursive; however, it does not mean that they cannot be represented somehow on a test using a combination of pre-writing items, a main writing task, and post-writing items, to tap into these different cognitive activities and metacognitive strategies employed and deployed in the composing process.

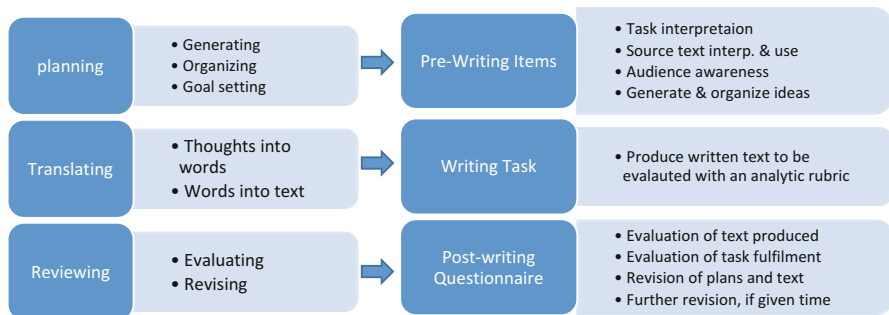


Fig. 9.1 Translating cognitive processes into test sections and items

As shown in Fig. 9.1, the planning stage in Flower and Hayes's (1981) model includes key components such as generating ideas, organizing ideas, and goal setting, which are also key elements in the strategic competence dimension of Bachman and Palmer's (1996, 2010) model of communicative language ability. In order to assess the ability of planning prior to the actual writing, pre-writing items may simulate the think-aloud process to elicit and examine the cognitive and metacognitive activities involved in the planning stage. Some items in the pre-writing section, for example, may ask students to articulate clearly their interpretations of the purpose, audience, and genre of the writing task in relation to the rhetorical situation given. Other items may ask students to list major ideas they think are relevant and important, and organize these ideas in an outline or bullet-point format. When the writing task involves and is based on students' processing of reading materials, pre-writing items could also measure students' understanding of the input and elicit their plans on *how* they intend to use such input for the purpose of the writing task.

Based on such planning, students can then translate these ideas and plans into an actual written text. Examining the written product eventually submitted in relation to their responses to the pre-writing items can reveal how original ideas and plans have been implemented, modified, or adjusted to varying degrees of success. This evidence may also be used to measure students' strategic competence, which is largely undealt with so far in existing writing assessment practices. Moreover, the written product could be scored using an analytic rubric designed to capture the multiple traits and dimensions of the construct of academic writing ability. In particular, the design of this analytic rubric should seek to restore the often-missing social dimension in L2 writing rubrics, highlighting the importance of the representation of the author and the audience in academic written interaction, as argued by Snow and Uccelli (2009) in their model of academic language use.

Conventionally, the writing section on most of the existing language tests would end here with the completion and submission of the final written product. Nevertheless, such a design does not faithfully reflect the recursive composing process. Successful writing almost always involves extensive revision, rewriting, and editing. Of course, writing assessment researchers and practitioners are not unaware of this mismatch. However, many believe it is simply impossible to address the recursive writing process in testing conditions. Hamp-Lyons and Kroll (1997), for example, pointed out that "other [non product-oriented] models that play a critical role in the field of composition studies may seem unhelpful, because they are not so much models of writing as a product as they are models of writing as a process," and noted specifically how the writing process model "is problematic for the design of academic writing assessment" (p. 7). Although no further explanation was given on why they believed the process model was "unhelpful" and "problematic," it was implied that tests and assessments can only be about products, despite the well-established process-oriented practices endorsed by writing teachers in various writing classrooms.

Unarguably, no test could fully emulate the authentic writing process due to practicality issues, particularly time constraints. However, this should not suggest that the assessment of writing ability could/should not go beyond that of the written product. The writing process, for example, could be captured, at least to a certain extent, by a pre-writing section that allows writers to demonstrate their planning and a post-writing section that prompts them to reflect on their writing processes as well as their plans for subsequent revision. In particular, this post-writing section could include items and tasks that ask students to (1) self-evaluate their writing and communicative success, and the overall task fulfillment, and (2) reflect on their own writing processes and strategies, including for example, how often, if at all, they evaluate their writing plans and products and revise their plans and texts while composing. Additional questions can be designed to probe into the subsequent revision plans by asking what types of revisions, if at all, they would focus on if they were given more time and resources. Such data, although self-reported, could still give us valuable information about the students' strategic competence, cognitive ability, and metacognitive strategy use that inform and influence their writing practices and performance.

