Thinking Allowed

Second language writing development: A research agenda

Charlene Polio Michigan State University, USA
polio@msu.edu

1. Defining writing development

In 1998, Wolfe-Quintero, Inagaki & Kim published a monograph describing measures used in assessing writing development. Despite more recent research on linguistic development (e.g., Bulté & Housen 2012; Verspoor, Schmid & Xu 2012; Connor-Linton & Polio 2014), the volume is still a valuable resource and good starting point for anyone wanting to select measures of fluency, accuracy, and complexity for second language (L2) writing research. The volume, however, was limited to research on LANGUAGE development within the context of writing, and this is certainly one way to think about writing development. A recent edited volume by Manchón (2012), however, has greatly expanded conceptions of writing development showing that it involves much more that linguistic development, for example, genre knowledge (Tardy 2012) and goal setting (Cumming 2012) as well. Writing development can also focus on various aspects of the writing process and how writers' change their approach to text production as they become more expert writers (e.g., Sasaki 2004; Nicolás-Conesa, Roca de Larios & Coyle 2014). This expansion of focus, of course, makes development more difficult to define.

Wolfe-Quintero et al. (1998) defined language development as 'characteristics of a learner's output that reveal some point or stage along a developmental continuum' (p. 2). This is somewhat of a circular definition, but was provided so as to distinguish development from proficiency, which they say is a broader concept used to place learners into groups. Manchón (2012) refrains from offering a specific definition of development, but most of the chapters in her book seem to define it as some change over time related to some aspect of writing. Given the lack of specific definition, I offer one here that I will use for the remainder of this article. I am defining writing development as change over time in any of the following areas related to written text production: language (e.g., complexity, accuracy, fluency, cohesion, mechanics); knowledge of different genres; text production processes; metacognitive knowledge and strategy use; and writing goals and motivation. Of course, this list is not exhaustive, and there are other areas that may exhibit change over time. In addition, I would like to add a caveat. Development may not always coincide with improved quality or movement toward a target norm. Without belaboring what I mean by quality, the most obvious example here might be related to linguistic complexity. Over time, learners' language might become more complex, but at some point, learners must vary their sentence structure, and too many long

and complex sentences can cause essays to be judged as lower in quality. Furthermore, given the recent focus on multilingualism in second language acquisition research, exemplified in a study of writing development by Kobayashi & Rinnert (2012, 2013), it also important to think of writing development as bidirectional in that L2 writing can also influence first language (L1) writing development. Imagine the common case of graduate students studying in an English-speaking country reading and writing about their field in English but not in their L1. One might expect changes in their L1 but not necessarily ones that were positive (e.g., a Chinese speaker not being able to recall Chinese characters). Nevertheless, development mostly includes progress toward some target, even though that target will vary according to context and the purposes for writing.

2. Methodological challenges

My definition of development as 'change over time' raises a number of challenges to consider when proposing any study of development, be it related to written or spoken language. I focus here on the issues of sample size, task and genre equivalence and authenticity, and the observer effect.

First, it is uncontroversial that longitudinal studies are difficult and time-consuming. The ability to track a large number of writers for any period of time longer than one semester or year is often impossible. One exception is the work of Miyuki Sasaki, who summarized much of her work in Sasaki (2009). She followed 22 Japanese writers of English for three and a half years. This is one of the most impressive longitudinal studies of L2 writers, and she included students who studied abroad for different amounts of time and some who did not. By the standards of experimental studies, the groups had very small numbers of participants, but a rather large number of participants for such a long-term study.

Most longitudinal studies of L2 writing development are case studies (e.g., Li & Schmitt (2009), who tracked lexical phrases in academic writing; Macqueen (2012), who tracked lexicogrammatical patterns; Nassaji (2007), who tracked spelling development). While such focused qualitative studies can be useful, they lack the strengths of quantitative studies (e.g., generalizability and group comparisons) and lack the strength of other qualitative studies (e.g., a holistic view of learning to write in an L2 from multiple perspectives). Note that both Li & Schmitt and Macqueen included interviews to supplement writing data, which is a methodological improvement on case studies limited to text analysis. Tardy (2012), discussed in Section 4, followed students for two years using several data sources to study genre development, but studies collecting a wide variety of data for prolonged periods are rare. Trade-offs between quantitative and qualitative approaches exist throughout applied linguistics studies, and research on writing development is no different. One solution is to supplement large cross-sectional data of large groups with case studies tracking individual learners (e.g., Kobayashi & Rinnert 2013) or to take a mixed-methods approach (e.g., Hashemi & Babaii 2013 and Brown 2014).

