Self and Peer Assessment in K-12 Chinese Language Classrooms: Teachers' Perceptions and Implementation

Xuexue Yang[a], Yanjiang Teng[b], Zheng Gu[c], Dongbo Zhang*[d]

Article Info

Abstract

Little research has been conducted on self and peer assessment (hereafter, SA and PA) in Chinese language education in the United States, despite the fact that both forms of assessment have multiple benefits for language learning and it is argued they should be an integral component of language teachers' repertoire of classroom assessment. Particularly, little is known about Chinese language teachers' perceptions of SA and PA and their classroom SA and PA practices. To address this knowledge gap, this study aimed at exploring Chinese language teachers' perceptions and implementation of SA and PA and establishing a ground for future research in this field. Using a qualitative methodology, we interviewed ten K–12 Chinese language teachers in US schools. By examining the interview responses of those teachers, this study revealed some important findings. Notably, all the teachers realized the benefits of SA and PA for student learning. Yet, there was rare use of them in most teachers' classrooms. We discussed this gap by drawing upon the probed responses of the teachers as well as our inferences. The findings of this study inform SA and PA of Chinese language in K–12 classrooms and professional development for Chinese language teachers in US schools.

[a] Ph.D candidate, Michigan State University, Michigan, U.S.A
[b] Ph.D. Michigan State University, Michigan, U.S.A
[c] Ph.D. candidate, Michigan State University, Michigan U.S.A
*[d] Corresponding author: Prof., University of Exeter, Exeter, U.K. D.Zhang4@exeter.ac.uk

Introduction

Among the many types of classroom assessment that have a strong formative purpose, self and peer assessment (hereafter, SA and PA), which offer opportunities for students to assess themselves and their fellow students, respectively, have attracted increasing attention from educational researchers and practitioners, including language teachers (Blanche & Merino 1989; Butler, 2016; De Saint Léger, 2009; Dochy, Segers, & Sluijsmans, 1999; Edwards, 2014; Panadero, Brown, & Courtney, 2014; Oscarson, 2014). It is argued that by letting learners self-assess or assess others, they can become better aware of and reflexive on the learning process, and eventually become self-regulated learners (Sambell & McDowell, 1998; Sluijsmans & Prins, 2006). In other words, through SA and PA, learners are given an opportunity to take ownership of their learning and be engaged actively in the learning process (Blanche & Merino 1989; Butler, 2016; Edwards, 2014; Oscarson, 1989, 2014).

In English as a Second/Foreign Language (ESL/EFL) assessment, while the benefits of SA and PA have been recognized (e.g., Butler, 2016; Butler & Lee, 2010; Hung, Samuelson, & Chen, 2016; Matsuno, 2009; Patri, 2002), and it is often suggested that teachers incorporate them into their repertoire of classroom-based assessment (Blanche & Merino, 1989; Butler, 2016), these two types of alternative assessment are often not commonly used (Edwards, 2014; Oscarson, 2014). Many factors could come into play in impacting the use, or the nonuse, of PA and SA in L2 classrooms, an important one being teachers' beliefs and perceptions, which have been found to have a close relationship with what teachers do, or do not do, in the classroom (Borg, 2006; Dixon, Hawe, & Parr, 2011; Harris & Brown, 2013; Pajares, 1992).

In foreign language education in the United States, including Chinese language education, attention to SA and PA is limited, despite sporadic discussion on the benefits of them and a need that teachers need to be empowered to conduct them – and other forms of formative assessment – to facilitate learners' metacognitive processes of learning and language development. In addition, the limited research on SA and PA was largely from the perspective of learners and learning (e.g., Brown, Dewey, & Cox, 2014; De Saint Léger, 2009; Wang, 2017); few studies addressed SA and PA from the teacher's perspective. To this end, drawing upon the data collected for a study that investigated the assessment literacy of K–12 World Language teachers in the United States, this paper aims to initiate an exploration on Chinese language teachers' perceptions and use of SA and PA, if any, of these two forms of student-centered assessment in their classrooms.

