Articles

Building Teachers' Assessment Capacity for Supporting English Language Learners Through the Implementation of the STEP Language Assessment in Ontario K-12 Schools

Saskia Van Viegen Stille, Eunice Jang, & Maryam Wagner

The Ontario Ministry of Education recently implemented the Steps to English Proficiency (STEP) language assessment framework to build educator capacity for addressing the needs of English language learners (ELLs) in K-12 schools. The STEP framework is a set of descriptors-based language proficiency scales that specify observable linguistic behaviours from which educators can make inferences about students' English language development. Teachers use these proficiency scales to assess, document, and track students' language proficiency development based on daily interactions with students in classrooms. The purpose of this article is to report on teachers' perceptions of and experiences with the STEP proficiency scales during a three-year pilot implementation and validation study of the initiative. Based on analysis of these findings, we articulate implications for building teachers' assessment capacity using observational language assessment scales for K-12 ELLs.

Le Ministère de l'Éducation de l'Ontario a récemment mis sur pied un cadre d'évaluation des compétences linguistiques (Steps to English Proficiency - STEP) pour accroître la capacité des enseignants à répondre aux besoins des apprenants d'anglais dans les écoles K-12. Le cadre STEP est un ensemble d'échelles de compétences linguistiques basées sur des descripteurs qui décrivent des comportements linguistiques à partir desquels les enseignants peuvent faire des inférences quant au développement des élèves en anglais. Les enseignants se servent de ces échelles de compétence pour évaluer, documenter et suivre le développement langagier de leurs élèves dans leurs interactions quotidiennes avec les élèves en classe. L'objectif de cet article est de faire état des perceptions et des expériences des enseignants relatives aux échelles de compétences STEP pendant les trois années de la phase de mise en oeuvre initiale et d'étude de la validité de l'initiative. Nous nous appuyons sur les résultats de notre analyse pour formuler des implications relatives à l'accroissement de la capacité des enseignants par l'emploi des échelles de compétences linguistiques auprès d'apprenants d'anglais en K-12.

English-medium schools in the province of Ontario have a large number of students learning the language of instruction while at the same time learning content curriculum. For example, among the 2.1 million students in the province, 27% were born outside of Canada, and a great proportion of these students speak a language other than English or French at home—the two official languages of Canada (Gallagher, 2014). Further, educators are increasingly recognizing the language learning needs of Canadian-born children for whom English is an additional language. Both immigrant and domestic English language learners (ELLs) face the challenge of learning academic subjects in a new language that is necessary to set them on a positive trajectory for curriculum learning and academic success. Over the past 10 years, the Ontario Ministry of Education has developed policies and resources to build educator capacity for addressing these students' learning needs. Among these resources is the Steps to English Proficiency (STEP) language assessment framework. The STEP framework primarily consists of grade-specific, descriptors-based proficiency continua. The STEP continua specify observable linguistic behaviours from which educators can make inferences about students' English language development. Teachers use these proficiency scales to assess, document, and track students' language proficiency development based on daily interactions with students in classrooms. The purpose of this article is to report on teachers' perceptions of and experiences with the STEP proficiency scales during a three-year pilot implementation and validation study of the initiative. Based on analysis of these findings, we articulate implications for building teachers' assessment capacity using observational language assessment scales for K-12 ELLs.

Support for K-12 English Language Learners in Ontario Schools

In Canada, education is provincially mandated; therefore, each province is responsible for the design and implementation of initiatives to support and guide learning in curricular contexts. Ontario is one of Canada's largest provinces and is its most populated province (Statistics Canada, 2014). As a consequence, there are many differences in how Ontario schools provide support for English Language Learners (ELLs) due to demographics and the diversity of board-level priorities across the province. For instance, school boards with a large population of ELLs may have schools with one or more full-time English as a Second Language (ESL) teacher(s), whereas school boards with a smaller population may have itinerant ESL teachers who provide support to a family of schools. Similarly, some students have the opportunity to attend their local school and receive ESL support, whereas others may need to be transported by bus to a hub school that provides ESL support within the district. Language teaching and learning support provided to individual students in Ontario schools depends on both the learner's level of English language proficiency and school resources. At the elementary level, students at

SASKIA VAN VIEGEN STILLE, EUNICE JANG, & MARYAM WAGNER

beginning stages of proficiency are often placed in self-contained ESL classes for part of the school day and integrated into mainstream classes for the remainder of the day. Students at higher levels of proficiency tend to be fully integrated into mainstream classes, with monitoring and support provided by an ESL teacher working in collaboration with the classroom teacher. At the secondary level, students can take up to five ESL credit courses in place of mainstream English courses, with the possibility of adding non-ESL English courses when they are prepared to do so. Some secondary schools also offer locally developed subject-area courses such as ESL history and science for ELLs depending on the level of demand and the resources available.

School boards, and in some cases individual schools, have considerable discretion concerning the delivery of ESL programs, leading to different levels of support being provided to students with similar needs (Auditor General of Ontario, 2005). Furthermore, there has been no consistent mode of assessing and tracking the developmental trajectories of ELLs in Ontario schools, nor has there been a consistent and common language for discussing students' progression across grades, schools, and districts. In 2005, the Auditor General of Ontario issued a report calling for a more consistent approach to meeting the needs of ELLs and both ESL and mainstream teachers in the province. The Ministry initiated the development of STEP as a means to achieve these purposes, providing a resource for ESL and mainstream classroom teachers and language assessors to use to identify, describe, and monitor students' English language proficiency development. Further, STEP was developed to build capacity for

- determining student placement;
- supporting planning and programming decisions;
- implementing differentiated instruction and assessment;
- selecting appropriate teaching and learning resources;
- making decisions regarding student participation in and support for large-scale assessment;
- engaging students in self-assessment and goal setting;
- identifying possible special learning needs;
- providing students and parents with accurate indications of the child's level of English language acquisition and literacy development;
- determining discontinuation of ESL support;
- promoting reflective teacher practice;
- providing an opportunity to focus teacher reflection and professional dialogue (Ontario Ministry of Education, 2012, p. 3)

STEP Language Proficiency Descriptor Scales

The STEP framework is a set of descriptors-based language proficiency scales to be used by teachers based on their daily interactions with students in their

classes. The framework includes four sets of scales, one for each grade cluster: primary (grades 1–3), junior (grades 4–6), intermediate (grades 7–8), and secondary (grades 9–12). The STEP scales comprise descriptors targeting three areas of language use: Oral Communication, Reading and Responding, and Writing based on the ESL curriculum (please refer to the Ontario Ministry of Education 2007 curriculum documents). Separate scales are provided for students in English Literacy Development (ELD) programs, reflecting the fact that the latter group has experienced gaps in their educational background and/or opportunities for literacy development in their home language.[1] The descriptors in each scale articulate a cumulative progression through increasingly complex forms of communicative competence, comprising six "steps" of language proficiency development. Sample descriptors for each of the three language modalities, and across six STEPs for students in Grades 1 to 3, are shown in Table 1.