Second, even if we can secure a large number of participants, collecting data from participants on a range of topics, tasks, and genres (preferably counterbalanced) is ideal. We

know that topic and genre influence language (e.g., Lu 2011; He & Shi 2012), and it is unlikely that learners follow the same writing process across all tasks. Furthermore, writing tasks used in quantitative studies are often relatively short timed essays that may not reflect writing tasks used in academic classes. Researchers are left with two options. The first is to create a group of similar timed-writing writing prompts on a variety of topics and counterbalance them throughout the longitudinal data collection period. For example, if a researcher wants to collect writing from students over the period of a year at six points, he or she could create six prompts on six different topics for which students have to support a point of view. The students could write for 60 minutes without outside sources, and the topics would be counterbalanced among the participants. In this way, genre and access to outside help are controlled for and topic effects are masked by the counterbalancing. Most longitudinal studies of writing development do not control for topic or genre type, making it difficult to compare writing at different times. Polio & Park (2016) reviewed recent longitudinal research on linguistic development in writing and found only one study (Yasuda 2011) that counterbalanced the topics and tasks. Tightly controlled studies, however, run the risk of lacking ecological validity in that the writing is timed, does not allow students to work through a typical extended writing process, and generally uses impromptu tasks without outside sources.

Finally, studies that investigate aspects of writing other than writers' texts will have challenges that many qualitative researchers face, namely, how to observe a process without altering it. For example, collecting introspective or retrospective data from participants draws their attention to aspects of writing under investigation. For example, if one wanted to study how students' editing strategies developed, simply asking them about such strategies might cause the students to change what they do. These issues are relevant to any study of writing processes but effects might be magnified if such data collection techniques are used repeatedly over time. With these problems laid out, I turn to specific research tasks with the caveat that these methodological issues persist. I have grouped these tasks into four different areas: linguistic development, holistic development, intervention studies, and conceptions of development in curricular and assessment contexts.

3. Linguistic development

Polio & Park (2016) provide a synthesis of research on linguistic development in writing beginning with an historical perspective and detailing the current research. The majority of the studies were descriptive and focused on specific grammatical features, specific constructs, specific contexts, and modality differences. In addition, we reported on descriptive studies that were theoretically motivated and experimental intervention studies. Based on the gaps found in the literature, I propose three research tasks related to linguistic development in L2 writing.

> **Research task 1**
> **Analyze language development within specific theories of second language acquisition to test whether theories can be extended to written language.**

Early second language research attempted to tie stages of language development to theories of SLA. For example, Gass (1979) examined relative clause development and tied it to universals in language typology, and Pienemann (1989) investigated word order acquisition across several languages explaining development according to processing constraints. These studies focused on oral language development and may not be applicable to writing given less strict time constraints, but it might be useful to consider studies that are more theoretically motivated. Three specific approaches include usage-based approaches, dynamic systems approaches, a subset of usage-based approaches, and systemic functional linguistics, a theory of language as opposed to a theory of SLA.

There are many different conceptions of usage-based SLA (Tyler 2010), but many studies focus on matters of the frequency of structures, vocabulary, and chunks of language in the input. Most descriptive studies of written linguistic development ignore how exposure to oral and written language may affect development. One possible study would be to examine learners' writing over an extended period of time and to compare their language to a corpus representing the language they were exposed to. While this might be extremely difficult, it could be more easily done in a foreign language setting where input is likely limited to teacher talk and classroom materials. Li & Schmitt (2009) tracked the development of lexical phrases but relied on the participant's recollections of where she learned the phrases. A corpus-based study of classrooms and students' writing in those classrooms could more reliably track and link such phrasal development. Verspoor & Smiskova (2012) compared students' writing development from low and high input groups but did not actually examine the input itself to confirm that the quantity and quality of the input was indeed different. Note that any longitudinal study in which students' writing and oral and written input were collected would be very labor intensive, but such a data set could then be used for a variety of studies.

Dynamic systems approaches acknowledge the importance of exposure and of language use for learning, but they also focus on variability within individuals and variation among learners. Several studies of written language (Larsen-Freeman 2006; Verspoor et al. 2012; Verspoor & Smiskova 2012) have been conducted within this framework. These studies have shown that different features of language develop at different points and do not develop linearly. These studies emphasize that development is complex and, as Larsen-Freeman (2006) stated 'The messiness is not "noise", but rather a natural part of dynamically emergent behavior assembled by the individual with a dynamic history of engaging in such tasks, with his or her own self-identified (or jointly identified) target of opportunities for growth' (p. 615). While research in this framework is usually very carefully done and shows interesting patterns that we might miss otherwise, it does not attempt to explain why differences within and across individuals occur. The Verspoor & Smiskova (2012) study was a first step in that it attempted to compare writers from different contexts. It did not control for task or genre, two factors that might causes variation. Thus, a useful study might follow some students' development of various linguistic features but compare development on different tasks. It could be that the complexity of certain tasks drives development, for example tasks deemed more cognitively complex such as those that involve more steps or more reasoning demands (Robinson 2001). Similarly, a comparison of variability and variation between speaking and writing (on similar tasks) could be illuminating. It is quite possible writing increases variability because the learner has more time to draw on multiple cognitive resources, namely explicit and implicit

knowledge. It would be interesting to see the different rates of development of various features in the two modalities. (See Research task 2.)