## 9.5 Discussion and Conclusion

In response to the observation that current writing assessment designs and practices are inadequately informed by writing theories, an alternative design informed and guided by theoretical models of writing is proposed. As Cumming (1998) pointed out decades ago, "writing is text, is composing, and is social construction" (p. 61). Existing writing assessments, however, focus primarily on the assessment of written product, leaving out the composing process and the social construction. The design proposed here, therefore, expands on the current coverage of the construct by incorporating the cognitive processes (as informed by Flower and Hayes's cognitive process model) and strategic competence (following Bachman and Palmer's communicative language ability model) involved in composing and written interaction, and foregrounding the nature of writing as a social construction of meaning and relationships (as highlighted in Snow and Uccelli's nested model of academic language). Admittedly, in a large-scale testing context, not all aspects of the social functions of writing can be fully captured, particularly in terms of collaborative writing or using writing as a site for social and political actions (Cumming, 1998). In this chapter, therefore, the social aspect of the construct primarily focuses on the importance of situating meaning making in specific social cultural contexts and in relation to different communicative purposes and audiences. It also highlights the function of writing as a site for the author to build relations with readers, with prior texts, and gain voice and identity within a particular sociocultural context (cf. Bazerman, 2015; Beach et al., 2015; Snow & Uccelli, 2009).

Specifically, the pre-writing and post-writing items aim to make explicit the implicit cognitive and metacognitive activities involved in the composing process. This itemized design is also more practical than Beck et al.'s (2015) use of think-alouds, making it applicable to various testing and assessment contexts. In addition, a streamlined process-oriented design could also serve to raise students' awareness about the kind of thinking, planning, monitoring, and revising that are necessary for successful writing, making test taking a learning process in and of itself. Furthermore, highlighting the nature of writing as context-specific social construction of meaning and relationship in the design of the writing tasks and rubrics also help raise L2 writers' awareness about the "dialogic, [goal-oriented], and audience-directed quality of powerful writing, and . . . hone [their] understanding of how academic language choices are shaped by social contexts" (Beck et al., 2015: 680).

In addition to the positive impact on test takers and their test taking experience, such an alternative assessment design could also benefit other stakeholders, especially users of the assessment results and decision makers. Test takers' responses to the pre- and post-writing items would provide additional information about their writing performance and ability, adding discriminative power to the writing test as a whole. It is likely, for example, to have multiple, or sometimes even a large number of, test takers receiving the exact same score/rating on the essays they produce in response to a conventional writing prompt. It would be impossible to interpret, based on these essay scores alone, how one test taker may still differ from another. Data collected from pre- and post-writing items, however, could reveal varying levels of composing competence and strategy use, contributing to more nuanced and accurate interpretation of their writing ability. Such information could serve as the basis for important decision making by test users, including for example, placement decisions into various writing courses and curricula that target different instructional approaches and foci. Of course, when used by classroom teachers for diagnostic purposes, such information could greatly enhance pedagogical effectiveness and support differential treatment of individual writers' needs and challenges.

While such an alternative design creates opportunities for writing assessments to better represent the construct and bring about positive washback effect, it also poses a few challenges on the actual test development and administration. One such challenge is that it requires the test developers and item writers to have a solid understanding of relevant writing theories that could properly inform their task design and item writing. Without such theoretical knowledge, it is likely that the design of the items and tasks may misguide the test takers and distort the (meta)-cognitive processes, hence negatively influence students' writing performance. To address this issue, therefore, it is important that professional development and training be offered to test developers and item writers prior to the actual test development.

Another major challenge concerns the complexity of scoring. Such a contextualized, process-oriented design defies the use of any generic existing writing rubrics. Instead, it calls for the use of a combination of various scoring approaches and tailor-made rubrics to evaluate responses from different sections and items. In general,

holistic primary-trait scoring may be used for specific pre- or post-writing items that tap into various context-specific interpretative or (meta)cognitive skills, whereas multiple-trait and analytic scoring can be used to evaluate the main written product composed in response to the task-specific writing prompt. The design of these scales or rubrics will also need to be context- and task-based, although they may still incorporate categories we often find on existing generic writing rubrics.