Self-Assessment (SA) and Student Learning

SA, as the name suggests, refers to students assessing themselves with respect to what they know and can do. Through direct involvement in assessment processes, students can develop a sense of responsibility for their learning (Sambell & McDowell, 1998; Sluijsmans & Prins, 2006). It is argued that SA can enhance students' self-regulated learning (Ross, 2006; Oscarson, 2014), or the "autonomy and control by individual who monitors, directs, and regulates actions toward goals of information acquisition, expanding expertise, and self-improvement" (Paris & Paris, 2001, p. 89). According to Ross (2006), student self-regulated learning through SA embodies three processes: students producing self-observation by focusing on specific aspects of their performance; students making judgments about how well their general and specific goals were met; as well as students' self-reactions and interpretations of the degree of goal achievements.

Given the importance of self-regulated learning in students' academic achievements (Zimmerman, 1990), a relationship between conducting SA and students' academic achievement and improvement have been identified (Andrade & Boulay, 2003; Andrade & Valtcheva, 2009; Baleghizadeh & Masoun, 2014;

Butler, 2016; Wan-a-rom, 2010). Baleghizadeh and Masoun (2014), for example, divided adult EFL learners into an experimental group and a control group. Those in the experimental group received a SA questionnaire biweekly whereas those in the control group did not. Both groups also completed an EFL self-efficacy questionnaire. It was found that the SA practices significantly enhanced the experimental group's self-efficacy as compared to the control group. SA has also been found to improve different language abilities among learners. Andrade and Boulay (2003), for example, revealed a positive relationship between SA and the quality of students' work in writing classes. SA was also demonstrated to contribute to students' L2 reading performance (Baniabdelrahman, 2010). De Saint Léger (2009) reported that SA strengthened the oral skills and vocabulary knowledge of adult learners of French as a Foreign Language. In a study on sixth-grade students in South Korea, Bulter and Lee (2010) found those young EFL learners' accuracy of doing SA significantly improved after one semester's intervention. More importantly, the study showed positive effects of the intervention on the learners' English performance and their confidence in English learning, which lent support to Oscarson's (1997) argument that SA enables L2 learners to improve their confidence in performing a task through learner-centric ways.

Despite the aforementioned benefits of SA, its implementation in language classrooms, however, has been limited (Butler, 2016; Oscarson, 2014; Panadero, et al., 2014), due possibly to the many concerns that students, and more importantly, teachers have about SA. Students, for example, expressed their concerns about "their competence in making judgements" (Harris & Brown, 2013, p.109) during the process of evaluating themselves. In Bulter and Lee's (2010) study, primary school EFL learners in South Korea reflected that they were not comfortable with assessing themselves (and their peers) even though they were provided with clear assessment criteria or rubrics. This finding perhaps also explains the little agreement on the consistency of SA or the low accuracy and reliability previously reported on younger learners (e.g., Blatchford, 1997; Brown & Harris, 2013; Butler, 2016; Oscarson, 1997, 2014). A more direct reason for the limited appearance of SA in classrooms perhaps pertains to teachers' perceptions and concerns (e.g., threat of SA to teacher authority and waste of instructional time), even though they may be cognizant of the benefits of SA, such as those mentioned earlier (Butler, 2016; Oscarson, 2014; Panadero et al., 2014). We will revisit this issue in further detail later.

Peer Assessment (PA) and Student Learning

PA refers to "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners" so that they can help each other, identify their strengths and weaknesses, and develop metacognitive and collaborative skills (Topping, 2009, p. 20). PA can involve a variety of students' skills and performance, such as writing, oral presentations, and portfolios; and can be conducted in various forms, such as peer feedback and grading (Brown & Abeywickrama, 2010; Edwards, 2014; Topping, 2009)

Many benefits of PA, as in the case of SA, have been identified for classroom learning. For example, PA could help improve student motivation (Edwards, 2014; Topping, 2009). It contributes to audience awareness, problem-solving skills, faster and timely feedback, communication and negotiation skills, and more opportunities for using the target language and developing language proficiency (Cheng & Warren, 2005; Edwards, 2014; Hung et al., 2016). It can also help students understand the quality of each other's work, foster students' higher order thinking and collaboration skills (Edwards, 2014; Sluijsmans & Prins, 2006; Topping, 2009). From a sociocultural perspective on language learning, "peer assessment activities provide opportunities for learners to be 'scaffolded' in learning through interaction with more knowledgeable peers" (Edwards, 2014, p. 2). Hung and others (2016) found that through the process of PA of speaking performance, young EFL learners in Taiwan were engaged as active learners, and the entire class was vibrant. This finding led the authors to argue that the interaction process during PA is particularly valuable in cultures where assessments are predominately teacher-centered.