Finally, most of the studies of L2 linguistic development are not done within any specific theory of language. Some studies, but few on writing, are conducted within a formalist framework. Han (2000) is an exception in that she brings in formalist descriptions in her analysis of writers' use of the pseudopassive. The remainder of the studies often tacitly assume a functional perspective without any discussion of a specific theory. The exception has been the various studies conducted within a systemic functional approach to language. Byrnes (2009) traced the development of grammatical metaphor, a concept taken from systemic functional linguistics, in English learners of German. (See also Liardét (2013), who studied Chinese learners of English). While the theory may not appeal to everyone, having a specific theory of language that guides an analysis of language makes it easier to make cross-study comparisons as well as cross-language comparisons (e.g., see discussion of Byrnes & Sinicrope (2008) below).

Research task 2
Determine how oral and written language do or do not coincide for different populations.

In Polio (2012a, 2012b), I argued that a core issue in L2 writing research was the relevance of the medium itself. Because writing can include slower language production than speaking, learners can focus on form more easily than in speaking and, in fact, writing may facilitate the acquisition of speech. Williams (2012), in particular, makes a strong argument for this while Tarone & Bigelow (2005) show that a lack of L1 literacy skills can affect oral language processing. Harklau (2002) suggested that some learners may try out forms in writing before speaking, and Weissberg (2000), in his profiles of five Spanish learners of English, shows that for four of the five learners, most structures appeared first in writing. Although Weissberg showed that individuals vary in how their oral and written linguistic development relate to each other, if we find, as Weissberg did, that certain structures are more likely to emerge in speaking or writing, teachers can exploit this insight and focus on helping students to transfer those structures. (See Research task 6.)

Kuiken & Vedder (2012) reviewed studies of oral and written language, but only one study, Bulté & Housen (2009) examined development. They found that learners did not progress in the same way across the two modalities on measures of lexical development. Byrnes & Sinicrope (2008) studied relative clause development and were able to compare their finding to studies of oral development in relative clauses, finding that development was different. This could be done only because the development of relative clauses in oral language is fairly well attested. What is lacking are longitudinal studies with comparable counterbalanced tasks that would confirm that differences are not simply task related. In addition, studies comparing oral and written development could be conducted in both foreign language and L2 contexts where the relative amounts of oral and written input might be different as well as students' amount of explicit attention to form. The studies would have teaching implications and might hint at how the two skills are or are not related in specific populations of learners. For example,

heritage learners (i.e., students who are exposed to a language to some degree at home) often come to foreign language classes with strong oral and weak writing skills. How these two sets of skills do or do not interact and affect each other for this type of learner is not clear.

Research task 3
Update Wolfe-Quintero et al. (1998) and conduct more validation studies on measures used in L2 writing.

Much progress has been made since the publication of Wolfe-Quintero et al. (1998) in the area of measures used in L2 writing research. For example, when the volume was published, the most common measure of complexity was clauses per T-unit, perhaps because the addition of dependent clauses made sentences look more complex. But since Norris & Ortega's (2009) call for a multidimensional and organic (i.e., ecologically valid) measure, complexity has been widely researched and it has been acknowledged that complexity is multifaceted and can also include, for example, nominal complexity and coordination (e.g., Bulté & Housen (2014)). In addition, lexical complexity was often measured using type-token ratios, but because this measure can be affected by text length, studies (e.g., Kormos 2011) now use other measures (e.g., D-value: Richards, Malvern & Graham 2008) that correct for length. Nevertheless, a systematic investigation of the measures needs to be redone.

Much less progress has been made in measuring accuracy and fluency, or even clarifying what is meant by fluency in the context of writing. (See discussion in Norris & Ortega 2009 regarding fluency as well as in Abdel Latif 2013.) Studies using keystroke logging software (e.g., Miller, Lindgren & Sullivan 2008) can examine features more similar to those discussed with regard to oral fluency such as pausing or self-corrections. (See the extensive discussion of oral fluency in Segalowitz 2010.) With regard to accuracy, Wolfe-Quintero et al. (1998) has been the only study that extensively investigated the validity of accuracy measures, but as they explained, their results were difficult to interpret because of problems with the studies that they reviewed. We know of only two studies that longitudinally tracked changes in accuracy. Gunnarsson (2012) examined the relationship among accuracy, fluency, and complexity in the writing of L2 learners of French, but she examined four specific features and because of a small sample size and a low number of occurrences, firm conclusions could not be drawn. Polio & Shea (2014) studied students' writing over the course of a semester and found no change in any of the accuracy measures. More and more research is suggesting that accuracy alone is not a sufficient indication of development (cf. Verspoor et al. 2012; Bulté & Housen 2014), which makes the question of which accuracy measure to use as an indication of development more difficult to answer.