In addition to proper choices of scoring approaches, score reporting could be yet another challenge. Should we report scores based on sections, reflecting test takers' ability to control the writing process? Or should we report scores based on skill areas, such as test takers' audience awareness, which could be reflected in their interpretation of the task, their writing plans, the actual written product, as well as in their plans for subsequent revisions? The decision, of course, will have to be made based on the purpose and focus of the assessment, together with considerations of the different stakeholders' needs and intended uses of such score reports.

One more decision to be made and justified is whether or not to penalize students' less-than-optimal planning in the pre-writing section, knowing that initial plans are likely to change during the recursive writing process. Likewise, precise interpretations of any observed discrepancy between a pre-writing plan and an actual written product could be a real challenge, as it would be difficult to tell whether the discrepancy is a result of the writer's conscious modification of initial plans during the writing process, or a reflection of his/her inability to execute the plans in the actual act of composing. A potential solution to such a problem is to design the post-writing items in a way that would elicit students' explicit reflections on the choices they made prior to and during the writing. This would allow us to gather information similar to that obtained from a think-aloud session, despite the retrospective route.

All the aforementioned challenges, however, do not outweigh the value-added benefits derived from the use of such an alternative assessment design, especially in EFL contexts that have long had a skewed representation of the writing construct both on their tests and in various writing classrooms and programs. Hopefully, with a new mindset that goes beyond the conventional product-oriented testing practice, together with the technological affordances available to us in this new era, we are able to design new assessments that more accurately reflect our current understanding of the construct under examination, instead of prioritizing only the measurement or psychometric issues. Only in this way will we be able to materialize the next generation of writing assessment, one that reflects an understanding of writing assessment as "both humanistic and technological" and "a complex of processes in which multiple authors and readers are involved and revealed" (Hamp-Lyons, 2001: 117).

## References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Banerjee, J., Yan, X., Chapman, M., & Elliot, H. (2015). Keeping up with the times: Revising and refreshing a rating scale. *Assessing Writing*, 26, 5–19.
- Bawarshi, A. S., & Reiff, M. J. (2010). *Genre: An introduction to history, theory, research, and pedagogy*. Parlor Press.
- Bazerman, C. (2015). What do social cultural studies of writing tell us about learning to write? In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 11–22). Guilford Press.
- Beach, R., Newell, G. E., & VanDerHeide, J. (2015). A sociocultural perspective on writing development: Toward an agenda for classroom research on students' use of social practices. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 11–22). Guilford Press.
- Beck, S. W., Llosa, L., Black, K., & Trzeszkowski-Giese, A. (2015). Beyond the rubric: Think-alouds as a diagnostic assessment tool for high school writing Teachers. *Journal of Adolescent & Adult Literacy*, 58(8), 670–681.
- Behizadeh, N., & Engelhard, G. (2011). Historical view of the influences of measurement and writing theories on the practice of writing assessment in the United States. *Assessing Writing*, 16(3), 189–211.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Erlbaum.
- Bloome, D., Carvalho, G. T., & Rue, S. (2018). Researching academic literacies. In A. Phakiti, P. D. Costa, P. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (1st ed., pp. 887–902). Palgrave Macmillan.
- Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities (11–18)*. Macmillan.
- Cai, J. (2002). The impact of CET-4 and CET-6 writing requirements and rubrics on Chinese students' writing. *Journal of PLA University of Foreign Languages*, 25(5), 49–53.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
- Chan, S., Inoue, C., & Taylor, L. (2015). Developing rubrics to assess the reading-into-writing skills: A case study. *Assessing Writing*, 26, 20–37.
- Chen, W. (2017). On NMET writing tasks and their future development. *China Examinations*, 302, 44–47.
- Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, 8(3), 165–191.
- Crusan, D. (2014). Assessing writing. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1–15). Wiley Blackwell. <https://doi.org/10.1002/9781118411360.wbcla067>
- Cumming, A. (1998). Theoretical perspectives on writing. *Annual review of applied linguistics*, 18, 61–78.
- Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C., & Bivens-Tatum, J. (2008). *Cognitive models of writing: Writing proficiency as a complex integrated skill* (ETS RR-08-55). Retrieved from <https://doi.org/10.1002/j.2333-8504.2008.tb02141.x>
- Deng, J. (Ed.). (2017). *Guide to test for English Majors, Band Eight*. Shanghai Foreign Language Education Press.
- Dong, M., Gao, X., & Yang, Z. (2011). A longitudinal study of writing tasks on the NMET national test papers (1989–2011). *Educational Measurement and Evaluation*, 10, 47–52.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387.