Designed specifically for use in the Ontario educational context, the descriptors focus on linguistic performances that are observable by teachers during curriculum learning tasks. In Table 1, the column containing "Elements" illustrates the connection between the curricular focus (or element) and the linguistic behaviours that may be observed over time as students learn content. Teachers' classroom-based tasks provide a context and a point of reference for observing and gathering evidence to assess learners' language proficiency development. For these reasons, the descriptors relate to and articulate the communicative demands of Ontario curriculum. Our validation research (Jang, Cummins, Wagner, Stille, & Dunlop, 2015; Jang, Wagner, & Stille, 2011) provided evidence that these descriptors act as generally distinct stages, indicating the stability of the scales in distinguishing among six proficiency levels. Additionally, the research demonstrated that average difficulty increases when moving from Step 1 to Step 6, and that the scales reflect the developmental nature of language and curriculum learning across grade clusters (e.g., Step 1–2 for Grade 1, Step 3–4 for Grade 2, etc.).

Language Proficiency Descriptor Scales in the Educational Context

Development of language proficiency descriptor scales emerged both from a need to describe what students can do at various levels of proficiency development, and from the standards-based movement in the United States wherein standards articulate expectations of what students should know and be able to do (Bailey & Huang, 2011). Descriptors-based language proficiency scales are part of a global assessment movement, which use these frameworks to assess students' language proficiency development. For example, the descriptors-based Common European Framework of Reference (CEFR) for Languages (Council of Europe, 2001) has been translated into many European languages and has been widely implemented in a variety of textbooks, cur-

Table 1

Sample Primary Level (Grades 1–3) Descriptors Across Six STEPs (Ontario Ministry of Education, 2012a)

Modality	Element	STEP 1	STEP 2	STEP 3	STEP 4	STEP 5	STEP 6
Reading & responding	Understand and respond to texts, using strategies	Demonstrate understanding by retelling information from a highly visual text by using a combination of visuals, L1, and English. Make connections by matching words and pictures Use concepts of English print (e.g., directionality of print, English alphabet, sound/symbol patterns, upper and lower case letters).	Demonstrate understanding by making predictions before, during, and after reading, using drawings, English words and phrases, and L1. Make connections by relating content in read-alouds to personal experiences, using drawings, English words and phrases, and L1.	Identify the main ideas Make connections to other texts and personal experiences, orally or in writing by responding to teacher prompts.	Make connections by expressing personal opinions about ideas presented in texts. Identify main ideas and supporting details. Identify point of view.	Demonstrate deeper understanding of complex texts by making inferences. Make connections by comparing and contrasting texts. Identify point of view with supporting details.	Demonstrate understanding by synthesizing information (e.g., combine information about healthy food to prepare a nutritious lunch). Make connections, using a variety of strategies and resources. Identify and contrast alternative points of view within a text.
Writing	Develop and organize content	Generate ideas by answering simple questions about personal experiences in L1 or English (e.g., using a word web, drawings, or illustrations).	Generate ideas by talking or brainstorming with peers and teachers in L1 or English (e.g., using an idea web, picture, and/or dual language dictionaries).	Generate ideas with peers using English and/or L1 by talking, brainstorming, and using graphic organizers (e.g., Think-Pair-Share).	Plan writing by considering questions related to purpose and audience (e.g., Why am I writing? Who am I writing for? What am I writing about?). Record key words to identify important information.	Gather information to develop ideas from a variety of sources (e.g., family, peers, age-appropriate print and multimedia resources).	Identify relevant information for a writing topic from multiple resources (e.g., print, multimedia resources, and websites).
Oral communication	Listen and respond for a variety of purposes	Respond to a personally relevant question with a single word or phrase in English or L1 (e.g., What's your name?). Follow simple instructions for classroom routines, using visual cues and pretaught English words and phrases.	Respond to simple questions in English or L1. Restate simple instructions. Identify key information from classroom discussions with teacher prompts.	Retell a read-aloud story or spoken text. Repeat multistep instructions (e.g., First finish the painting; then get a book; then sit on the carpet).	Retell main ideas from an oral narrative or video clip, using pre-taught key vocabulary.	Confirm understanding of teacher's lesson or an extended oral text (e.g., video) by completing a graphic organizer.	Listen to a class discussion and demonstrate understanding through active participation.

ricula, and examinations across Europe (Alderson, 2005). Other descriptors-based proficiency assessments include the Canadian Language Benchmarks (CLB), the American Council of the Teaching of Foreign Languages (ACTFL), and World-Class Instructional Design and Assessment (WIDA) English Language Proficiency (Jang et al., 2011). STEP is distinguished from these scales as it is used to track and assess school-aged learners' language proficiency development.

Much scholarly research has focused on development of descriptor scales, illustrating both the challenges in their development and the potential for these scales to describe language proficiency development in a way that is useful to teachers (Byrnes, 2008; Davison, 2007; McKay, 2000; North, 1993; North & Schneider, 1998; Scott, 2009; Scott & Erduran, 2004). Rather less attention has been paid to how these descriptors-based proficiency scales have influenced teacher instruction and assessment practice or to their usefulness for students. Several recent studies have focused on highlighting the diagnostic and formative purposes of descriptors-based proficiency scales for language teaching and learning (Brindley, 1998; Colby-Kelly & Turner, 2007; Davison, 2004; Hamp-Lyons, 2007; Rea-Dickens, 2004, 2007; Teasdale & Leung, 2000).