4. Focus on broader conceptions of writing development

Again, drawing on the work of Manchón (2012), it is evident that writing development concerns much more that lexical-grammatical development. What's more is that we don't

know much about how individual issues such as motivation, goal setting, and beliefs about writing affect lexical-grammatical development. Research tasks 4 and 5 address these gaps.

> **Research task 4**
> **Examine how students progress over time in terms of genre knowledge and genre-specific text production.**

Tardy (2012) provides a cogent discussion of genre in relation to L2 writing development explaining that genre is a type of social practice and that L2 writers may bring a set of L1-based expectations to the task of producing a genre in their L2 that do not match their new social setting. Much research on writing development, particularly linguistic development, relies on timed writing which may bear little resemblance to writing done for academic courses. Thus, longitudinal research on specific genres and tasks, both with and without explicit instruction, may provide insight in how to teach specific genres.

Two examples of studies that examined specific genres are Yasuda (2011) and McDonough, Crawford & Vleeschauwer (2014). Yasuda studied Japanese English as a foreign language (EFL) students as they were taught to write emails over the duration of a class that used a variety of simulated email assignments. Her study was particularly unique in that it also examined how genre learning contributed to L2 learning. For example, she studied lexical complexity and found no change in diversity but found that students used more sophisticated lexical chunks. McDonough et al. examined summary writing in Thai EFL students over 17 weeks and examined the use of word strings from the source texts. They found that although students used more copied word strings, the strings were shorter. As with many studies done in instructional contexts, neither of these studies used control groups. It would be helpful to conduct similar studies with control groups to better determine the effects of specific genre instruction as opposed to general writing instruction. Another approach to examining writers' development of specific genres would be to follow L2 writers as they learn a genre in real life. For example, how do international students learn email conventions? Furthermore, while genre analyses of various workplace genres (e.g., Flowerdew & Wan 2006) exist, few studies follow new employees as they learn workplace genres. One exception is Parks (2001), who followed nurses over 22 months to determine how their writing of medical genres changed from what they learned in school to workplace conventions.

Finally, although there is a huge amount of research on how students incorporate outside sources into their academic writing, most studies do not follow students as they learn this complex skill. Among the exceptions are Wette (2010), who studied how students used sources after eight hours of instruction. Although she found a small improvement, the students still had trouble integrating sources. Incorporating sources involves linguistic and field-specific conventions, but most studies simply interview students about how they have acquired the knowledge they have instead of tracing it longitudinally. Investigating what knowledge students bring to the task of writing with sources and how that knowledge changes might help us determine what does and does not need to be explicitly taught.

> **Research task 5**
> **Conduct holistic case studies of multilingual writers at critical transitions in their writing experiences.**

Leki & Carson (1997) conducted a study in which they interviewed students about writing in their English as a second language (ESL) classes and in their content classes. As part of this study, they were able to interview students at the beginning and end of a semester. Leki (1995) also conducted an in-depth semester-long case study of five international students, some of whom were required to take ESL. Both of these studies were important as they dealt with students' writing needs in their content classes and how ESL classes may or may not prepare students for content classes. It would be particularly interesting to follow individual students across semesters through critical transitions. Extended case studies would allow us to target writers at specific points that might influence writing development. For example, little has been written about international students in writing classes for native speakers. One example of a longitudinal case study would be to follow a student from arrival and placement into an English for academic purposes (EAP) writing class, then into a native-speaker writing class, and then to content classes. The point here is that different instructional practices in the EAP and the native-speaker composition classes could affect the writer's process, knowledge, and perceptions of writing. Furthermore, we would learn about disjunctures between expectations among the three contexts and how the student deals with changes in expectation. Ultimately, we could evaluate how and if the three different types of classes, as well as the students' L1 and L1 writing instruction, contributed or not to the student's development in the broadest sense. Such a study would involve student interviews, class observations, teacher interviews, analysis of students' writing, text-based interviews, and observations of other opportunities for assistance (e.g., writing center conferences).