- Gu, X., & Yang, Z. (2009). A study of the CET writing test items over the past two decades. *Foreign Languages and Their Teaching*, 243, 21–26.
- Hamp-Lyons, L. (2001). Fourth generation writing. In T. Silva & P. K. Matsuda (Eds.), *On second language writing* (pp. 117–129). Lawrence Erlbaum Associates, Inc.
- Hamp-Lyons, L. (2016a). Farewell to holistic scoring? *Assessing Writing*, 27, A1–A2.
- Hamp-Lyons, L. (2016b). Farewell to holistic scoring. Part Two: Why build a house with only one brick? *Assessing Writing*, 29, A1–A5.
- Hamp-Lyons, L. (2016c). *Unanswered questions for assessing writing in the HE. What should be assessed and how?* Paper presented at the UKALTA Language Testing Forum (LTF 2016).
- Hamp-Lyons, L., & Kroll, B. (1997). *TOEFL 2000 – Writing: Composition, community, and assessment* (TOEFL monograph series report no. 5). Educational Testing Service.
- Hatfield, W. W. (1935). *An experience curriculum in English: A report of a commission of the National Council of Teachers of English*. D. Appleton-Century Company, Incorporated. Retrieved from <https://archive.org/details/experiencecurric00nati> (Original work published 1935).
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Lawrence Erlbaum Associates, Inc.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive process in writing: An interdisciplinary approach* (pp. 3–30). Lawrence Erlbaum Associates.
- Hildyard, A. (1992). Written composition. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (pp. 1528–1540). Macmillan Publishing Company.
- Horowitz, D. M. (1986). What professors actually require: Academic tasks for the ESL classroom. *TESOL Quarterly*, 20(3), 445–462.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12(1), 17–29.
- Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly*, 30(4), 693–722.
- Jin, Y., & Fan, J. (2011). Test for English majors (TEM) in China. *Language Testing*, 28(4), 589–596.
- Johns, A. M. (1991). English for specific purposes: Its history and contributions. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 67–77). Heinle & Heinle.
- Johns, A. M. (2008). Genre awareness for the novice academic student: An ongoing quest. *Language Teaching*, 41(2), 237–252.
- Johns, A. M. (2011). The future of genre in L2 writing: Fundamental, but contested, instructional decisions. *Journal of Second Language Writing*, 20(1), 56–68.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275–304.
- Lee, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157–172.
- Lee, M. R., & Street, B. V. (2006). The “academic literacies” model: Theory and applications. *Theory into Practice*, 45(4), 368–377.
- Lu, Z. (2010). Possible path of reforming English tests for College Entrance Examination based on application ability. *Educational Measurement and Evaluation*, 1, 15–26.
- Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4, 43–66.
- National College English Testing Committee. (2016). *Guide to College English Test (2016 revised version)*. Retrieved from <http://cet.neea.edu.cn/res/Home/1704/55b02330ac17274664f06d9d3db8249d.pdf>
- National Education Examinations Authority. (2018). *Guide to Unified National Graduate Entrance Examination-English I & II (for non-English major)*. Higher Education Press.

- National Education Examinations Authority. (2019). *Guide to National Matriculation English Test*. Retrieved from <http://www.neea.edu.cn/res/Home/1901/d15ec0514666ac280810099f9595b5557.pdf>
- Pan, M. (Ed.). (2016). *Guide to test for English Majors, Band Four*. Shanghai Foreign Language Education Press.
- Rohman, D. G. (1965). Pre-writing: The stage of discovery in the writing process. *College Composition and Communication*, 16(2), 106–112.
- Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 172–194). Cambridge University Press.
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge University Press.
- Spack, R. (1988). Initiating ESL students into the academic discourse community: How far should we go? *TESOL Quarterly*, 22(1), 29–51.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

**Cecilia Guanfang Zhao** obtained her Ph.D. in TESOL from New York University and is currently Associate Professor in the Department of English at University of Macau. Her research interests are in the areas of second language writing, language assessment, and writing assessment in particular.