Language proficiency descriptor scales such as the CEFR (Council of Europe, 2001), the WIDA English Language Proficiency Standards and Assessing Comprehension and Communication in English State to State (ACCESS) assessment (Kenyon, MacGregor, Li, & Cook, 2011; WIDA, 2012), the ESL Standards for Pre-K-12 students in the United States (Teachers of English to Speakers of Other Languages, Inc., 2006), the ACCLES in Hong Kong (Davison, 2004), and the National Languages and Literacy Institute of Australia (NLLIA) ESL Bandscales (McKay, 2007; Scott & Erduran, 2004) have been used to assist teachers in understanding and interpreting learners' language proficiency development. The development and use of these scales have highlighted issues and challenges arising out of the intersection between language assessment and classroom curricula and pedagogy. In particular, research has drawn attention to issues in construct definition and the challenges in operationalizing language proficiency development in the context of content-area learning (Davison & Leung, 2009; Little, 2010; McKay, 2000; Scott, 2009). For instance, higher levels in second-language assessment scales tend to reflect learners' cognitive skills and educational experiences (Hulstijn, 2011), making it difficult for teachers to distinguish students' language proficiency development from higher-order thinking skills. Similarly, alignment with curriculum can make it difficult to reliably distinguish students' English language proficiency levels, and teachers may have difficulty distinguishing students' language proficiency development from their subject knowledge (Jang et al., 2015). Moreover, as social and socially situated activities, classroom-based assessments are not wholly reflective of individual cognitive processes, but also reflect social, affective, and academic circumstances and learners' in-

SASKIA VAN VIEGEN STILLE, EUNICE JANG, & MARYAM WAGNER

structional learning experiences. Language acquisition and language use in the classroom can therefore be seen as a unified process (Lantolf, 2009).

Along with increased use of descriptors-based language proficiency scales for assessment, language testing literature has called for research into classroom-based language assessment (CBLA) to explore the relationship between standardized testing and observation-driven assessments of language proficiency (Leung & Mohan, 2004; McNamara, 2001; Rea-Dickens, 2004; Shohamy, 2004). Unlike normative standardized tests, which measure learners' language proficiency development at one point in time against a predetermined population norm, CBLA takes place during everyday teaching and learning activities, allowing educators to evaluate learners' growth in language competencies over time and on multiple occasions, using diverse modes and forms of communicative interactions in the classroom. Examining and reflecting upon this evidence of learners' communicative performances, educators can then use language proficiency descriptor scales to judge learners' proficiency levels and identify areas of future learning. Situated within the classroom and embedded in educators' ongoing instructional practice, CBLA promotes authentic assessment in a naturally occurring language learning context (Brindley, 2001; Chalhoub-Deville, 2003; Scott, 2009; Shepard, 2002; Wigglesworth, 2008). This congruence between learning, teaching, and assessment corroborates a degree of ecological validity in that students are assessed in the way they have been taught and within the context of their language use (Whitehead, 2007).

Teachers' Use of Descriptors-Based Language Proficiency Scales

The use of descriptors-based proficiency scales in the educational context has led to an interest in challenges that teachers face in performing classroom-based language assessment. Studies exploring teachers' assessment practices and their use of proficiency scales have demonstrated teachers' needs associated with implementing effective language assessment in their classrooms, including the need for clear and interpretable proficiency scales (Davison & Leung, 2009; Llosa, 2011; Rea-Dickens, 2004). Research has also documented several challenges, including variability in teachers' assessments based on their views of assessment criteria (Butler, 2009), their perceptions of student motivation (Butler, 2009), their assessment literacy (Fulcher, 2012; Inbar-Lourie, 2013; Malone, 2013; Taylor, 2009, 2013), and the complex nature of the classroom context (Brindley, 1998). For instance, investigating teachers' language assessment practices, Davison (2004) described several of these challenges. First, assessment criteria are interpreted differently by teachers according to their personal background, previous experience, and expectations regarding the relative importance and meaning of language assessment criteria, causing teachers to differ from one another in their interpretation of, response to, and use of these criteria. Second, teacher-based assessment is not

a technical activity: it relies upon teachers' professional judgement and interpretation. Although descriptor scales may provide a basis for understanding learners' linguistic development, teachers understand and interpret these criteria in relation to their experience with real students and their linguistic performances during everyday classroom activities. Third, teachers' interpretation, negotiation, and discussion of their assessment decisions contribute to the validity and reliability of teachers' judgements. Finally, teachers may vary in the extent to which they will accept externally imposed criteria as a basis for their professional judgement about learners' language development. These multiple challenges are compounded by the diversity and variability of the contexts in which they are used.

Much of the use of descriptors-based proficiency scales depends upon the role of teachers. Teachers' roles and responsibilities often include, for instance, planning assessment activities, collecting samples of student work, and interpreting and making judgements about students' linguistic performances; monitoring, adapting, and modifying assessment depending on teaching and learning goals; and giving immediate and constructive feedback to students (Davison & Leung, 2009). Considering the interactionalist perspective that communicative language ability is the result of interactions between intrapersonal linguistic traits and the situational context (Chalhoub-Deville, 2003; Chapelle, 1998; Wiliam, 2010), the context in which language learning takes place is an important aspect in the use of descriptors-based proficiency scales. Teachers need to be able to create communicative contexts by designing tasks that elicit the specific observable behaviours from which meaningful inferences about students' language ability can be made. Teachers therefore require some degree of assessment literacy to support their effective use of descriptor scales.

Teachers' language assessment competence refers to the knowledge, skills, and abilities that teachers need to implement language assessment activities and interpret students' language skills and development (Fulcher, 2012). These competences are important, as teachers' interpretations and decisions about students' language abilities may have impact beyond the confines of the classroom (Edelenbos & Kubanek-German, 2004; Jang et al., 2015). For example, policy-related decisions such as the necessity to offer accommodations during standardized testing, or even whether it is appropriate for some language learners to write standardized assessments (particularly during the initial phase of their language acquisition), may be determined by teachers' interpretations of students' language abilities. Other language-related decisions such as provision of supports or advancement may also be influenced or determined by teachers' evaluation. While CBLA promotes teachers' agency in assessment in general, it recognizes a great need to develop teachers' assessment competencies. These concerns point to the role that descriptors-based proficiency scales might play in supporting teachers' development of assessment competencies, potentially increasing teachers' knowledge of and

SASKIA VAN VIEGEN STILLE, EUNICE JANG, & MARYAM WAGNER

communication about language proficiency development (Cummins et al., 2009; Jones & Saville, 2005), and assisting with teachers' professional judgements about learners' language development needs (Cumming, 2009; Davison & Leung, 2009).