In the above example, however, typical EAP students in North America may not be actively engaged in writing in his or her L1. Longitudinal cases studies could also target writers who were truly engaged in writing in more than one language. Consider the case of students who attend university in an English-speaking country, receive a doctorate, and then return to their home country and are expected to write in both English and their L1. This return to writing in their L1 is a critical transition, and we know little about how such scholars' L1 writing is influenced. Another much less researched example might be foreign language majors who have to make the transition from language to literature classes, where writing conventions in the L2 might not be made explicit and students are left to draw on their L1 discourse knowledge. While there is much research showing how multilingual writers draw on multiple linguistic repertoires such as Cenoz & Gorter (2011), who examined crosslinguistic relationships in trilingual writers' essays at one point in time, we do not how these languages do or do not develop together. Detailed longitudinal case studies might illuminate influences on the development of the different languages, particularly during changes in instruction or transitions in contexts for writing. It is through these kinds of studies that we might better understand the bi-directional relationship of more than one language on writing.

5. Intervention studies

Intervention studies in L2 writing research generally investigate the effects of instruction or the effects of some type of feedback at some point in the writing process. Polio & Park (2016) reviewed various longitudinal studies conducted in a variety of instructional contexts, including study abroad (Serrano, Tragant & Llanes 2012), secondary school foreign language classes (Benevento & Storch 2011) and genre-based EFL classes (Yasuda 2011) among many others. The studies used pre and posttests to examine development but did not have control groups. Conversely, some well-conducted intervention studies of feedback, namely Van Beuningen, De Jong & Kuiken (2012), were short-term studies whose effect could have dissipated over an extended period of time. What is missing from the literature are well-designed longitudinal experimental studies (but see Hartshorn et al. 2010). Of course, such studies have both the problems of longitudinal studies (e.g., attrition) and the problems of experimental studies (e.g., random assignment, controlling for variables), so they are particularly difficult. Research task 6 focuses on replicating previous studies while Research task 7 focuses on new areas of inquiry.

> **Research task 6**
> **Conceptually replicate previous intervention studies by conducting longitudinal studies.**

One approach to this research task is to consider well-designed intervention studies then replicate them using a longitudinal design. Such studies would not be exact replications nor approximate replications (see Porte 2012 for a discussion of types of replications) but rather conceptual replications that research questions similar to a previous study but change the research design. The purpose of such studies would be to determine the lasting or cumulative effects of a repeated treatment. The Van Beuningen, de Jong & Kuiken (2012) study is an excellent example of a well-designed study that examined written corrective feedback but used only one treatment. They found an effect for feedback, but if this effect is not sustained, it may be a waste of time for teachers to spend so much time providing grammar feedback. Another example is Shintani & Ellis (2013), who studied the effects of both corrective feedback and metalinguistic explanation after one treatment. They found that metalinguistic explanation helped learners improve their accuracy after one treatment but that the effect was not sustained in a delayed posttest. If the feedback were repeated over an extended period of time, there could be a long-term effect on students' accuracy. Conversely, any of the longitudinal studies mentioned above that focus on instructional context would be useful to replicate with control groups so as to ascertain the source of student development.

> **Research task 7**
> **Conduct intervention studies that focus on linguistic dimensions of L2 writing other than accuracy.**

A great majority of intervention studies focus on corrective feedback and accuracy. This is likely because of Truscott's (1996) call to abandon grammar feedback and because giving language feedback is a time-consuming task. But such a focus may be misguided. Some research (Verspoor et al. 2012; Bulté & Housen 2014) has suggested that L2 learners' language first becomes more complex and then more accurate, but we know little about what drives complexity. For example, the nature of the writing prompt can affect syntactic complexity; L2 writers produce more complex language in argumentative than narrative essays (Lu 2011; Yoon & Polio 2016). What we do not know is whether or not writing certain genres over an extended period of time results in actual language development.

Some studies have examined the effects of task complexity on written language (e.g., Kuiken & Vedder 2008), but the purpose of these studies was not to examine development. Interventions related to the use of more complex tasks could be implemented over an extended period of time to determine if such intervention results in more syntactically and lexically complex language. Other features such as lexical phrases or cohesive devices could be studied as well. Furthermore, as mentioned earlier, some studies have examined differences in oral and written language, but few have attempted interventions in one modality to determine the effects on the other modality.

6. Conceptions of writing development in instructional and assessment contexts

There has been some attempt to look at language development across many languages (e.g., Pienemann 1989) but no attempt to look at writing development broadly across a range of contexts. Teachers, curriculum writers, and even test designers do their jobs without clear descriptions of writing development, but they may have some intuitive sense of how students' writing develops. The previous research tasks are intended to inform teaching and assessment, but perhaps we could approach development in reverse, specifically, perhaps we could examine how development is manifested in a variety of curricula, standards, and assessment contexts. This approach is addressed in Research tasks 8 and 9.