Research Methods

Our research team was commissioned by the Ontario Ministry of Education to conduct a validation study of the STEP proficiency scales, beginning in September 2008 and continuing to September 2011. The objectives of the study were to (a) empirically verify the STEP scales by collecting samples of student performance that exemplify the levels of proficiency on the STEP scales, (b) examine the linguistic and cultural sensitivity of the STEP descriptors, and (c) evaluate the impact of the STEP assessment on teaching and learning.

We gathered multiple types of data that documented teachers' use of the STEP scales from 42 ESL and classroom teachers and 159 students across three school districts.[2] Table 2 displays the distribution of the teachers across the different grade clusters. Most of these teachers (87%) had 6 or more years of teaching experience. The participating teachers were uniformly distributed among subject specialists, mainstream teachers, and ESL teachers. Teachers were required to observe and sample students' language performances during student learning activities, and interpret and evaluate these performances to make decisions about students' language proficiency along the STEP continuum in each of the three language modalities. Specifically, teachers tracked students' mastery of described observable linguistic behaviours as described in the STEP continua for the Oral Communication, Reading and Responding, and Writing modalities. This decision-making process was based on the evidence that teachers gathered about students' language learning over several weeks of instructional time, which were based on teachers' informal observations, discussions with students, and planned assessment activities (e.g.,

Table 2
Distribution of Participating Teachers Across Grade Clusters

Grade cluster	Number of teachers
Primary	19
Junior	7
Intermediate	7
Secondary	8
Grade not specified	1
Total	42

writing assignments, presentations). This article focuses on the impact of the STEP scales on teaching practice by drawing on the various types of data gathered during this period of teachers' evidence-gathering and decision-making about students' linguistic behaviours. The data sources included classroom observations, teacher and student interviews, samples of students' written work, and digital video documentation of teachers' assessment activities using STEP. Herein, we describe the process of collecting these data and their analyses in more detail.

During our classroom visits, teachers were encouraged to showcase the specific needs and interests of their students as well as their classroom practices. Classroom observations were conducted using a preconstructed protocol in order to observe teachers' classroom and feedback practices. In particular, we were interested in observing the facilitation of classroom-based tasks and activities and the extent to which (a) teachers scaffolded learning, (b) opportunities existed for teacher-student interactions and provision of feedback to learners, (c) language and literacy behaviours could be observed across different modalities, and (d) teachers employed strategies to assist and reinforce language learning. These activities were documented using field notes, as well as audio and video recordings. All of the observational data were transcribed and coded (using NVivo version 9) according to these four categories. We further coded the data thematically using both inductive (to seek new themes) and deductive (to identify more specific features of the aforementioned categories) approaches.

Teachers were interviewed in order to gather more information about their classroom practices, their processes of gathering evidence about students' language learning and development, and their post-STEP reflections. We aimed to understand how STEP influenced these educators and asked specific questions about their teaching practices, understanding of ELL's language development, assessment practices, and challenges they may have encountered during the process of its use. These data were similarly transcribed and thematically analyzed.

Students' work samples were gathered and used to further understand how teachers interpreted the evidence of students' performance, and the indicators they used to make decisions about students' mastery of STEP observable linguistic behaviours. The samples were also used to understand teachers' feedback practices and the ways in which they identified gaps in students' learning and guided learners towards addressing them. These samples included work across each of the three STEP modalities, and were matched to students' STEP profiles and background information about the students.

The multiple data from teachers and classrooms across different grade levels provided the opportunity to elaborate and complement the results of one set of data analyses with those of another. Consequently, the process of data analysis was recurrent and iterative; we revisited codes and themes

SASKIA VAN VIEGEN STILLE, EUNICE JANG, & MARYAM WAGNER

across all of the data sources to revise, reanalyze, and further augment our understanding and interpretation of the results. Accordingly, we were able to synthesize our findings across the different data sources and types

Results

The analyses of the multiple types of data allowed us to generate findings about the use of STEP and its impact on teaching activities, including teachers' understanding of students' language development, teachers' understanding of the role of instruction and feedback in language learning, and teachers' use of the scales to support diagnostic and formative purposes of assessment. These data also highlighted challenges in classroom-based language assessment relating to the interdependence of language and literacy development in the context of school-based learning. We now report these findings in detail, drawing from a synthesis of the results from the different data sources and analyses.

Facilitating Formative Language Assessment

One of the major findings emerging from the data analyses was that the STEP scales operated as an overarching language assessment framework that supported curricular planning, instructional activities, and assessment of and feedback to students. The findings suggested that teachers' use of the STEP scales facilitated ongoing, informal assessments of students' language progress. Teachers used their judgements about students' progress to inform their teaching and assessment strategies, and to communicate with students about their language development. These formative and summative uses of STEP may have contributed to an assessment for learning culture within the classroom, helping teachers to better understand how to improve students' language learning and give learners constructive feedback (see, e.g., Biggs, 1998; Carless, 2007; Davison, 2007; Hamp-Lyons, 2007; Harlen, 2005; Taras, 2005).

A major theme that emerged from the analyses of the data across the multiple sources was that teachers perceived that the proficiency scales could be used for diagnostic and formative purposes. Using their own observational evidence as well as students' performance on learning tasks that were already a part of their instructional repertoire, experienced teachers who participated in the study reported that they were able to elicit sufficient evidence of students' linguistic performances to make judgements about students' proficiency level. Participating teachers also described their ability to use the STEP scales to describe learners' language development based on learners' linguistic performances that they observed in their classrooms. Therefore, it can be surmised that the descriptor scales helped teachers to identify students' initial language proficiency level, and each level provided teachers with some indication of the gradually expanding scope of the learners' strategic range and underlying linguistic competence. As teachers explained, this informa-

tion helped them to recognize and differentiate learners' individual language learning needs.

Teachers reported that their use of the STEP assessment tool involved gathering evidence of students' linguistic performances within the context of everyday curriculum learning activities. This evidence included anecdotal notes from observations of students' participation in class, as well as other classroom-based evidence such as student portfolios, samples of students' work, and teacher-developed tests. Reflecting upon this evidence of students' linguistic performances, participating teachers described that they could use the proficiency scales to help them select future learning goals for students and plan their instruction accordingly. Participating teachers described their interest in using the STEP scales to help them know each learner and better understand learners' language proficiency development. The following quotation from one teacher illustrates this point:

> When I look at his work, at his work folder, I identified him as being STEP 2, and [the descriptors] are very specific in terms of what he should have accomplished at STEP 2 … There is a checklist and when I flip over to STEP 3, it's nice to see the things that he needs to be working on, it's a specific set of things that we need to do. (Elementary ESL teacher)

The results also supported the finding that the STEP scales drew teachers' attention to the wide range of learning activities implicitly embedded in the descriptors, which supported STEP's formative purposes to inform instructional planning and practice. The alignment of the proficiency scales with curriculum, the comprehensive nature of the scales, and the incremental progression of language development in the STEP scales contributed to teachers' ability to integrate STEP with their instructional practice. Describing what learners can do at each level, the STEP scales articulated the linguistic performances students should be able to do to demonstrate progress. Teachers reported using this information to identify teaching strategies to help learners progress, including differentiating instruction, planning instructional strategies, and designing future curriculum-related and classroom-based assessment tasks. Another quotation from a teacher exemplifies this finding:

> STEP helped me choose a more appropriate form of support for the students and the classroom teacher. I found the descriptors helpful to give me direction in the sort of strategies I might use to help the children progress. (Elementary ESL teacher)

Participating teachers also articulated the ways in which they used the STEP scales to guide their planning and design of assessment activities. Evidence gathered during the teacher interviews highlighted how teachers chose assessment tasks that would enable them to observe specific descriptors. Using STEP, teachers reflected upon their instructional practice and whether

and how it provides sufficient opportunities for both language learning and assessment of learners' language proficiency development. The use of the STEP framework also prompted teachers to identify opportunities for assessment embedded within curricular activities, and to assess the degree to which these activities incorporated domains of language development. For instance, teachers reported the need to gather multiple sources of evidence as the basis for their language assessment activities. Gathering a diverse range of student work samples reflects the idea that an individual learner may exhibit a range of proficiencies in language use depending on the context or learning task. Similarly, the use of a variety of classroom activities for language assessment tasks provided learners with several opportunities to display their linguistic and pragmatic competencies, and provided teachers with both holistic and nuanced understandings of the individual learner and his or her language development. The following quotation from one teacher illustrates these concerns:

> This is only one sample, one piece of evidence, so when you're looking at STEP, we need to have multiple pieces of evidence of something before we can see that they've really done it, sometimes kids miss something, like I think she missed it [in this instance], so I would revisit it with her, talk about it with her a bit more but not focus on this evidence too much, but focus on all the other pieces of evidence that I have from her, because I don't think that this is a fair piece to show her reading comprehension, to say OK, this is your summative and I'm using this and only this. (Intermediate ESL teacher)

The data revealed that teachers used scaffolding and various accommodations to differentiate their STEP assessment activities and support individual learners' participation and learning in the context of the assessment process. Among the student work that we gathered, many samples included evidence of teachers' efforts to scaffold students' production of culminating curricular activities. For example, we gathered evidence of brainstorming activities, jot-notes, drafts of written work, and final copies of written work. These kinds of scaffolding strategies built students' background knowledge, key concepts, and vocabulary. For assessment of reading using STEP, many teachers provided samples of graphic organizers they used to provide scaffolding to improve learners' comprehension of a selected text. Similarly, evidence from video recordings of teachers' assessment activities demonstrated the use of strategies such as explicit modelling by the teacher, opportunities for practice with feedback, and skillful adjustments to accommodate the learner's oral proficiency level. These strategies contributed to students working in the metaphorical zone of proximal development (Vygosky, 1978) during assessment and helped the learner to perform at a slightly higher level than they could have done independently.

Participating teachers described the ways in which they use assessment information to communicate with students about their language learning. Teachers used the descriptors in the STEP scales to explain to learners what they had achieved and what they needed to focus on next. Using the descriptors to formulate specific feedback to students, teachers assisted students in reviewing and reflecting on their language proficiency development. Teachers described their perception that this information would be useful to help learners monitor their own language learning progress. This is illustrated by the following quotation from one teacher:

> It [STEP] is really good even for my students who aren't English language learners because it's a reminder to have checkpoints, to have goals for the kids so they can see "Oh, I really am learning this," it helps them see where their learning needs to go. (Intermediate teacher)

However, from observation and interview data, we found that teachers' feedback to students during the STEP implementation was minimal, though feedback is generally seen as a key function of formative assessment, with the goal of improving student learning (Davison & Leung, 2009). This gap points to teachers' professional learning needs concerning how to use STEP as a checkpoint, in addition to their formative feedback, to support instructional and assessment practices in the classroom.

Supporting Professional Learning

A second major theme emerging from the data analyses was related to potential of STEP as a tool to support professional learning. As teachers reported, the use of the STEP scales provided them with a shared language in which they could recognize and discuss learners' language use and development with their students and other teachers. By providing teachers with a common language and framework of reference, the scales increased the extent of meaningful collaboration and communication among teachers working with ELLs, particularly between ESL and mainstream or subject-area teachers. Teacher collaboration and sharing of expertise occurred as teachers discussed their planning, assessment activities, and decision-making with other teachers, which is illustrated by the following quotation from one ESL teacher:

> The classroom teacher and I have collaborated … we have compared observations, we have shared concerns, and I think that has been a great foundation for a good partnership between the two of us. It worked out very well. (Elementary ESL teacher)

Accordingly, it can be suggested that teachers' use of the STEP scales contributed to their professional learning, building their knowledge based about the relationship among language learning, instruction, and assessment. ESL and classroom teachers who used STEP reported finding it useful to reflect upon

SASKIA VAN VIEGEN STILLE, EUNICE JANG, & MARYAM WAGNER

the strategies they used for meeting the needs of ELLs in their classrooms. These positive outcomes are described in the following illustrative quotations:

> For my perspective, a classroom teacher, STEP was almost self-assessment [of my teaching. (Elementary teacher)

> STEP opens your eyes to … where students are going and keeps a perspective that ESL students are not necessarily going to progress as fast in other areas as the regular class will go, and that is OK … as a classroom teacher it brought to my perspective that it may take a year for them to get to the next level, it may take two years to get to that level, and that is OK. (Elementary teacher)

Teachers reported using evidence gathered during the STEP assessment to support their decisions relating to students' needs at school. For example, teachers reported that STEP assessment activities assisted them in collecting samples of student work and making interpretations of this work which could be useful for (a) making decisions about when to provide or stop providing ESL support; (b) tracking students' language progress, particularly as they move from one school to another; and (c) understanding which students are ready to write standardized literacy tests such as the Ontario Secondary School Literacy Test (OSSLT), and how these students' performance on the provincial tests should be interpreted. This quotation was typical of teachers' responses:

> We had a meeting to decide whether Luis and another student in the ESL program would be participating in [the standardized provincial literacy test] and I relied on the STEP assessment tools to make a case that he was in fact ready for it, that his English skills have been built sufficiently so that he could participate. (Secondary ESL teacher)

Teachers reported that they experienced difficulty in distinguishing between reading problems stemming from low levels of linguistic proficiency versus more general reading/learning difficulties. With enriched understanding of the stages of language development, teachers explained that they saw potential for STEP assessment data to clarify distinctions between language acquisition and the learning processes relating to content area curriculum. For example, tracking the STEP progress of one student over the 18 months, a teacher noted that the student's progress had stalled beyond what she would consider normal—he had moved up to the next STEP level on only 1 of approximately 12 descriptors in his oral development, although his reading development had progressed. The teacher described this circumstance: "If you see a student that has been here for two years, and they are still at STEP 1, then … [you know] there is something happening here." Teachers explained that STEP provides useful evidence of exceptionalities where corroborating

evidence exists, explaining, "You cannot go into these meetings with how you feel or what you think. You've got to have concrete evidence."

Understanding Issues Related to Assessing and Tracking Language Learning

A third theme that emerged from the research investigation was teachers' increased awareness of issues related to assessing and tracking students' learning in the classroom. Participating teachers reported on their awareness of situational factors that influence language development and interpretation of language development. These factors included students' prior learning experiences, family history, socioeconomic background, and cultural differences. In the context of implementing STEP, teachers wondered whether they were assessing speaking, cognition, personality, or cultural knowledge. For example, one teacher explained how he takes a student's background into consideration when conducting a STEP assessment:

> We would look at that [descriptor] and say, is it unfair to ask that of this student? She would be unable to answer and I would have to put her on STEP 1. I did not think the [descriptor] was fair to that child. We struggled with that. (Elementary ESL teacher)

Teachers noted that the cultural fairness of classroom contexts had a noticeable influence upon the assessment of a learner's language development. For example, cultural differences in oral communication styles might influence a learner's classroom behaviours and linguistic performances that are used for STEP assessment. Furthermore, learners need support to bridge the cultural and interactional patterns from those of their home culture to ones that are functional for learning in the Canadian context. The following quotation from one teacher exemplifies some of these concerns:

> We have to understand the connections [students are] making between stepping out of their first language and into the second language … there are all kinds of things that they need to do to get to that transition. Depending on the culture, students learn and connect in different ways. So making those connections into English depends on how … they integrate themselves into a different community apart from their culture, more experiences will happen, they will learn from that … and it will be easier for them to move on oral, reading, and writing as well. (Elementary ESL teacher)

Assessing Language Ability in Content Learning

A final key finding that emerged from the data analyses was related to the opportunities afforded by STEP for assessing students' language development alongside authentic content-based contexts. By prioritizing the K-12 learning context, STEP aligns the development of language ability with On-

SASKIA VAN VIEGEN STILLE, EUNICE JANG, & MARYAM WAGNER

tario curricula and embeds assessment into curriculum learning. While the data have revealed that the use of STEP for describing learners' language abilities within the context of curriculum learning has contributed to facilitating collaboration between ESL and classroom teachers, it was illustrated to be challenging for classroom teachers in distinguishing between literacy development and language proficiency, partly because the curriculum is not sufficiently detailed to the point of defining what language ability means (McKay, 2006). Accordingly, teachers felt that the STEP descriptors were too tightly aligned with curriculum, and expressed their concern that the STEP scales may be limited to assessing literacy development rather than English language proficiency.

Aligning the development of English proficiency with curricular expectations involves articulating levels of language progression based on actual learner performance in schools (Byrnes, 2008; North, 2007). The challenges we observed are not uncommon, as similar concerns have been raised in discussions of the CEFR, whose descriptors were developed by expert teachers (North, 1993; North & Schneider, 1998). If the construct of "English language development" remains indistinguishable from curricular literacy learning, embedding language proficiency scales into specific content-learning contexts may not be fully achieved. Such challenges emphasize the importance of collaborations between ESL and classroom teachers as well as assessment specialists and policy makers.

Discussion

The findings that emerged from this research illuminate how descriptors-based language proficiency scales can contribute to the formative purposes of classroom-based language assessment in the K-12 educational context. The use of the STEP scales promoted an assessment for learning culture among participating teachers (Black & Wiliam, 1998, 2008; Cumming, 2009; Davison & Leung, 2009; Gardner, Wynne, Hayward, & Stobart, 2008), increasing their knowledge about language proficiency development, assisting with their professional judgements (Cumming, 2009; Davison & Leung, 2009), increasing communication, and promoting teachers' collaboration with peers. We have reported elsewhere on teachers' critiques of the STEP framework (Jang et al., 2011), which include issues relating to teachers' time, professional learning needs, and availability of resources. Despite these issues, teachers' perceptions about the use and impact of STEP suggest that descriptor scales have the potential to be used alongside existing classroom assessment activities to assist with teachers' understandings about learners' capabilities and needs, and to develop teachers' knowledge, skills, and abilities in classroom-based language assessment. In Ontario, these developments are critical because curriculum policy articulates that all teachers, not just those designated as ESL teachers, are required to identify the needs and instructional strategies

necessary to support the language and learning of ELLs in their classrooms, and all teachers are required to record students' language progression in official school records. Teachers are often required to meet these objectives with sometimes little background knowledge of issues in language learning and assessment. Thus, district-level implementation and teachers' use of the STEP scales can potentially assist in building province-wide capacity to meet these policy requirements.

Relating to these policy requirements, we can identify several challenges concerning how to assist teachers in integrating the use of language proficiency scales into their instructional practice. For example, teachers need to understand and differentiate among students' linguistic and pragmatic competencies, and their overall cognitive and social development and content area knowledge. Connecting the proficiency scales to language use using evidence from everyday curriculum learning activities contributes to an appropriate and representative form of observation of students' linguistic performance and process of language acquisition. However, reflecting the interdependence of language, literacy, and content curriculum learning, evaluating students' language use and performances in the classroom means that assessing language includes assessing content to some degree. Further, teachers' understandings of these aspects of language proficiency development in the learning context relate to their ability to select and use appropriate tasks for assessing language proficiency development.