Research task 8
Analyze and compare descriptions of curricula, standards, and assessment realizations of writing development for common themes.

Reynolds (2012) proposed that given the lack of a common yardstick for writing development, we should look at various descriptions of writing development to see what features of writing are included and how these features are expected to develop over time. In a preliminary attempt to look at such description, he examined the Arizona guidelines (http://www.azed.gov/english-language-learners/elps/), the Canadian Language Benchmarks (http://www.language.ca/), and the Georgetown German program (http://german.georgetown.edu/page/1242716536825.html). Reynolds examined how these three descriptions characterized different levels of writing proficiency in terms of features such

as syntax, length, genre knowledge, vocabulary, cohesion, and intertextuality, among others. It might be useful to continue this line of research by also including, for example, the ACTFL proficiency guidelines and the Common European Framework of Reference. These various guidelines were developed for different populations, but if we compare them, it might help us understand more general expectations for writers and give us a starting point for researching task and genre difficulty. More specifically, one could determine expectations for writers as they progress through levels. These findings could then be compared to task-related descriptors, such as are found in the task complexity literature (e.g., Robinson & Gilabert 2007) and language-related descriptions of L2 writing development (as summarized in Polio & Park 2016). The main point here is to try to bring together descriptions of L2 writing development even though such descriptions emerge out of different contexts. Although this research task is somewhat exploratory, a comparison of how test and curriculum developers conceive of writing development could inform others needing to create curricula. Such a comparison might also suggest areas of research on writing development.

Research task 9
Compare assessment rubrics to actual development.

There is no doubt that large-scale tests such as the TOEFL or IELTS have been well validated with regard to the construct of writing quality and experts' assessment thereof. What is not clear is that students' scores will improve on these rubrics as their writing develops over time. This also speaks to the relationship between quality and development. It seems that test scores should be somewhat related to actual development. In other words, one way to validate a writing test is to determine whether or not students improve their scores as they continue to study. Because improvement in quality and development do not always coincide, it is not the only way to validate a test, but it is one way. A related study was conducted by Brown (1989) but with a reading test. He administered a reading test and determined which items showed improvement from pre- to post-instruction. From these items, he created a new reading placement test so that the test would be a more valid measure of how students reading developed over the semester. His point was that the placement test should be capable of measuring change in reading proficiency.

Although several testing researchers have discussed procedures for rubric revision (e.g., Knoch 2011; Harsch & Martin 2012), no published studies that I know of have revised rubrics based on what actually develops in students' writing. In Polio (2013), I revised a placement test analytic rubric to reflect raters' comments about what they saw as improving in students' writing over the course of a semester. The new rubric was more reliable and correlated better with a holistic rating.

As an example, as mentioned earlier, because complexity has been shown to improve over a semester while accuracy may not, we should question the emphasis on accuracy in some test rubrics, particularly placement or achievement test rubrics if we know that students will not improve their accuracy in the short term. This is obviously a complex and controversial issue, particularly because raters might continue to consider accuracy, but one which should be discussed.

7. Final thoughts

Writing development covers a broad range of areas, so there is ample research to be done, much of which could not be covered here, including research on non-essay types of writing such as synchronous computer-mediated communication (SCMC), social media posts, or texting. For example, some research has looked at translanguaging within one context (e.g., Oliver, Grote & Nguyen 2014, who examined Facebook). Other research has explored the relationship between SCMC and oral language (Blake 2009). These contexts are wide open for studies on how L2 writers develop over time and on how writing in such contexts might affect more formal writing.

Although there are some exceptions, the majority of the research on L2 writing development has been conducted with learners of English. Many of the studies conducted on English learners should be replicated with learners of other languages, but a few specific areas stand out. First, longitudinal cases studies could be conducted following zero beginners enrolled, perhaps, in an intensive university language program (e.g., such as a US university government Language Flagship Program) that was intended to get students to a high level of proficiency at the end of four years. Another type of study would be to follow students before and after studying abroad with a focus on writing, such as Sasaki (2004) did, but with learners of a language other than English. Finally, there is very little written about how students develop writing skills in character-based languages as most of the research on Japanese and Chinese is related to reading.

References

Abdel Latif, M. M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics* 34.1, 99–105.

Benevento, C. & N. Storch (2011). Investigating writing development in secondary school learners of French. *Assessing Writing* 16.2, 97–110.

Blake, C. (2009). Potential of text-based internet chats for improving oral fluency in a second language. *The Modern Language Journal* 93.2, 227–240.

Brown, J. D. (2014). *Mixed methods research for TESOL*. Edinburgh: Edinburgh Textbook for TESOL.