Initially, the articulated purpose of the STEP proficiency scales was to identify, monitor, and track the progress of English language learners in Ontario schools. Over time, the Ministry recognized that teachers' use of STEP could serve multiple purposes, including directing teachers' instructional goals and activities, guiding formative purposes of language assessment, supporting teachers' professional learning, and building system-wide capacity for supporting ELLs. To meet these multiple purposes of STEP in a manner consistent with the scholarly literature and evidence-based practice in classroom-based language assessment, we propose a framework for the development of teacher language assessment competence that highlights the kinds of knowledge and skills that teachers need to use descriptors-based language proficiency scales effectively in their classrooms. This framework is outlined below:

- **Assessment tasks**. Teachers' assessment activities should be sufficiently rich to support the learners in showcasing the full range of their communicative and linguistic abilities at, or slightly above, their level of competence.
- **Instructional strategies to scaffold assessment**. Assessment activities should be scaffolded with appropriate instructional strategies to assist the learners in accomplishing the task at, or slightly above, their level of competence; teachers' evaluation of this performance can account for

SASKIA VAN VIEGEN STILLE, EUNICE JANG, & MARYAM WAGNER

this scaffolding, based on the assumption that students progress from accomplishing linguistic performances with assistance at the lower levels of proficiency and move to independent accomplishments at the higher levels of proficiency.

- **Observational abilities**. Teachers need the ability to notice learners' linguistic and communicative performances during everyday classroom interactions, and need to gather observational data about these performances; this includes recording observational notes about the learner and gathering samples of his or her work to reflect on, in a manner that is sufficient to assist the teacher in making an appropriate judgement about the learner's language performance.
- **Interpretive abilities**. Teachers need the ability to interpret students' linguistic and communicative performances on everyday teaching and learning activities; they need the ability to distinguish among a learner's language development, his or her sociocultural competence, and his or her curriculum learning.
- **Use**. Teachers need the ability to use the proficiency scales to inform their instructional practice; for instance, using descriptors to guide instruction, to define what students need to know, to periodically gauge the learner's understanding, and to give the learner descriptive feedback to help him or her reach those goals.

Conclusion

Educators play a critical role in supporting the academic success of English language learners in Ontario schools, and these students have unique language and learning needs. The research reported here suggests that effective use of descriptors-based language proficiency scales for classroom-based language assessment can potentially promote professional learning for educators and build capacity for schools to meet the needs of ELLs. This research contributes to the articulation of classroom-based language assessment practices, drawing on the perspective of teachers to provide a nuanced perspective of relationship between language, language learning, and language assessment in the K-12 educational context. Documenting teachers' perceptions about the use and impact of descriptors-based language proficiency scales contributes evidence that can be drawn upon to theorize and operationalize the constructs underlying the processes of classroom-based language learning, and the assessment practices that measure these.

The Author(s)

Dr. Saskia Van Viegen Stille is an Assistant Professor in the Faculty of Education, Centre for English Language Learning and Teaching Research, at Simon Fraser University. Publications from her recent work appear in several edited books and journals, including *Language Assessment Quarterly*, *European Journal of Teacher Education*, and *TESOL Quarterly.*

Dr. Eunice Eunhee Jang is Associate Professor in the Department of Applied Psychology and Human Development in Ontario Institute for Studies in Education, University of Toronto. Dr. Jang's current research program is best situated in the areas of educational assessment in general, with an emphasis on language assessment of and for second language learners in particular.

Dr. Maryam Wagner is a SSHRC postdoctoral research fellow at Carleton University. Her research focuses on language assessment with specific interest in feedback, validity, second language writing, cognitive skill diagnosis, and the relationship between students' motivation and assessment use.

Notes

1   English for Literacy Development (ELD) is a term used in Ontario education to refer to English language learners with limited prior schooling, circumstances that have led to gaps in these students' academic literacy and numeracy. The term ELD is also used in Ontario education to refer to programs specifically designed to address these unique learning needs, including not only English language teaching, but also development of numeracy concepts and literacy skills.
2   The study focused on the STEP scales for ESL only, not ELD, because the ELD descriptors were still in the process of development at the time the study began.

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* London, UK: Continuum.

Gardner, J., Wynne, H., Hayward, L. & Stobart, G. (2008). *Changing assessment practice: Process, principles, and standards.* London, UK: Assessment Reform Group.

Auditor General of Ontario. (2005). *2005 annual report*. Toronto, ON: Author.

Bailey, A. L., & Huang, B. H. (2011). Do current English language development/proficiency standards reflect the English needed for success in school? *Language Testing*, *28*(3), 343–365.

Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice, 5*(1), 103–110. doi:10.1080/0969595980050106

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7–74.

Black, P., & Wiliam, D. (2008). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. doi:10.1007/s11092-008-9068-5

Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing, 15*, 45–85. doi:10.1177/026553229801500103

Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing*, *18*(4), 393–407. doi:10.1177/026553220101800405

Butler, Y. G. (2009). How do teachers observe and evaluate elementary school students' foreign language performance? A case study from South Korea. *TESOL Quarterly*, *43*(3), 417–444. doi:10.1002/j.1545-7249.2009.tb00243.x

Byrnes, H. (2008). Assessing content and language. In E. Shohamy & N. H. Hornberger (Vol. Eds.), *Encyclopedia of language and education, Vol. 7. Language testing and assessment* (pp. 37–52). New York, NY: Springer.

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57–66. doi:10.1080/07294360.2010.502218

Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language testing, 20*(4), 369–383. doi:10.1191/0265532203lt264oa

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32-–70). Cambridge, UK: Cambridge University Press.

Colby-Kelly, C., & Turner, C. E. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *Canadian Modern Language Review, 64*(1), 9–37. http://dx.doi.org/10.3138/cmlr.64.1.009

Council of Europe. (2001). *Common European reference framework for languages*. Strasbourg, France: Author. Retrieved from http://www.coe.int/T/DG4/Portfolio/?L=E&M=/documents_intro/common_framework.html

Cumming, A. (2009). What needs to be developed to facilitate classroom-based assessment? *TESOL Quarterly, 43*(3), 515–519.

Cummins, J., Jang, E. E., Stille, S., Wagner, M., Byrd Clark, J., & Trahey, M. (2009). *Steps to English Proficiency (STEP): Validation study* (Final research report presented to the Ministry of Education). Toronto, ON: Modern Language Centre, OISE.