Brown, J. D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly* 23.1, 65–83.

Bulté, B. & A. Housen (2009). The development of lexical proficiency in L2 speaking and writing tasks by Dutch-speaking learners of French in Brussels. Paper presented in the colloquium 'Tasks across modalities', Task-based Language Teaching Conference, Lancaster, UK.

Bulté, B. & A. Housen (2012). Defining and operationalizing L2 complexity. In Housen, Kuiken & Vedder (eds.), 21–46.

Bulté, B. & A. Housen (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing* 26.4, 42–65.

Byrnes, H. (2009). Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor. *Linguistics and Education* 20.1, 50–66.

Byrnes, H. & C. Sinicrope (2008). Advancedness and the development of relativization in L2 German: A curriculum-based longitudinal study. In L. Ortega & H. Byrnes (eds.), *The longitudinal study of advanced L2 capacities*. New York, NY: Routledge, 109–138.

Cenoz, J. & D. Gorter (2011). Focus on multilingualism: A study of trilingual writing. *The Modern Language Journal* 95.3, 356–369.

Connor-Linton, J. & C. Polio (2014). Comparing perspectives on L2 writing: Multiple analyses of a common corpus, *Journal of Second Language Writing* 26.4, 1–9.

Cumming, A. (2012). Goal theory and second-language writing development, two ways. In R. M. Manchón (ed.), 135–164.

Flowerdew, J. & A. Wan (2006). Genre analysis of tax computation letters: How and why tax accountants write the way they do. *English for Specific Purposes* 25.2, 133–153.

Gass, S. (1979). Language transfer and universal grammatical relations. *Language Learning* 29.2, 327–344.

Gass, S. & A. Mackey (eds.) (2012). *Handbook of second language acquisition*. New York: Routledge.

Gunnarsson, C. (2012). The development of complexity, accuracy, and fluency in the written production of L2 French. In Housen, Kuiken & Vedder (eds.), 247–276.

Han, Z. (2000). Persistence of the implicit influence of NL: The case of the pseudo-passive. *Applied Linguistics* 21.1, 78–105.

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing* 11.4, 329–350.

Harsch, C. & G. Martin (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing* 17.4, 228–250.

Hartshorn, K. J., N. W. Evans, P. F. Merrill, R. R. Sudweeks, D. Strong-Krause & N. J. Anderson (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly* 44.1, 84–109.

Hashemi, M. R. & E. Babaii (2013). Mixed method research: Toward new research designs in applied linguistics. *The Modern Language Journal* 97.4, 828–852.

He, L. & L. Shi (2012). Topical knowledge and ESL writing. *Language Testing* 29.3, 443–464.

Housen, A., F. Kuiken & I. Vedder (eds.) (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA*. Amsterdam and Philadelphia: John Benjamins.

Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing* 16.2, 81–96.

Kobayashi, H. & C. Rinnert (2012). Understanding L2 writing development from a multicompetence perspective: Dynamic repertoires of knowledge and text construction. In R. M. Manchón (ed.), 101–134.

Kobayashi, H. & C. Rinnert (2013). L1/ L2/ L3 writing development: Longitudinal case study of a Japanese multicompetent writer. *Journal of Second Language Writing* 22.1, 4–33.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing* 20.2, 148–161.

Kuiken, F. & I. Vedder (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing* 17.1, 48–60.

Kuiken, F. & I. Vedder (2012). Speaking and writing tasks and their effects on second language performance. In S. Gass & A. Mackey (eds.), 364–377.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics* 27.4, 590–619.

Leki, I. & J. Carson (1997). 'Completely different worlds': EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly* 31.1, 36–69.

Leki, I. (1995). Coping strategies of ESL students in writing tasks across the curriculum. *TESOL Quarterly* 29.2, 235–260.

Li, J. & N. Schmitt (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing* 18.2, 85–102.

Liardét, C. L. (2013). An exploration of Chinese EFL learners' deployment of grammatical metaphor: Learning to make academically valued meanings. *Journal of Second Language Writing* 22.2, 161–178.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly* 45.1, 36–62.

Macqueen, S. (2012). *The emergence of patterns in second language writing: A sociocognitive exploration of lexical trails*. Bern: Peter Lang.

Manchón, R. M. (ed.) (2012). *L2 writing development: Multiple perspectives*. Boston: Walter de Gruyter.

McDonough, K., W. J. Crawford & J. De Vleeschauwer (2014). Summary writing in a Thai EFL university context. *Journal of Second Language Writing* 24.2, 20–32.

Miller, K. S., E. Lindgren & K. P. H. Sullivan (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly* 42.3, 433–454.