Davison, C. (2004). The contradictory culture of classroom-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*(3), 305–334. doi:10.1191/0265532204lt286oa

Davison, C. (2007). Views from the chalkface: School-based assessment in Hong Kong. *Language Assessment Quarterly, 4*, 37–68. doi:10.1080/15434300701348359

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393–415. doi:10.1002/j.1545-7249.2009.tb00242.x

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of "diagnostic competence." *Language Testing, 21*, 259–283. doi:10.1191/0265532204lt284oa

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(1), 113–132. doi:10.1080/15434303.2011.642041

Gallagher, M. J. (2014). *Ontario: Reaching for equity and excellence in education*. Presentation at Comparative and International Education Society Conference, Toronto, ON.

Gardner, J., Wynne, H., Hayward, L. & Stobart, G. (2008). *Changing assessment practice: Process, principles, and standards.* London, UK: Assessment Reform Group.

Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *The international handbook of English language teaching* (Vol. 1, pp. 487–504). Norwell, MA: Springer.

Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *Curriculum Journal, 16*(2), 207–223. doi:10.1080/09585170500136093

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly, 8*(3), 229–249. doi:10.1080/15434303.2011.565844

Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing, 30*(3), 301–307. doi:10.1177/0265532213480126

Jang, E. E., Cummins, J., Wagner, M., Stille, S., & Dunlop, M. (2015). Investigating the homogeneity and distinguishability of STEP proficiency descriptors in assessing English language learners in Ontario schools. *Language Assessment Quarterly. 12*(1), 87–109.

Jang, E. E., Wagner, M., & Stille, S. (2011). Issues and challenges in using English proficiency descriptor scales for assessing school-aged English language learners. *Cambridge ESOL Research Notes*, 45, 8–13.

Jones, N., & Saville, N. (2009). European language policy: Assessment, learning and the CEFR. *Annual Review of Applied Linguistics, 29*, 51–63. doi:10.1017/S0267190509090059.

Kenyon, D. M., MacGregor, D., Li, D., & Cook, H. G. (2011). Issues in vertical scaling of a K-12 English language proficiency test. *Language Testing, 28*(3), 383–400. doi:10.1177/0265532211404190

Lantolf, J. P. (2009). Dynamic assessment: The dialectic integration of instruction and assessment. *Language Teaching, 42*(3), 355–368. doi:10.1017/S0261444808005569.

Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing, 21*(3), 335–359. doi:10.1191/0265532204lt287oa

Little, D. (2010). *The Common European Framework of References for Languages: A research agenda.* Paper presented at OISE/University of Toronto, Toronto, ON.

Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research. *Language Testing, 28*(3), 367–382. doi:10.1177/0265532211404188

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing, 30*(3), 329–344. doi:10.1177/0265532213480129

McKay, P. (2000). On ESL standards for school-age learners. *Language Testing, 17*(2), 185–214. doi:10.1177/026553220001700205

McKay, P. (2006). *Assessing young language learners.* New York, NY: Cambridge University Press.

McKay, P. (Ed.) (2007). *Assessing, monitoring and understanding English as a Second Language in schools: The NLLIA ESL Bandscales Version 2.* Brisbane, Australia: Queensland University of Technology and Independent Schools Queensland.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing, 18*(4), 333–349. doi:10.1177/026553220101800402

North, B. (1993). *The development of descriptors on scales of language proficiency.* Washington, DC: National Foreign Language Center.

North, B. (2007). The CEFR illustrative descriptor scales. *Modern Language Journal, 91*(4), 656–659.

North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing, 15*(2), 217–262. doi:10.1177/026553220101800204

Ontario Ministry of Education. (2007). English language learners: ESL and ELD programs and services. In *Policies and procedures for Ontario elementary and secondary schools, Kindergarten to Grade 12.* Toronto, ON: Queen's Printer for Ontario.

Ontario Ministry of Education. (2012). *English as a Second Language Observable Language Continua.* Toronto, ON: Queen's Printer for Ontario.

Rea-Dickens, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21*(3), 249–258.

Rea-Dickens, P. (2007). Classroom-based assessment: Possibilities and pitfalls. In J. Cummins & C. Davison (Eds.), *The international handbook of English language teaching* (Vol. 1, pp. 505–520). Norwell, MA: Springer.

Scott, C. (2009). Issues in the development of a descriptor framework for classroom-based teacher assessment of English as an Additional Language. *TESOL Quarterly,* 530–535. doi:10.1002/j.1545-7249.2009.tb00250.x

Scott, C., & Erduran, S. (2004) Learning from International Frameworks for Assessment: EAL descriptors in Australia and the USA. *Language Testing, 21*(3), 409–431. doi:10.1191/0265532204lt290oa

Shepard, L. (2002). Standardized tests and high-stakes assessment. In J. Guthrie (Ed.), *Encyclopedia of education* (Vol. 6, 2nd ed., pp. 2533–2537). New York, NY: Macmillan Reference.

Shohamy, E. (2004). Assessment in multicultural societies: Applying democratic principles and practices to language testing. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 72–92). Cambridge, UK: Cambridge University Press.

Statistics Canada. (2014). *Population by year, province and territory.* Retrieved from http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo02d-eng.htm

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–478. doi:10.1111/j.1467-8527.2005.00307.x

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics, 29*, 21–36. doi:10.1017/S0267190509090035

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing, 30*(3), 403–412.

Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing, 17*(2), 163–184. doi:10.1177/026553220001700204

Teachers of English to Speakers of Other Languages, Inc. (2006). *TESOL Pre-K–12 English language proficiency standards framework*. Alexandria, VA: Author.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Whitehead, D. (2007). Literacy assessment practices: Moving from standardised to ecologically valid assessments in secondary schools. *Language and Education, 21*(5), 434–452. doi:10.2167/le801.0

Wigglesworth, G. (2008). Task and performance based assessment. In E. Shohamy & N. H. Hornberger (Vol. Ed.), *Encyclopedia of language and education, Vol. 7: Language testing and assessment* (pp. 111–122). New York, NY: Springer.

Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education, 34,* 254–284. doi:10.3102/0091732X09351544

World-Class Instructional Design and Assessment (WIDA). (2012). *ACCESS for ELLs*. Retrieved from http://www.wida.us/assessment/access/