Nassaji, H. (2007). The development of spelling and orthographic knowledge in English as an L2: A longitudinal case study. *Canadian Journal of Applied Linguistics* 10.1, 77–98.

Nicolás-Conesa, F., J. Roca de Larios & Y. Coyle (2014). Development of EFL students' mental models of writing and their effects on performance. *Journal of Second Language Writing* 24.2, 1–19.

Norris, J. & L. Ortega (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics* 30.4, 555–578.

Oliver, R., E. Grote & B. Nguyen (2014). Australian Indigenous students developing literacy and code-switching skills through online communication: A study of Facebook as an education tool. In A. C. Cree (ed.), Literacies: The power to change, Volume 2 – Literacy in a changing world. Oxford: World Literacy Summit, 88–104.

Parks, S. (2001). Moving from school to the workplace: Disciplinary innovation, border crossings, and the reshaping of a written genre. *Applied Linguistics* 22.4, 405–438.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics* 10.1, 52–79.

Polio, C. (2012a). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing* 21.4, 375–389.

Polio, C. (2012b). The acquisition of second language writing. In S. Gass & A. Mackey (eds.), 319–334.

Polio, C. (September, 2013). Revising a writing rubric based on raters' comments: Does it result in a more reliable and valid assessment? Paper presented at Midwest Association of Language Testers. Michigan State University, East Lansing, MI.

Polio, C. & J. H. Park (2016). Language development in second language writing. In R. M. Manchón & P. Matsuda (eds.), *Handbook of second and foreign language writing*. New York: Routledge, 287–306.

Polio, C. & M. C. Shea (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing* 26, 10–27.

Porte, G. (2012). Introduction. In G. Porte (ed.), *Replication research in applied linguistics*. Cambridge: Cambridge University Press, 1–18.

Reynolds, D. (October, 2012). Second language writing development: Implications for (from) curriculum development. Paper presented at the Second Language Research Forum, Carnegie Mellon University, Pittsburgh, PA.

Richards, B., D. Malvern & S. Graham (2008). Word frequency and trends in the development of French vocabulary in lower-intermediate students during Year 12 in English schools. *The Language Learning Journal* 36.2, 199–213.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22.1, 27–57.

Robinson, P. & R. Gilabert (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching* 45.3, 161–176.

Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning* 54.3, 525–582.

Sasaki, M. (2009). Changes in English as a foreign language students' writing over 3.5 years: A sociocognitive account. In R. M. Manchón (ed.), 46–76.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Serrano, R., E. Tragant & A. Llanes (2012). A longitudinal analysis of the effects of one year abroad. *Canadian Modern Language Review* 68.2, 138–163.

Shintani, N. & R. Ellis (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing* 22.3, 286–306.

Tardy, C. M. (2012). A rhetorical genre theory perspective on L2 writing development. In R. M. Manchón (ed.), 165–190.

Tarone, E. & M. Bigelow (2005). Impact of literacy on oral language processing: Implications for second language acquisition research. *Annual Review of Applied Linguistics* 25, 77–97.

Truscott, J. (1996). The case against grammar journal correction in L2 writing classes. *Language Learning* 46.2, 327–369.

Tyler, A. (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics* 30, 270–291.

Van Beuningen, C. G., N. H. De Jong & F. Kuiken (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning* 62.1, 1–41.

Verspoor, M. & H. Smiskova (2012). Foreign language writing development from a dynamic usage based perspective. In R. M. Manchón (ed.), 17–46.

Verspoor, M., M. S. Schmid & X. Xu (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing* 21.3, 239–263.

Weissberg, B. (2000). Developmental relationships in the acquisition of English syntax: Writing vs. speech. *Learning and Instruction* 10.1, 37–53.

Wette, R. (2010). Evaluating student learning in a university-level EAP unit on writing using sources. *Journal of Second Language Writing* 19.3, 158–177.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing* 21.4, 321–331.

Wolfe-Quintero, K., S. Inagaki & H. Y. Kim (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawai'i.

Yasuda, S. (2011). Genre-based tasks in foreign language writing: Developing writers' genre awareness, linguistic knowledge, and writing competence. *Journal of Second Language Writing* 20.2, 111–133.

Yoon, H. J. & C. Polio (2016). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, doi: 10.1002/tesq.296.

CHARLENE POLIO is a Professor in the Department of Linguistics & Germanic, Slavic, Asian, & African Languages at Michigan State University. She is a co-editor of *The Modern Language Journal*. Her research interests on L2 writing include linguistic development and research methods. In addition, she has published research on classroom discourse in the foreign language classroom and differences between preservice and experienced teachers. Her latest book (with Debra Friedman) is *Understanding, evaluating, and conducting second language writing research* (Routledge, 2017).