Despite PA being identified as a beneficial assessment strategy for enhancing student learning; and many researchers suggesting that teachers use it in their classrooms (Edwards, 2014; Topping, 2009), the implementation of PA, like that of SA, has been limited (Dochy et al., 1999; Edwards, 2014; Harris & Brown, 2013). Like SA, a number of concerns about PA have been reported in the literature. Based on the data collected from college learners of English in Southeast Asia, Aryadoust (2016), for example, found students' oral presentations tended to be over-evaluated by peers of the opposite sex. In a study on PA (and SA) of oral presentations in an advanced Chinese class in an American university, Wang (2017) pointed out that students might overrate their peers due to "friendship," which could be attributed to their personality, cultural background, and the type of training they received. Another concern about PA pertains to students' age. Similar to the concerns about SA, PA among younger students might not be reliable; and sometimes, students could be discouraged if they perceive their performance to be "unfairly" rated by their peers (Dochy et al., 1999; Edwards, 2014; Harris & Brown, 2013).

Teachers' Perceptions and Classroom Use of Self and Peer Assessment

The benefits of SA and PA are not limited to student learning; they also help inform teachers' decision-making about their instruction based on the evidence collected from student assessment. Thus, teachers are often encouraged or required to incorporate SA and PA into their classroom assessment practices (Blanche & Merino, 1989; Brown & Abeywickrama, 2010; Edwards, 2014; Oscarson, 2014). Yet, as indicated earlier, classroom implementation of SA and PA has been very limited. Previous research on teachers and teaching, including language teachers and language instruction, revealed that what teachers know, think, and believe are closely related to what they do (and do not do) and how they teach in the classroom (Borg, 2006; Pajares, 1992). Thus, it is not a surprise that many factors pertaining to teachers, such as their perceptions of SA and PA and the teaching context (e.g., students' age and proficiency), have been found to be related to teachers' use (or nonuse) of SA and PA in their classrooms (Dixon et al., 2011; Harris & Brown, 2013; Sluijsmans & Prins, 2006).

Among the many concerns that teachers have about SA and PA is the low accuracy and reliability of these forms of student assessment (Cho, Schunn, & Wilson, 2006; Edwards, 2014; Harris & Brown, 2013; Oscarson, 2014; Sluijsmans & Prins, 2006). Harris and Brown (2013), for example, pointed out that teachers often worry about the accuracy of PA since it could be influenced by various factors such as students' self-expectation, friendship, proficiency level, and so on. Butler (2016) suggested that the low uptake of SA and PA could also be attributed to teachers' lack of understanding about when and how to integrate SA and PA into their assessment system.

Teachers' beliefs about role relationships in assessment could also have an influence on the implementation of SA and PA in their classrooms. Harris and Brown (2013), for example, contended that many teachers tend to consider classroom assessment as their own responsibility, and do not feel a need to actively involve students in the assessment processes. Butler and Lee (2010) noted that teachers who worked in a test-oriented environment, such as those in East Asian countries like South Korea, had little use of SA since they tended to believe that assessment should be associated with ability but not effort, the latter of which is often seen as an integral component of SA as a formative assessment method. In addition, in order to keep the traditional authority as a teacher, some teachers in those contexts of language teaching completely avoided using SA and PA in fear of power change in the classroom.

To sum up, SA and PA have been recognized as a beneficial assessment method to support student learning, as well as guide language teachers' planning and instruction across the literature. However, the studies on SA and PA and their implementation in language classrooms have focused primarily on ESL/EFL contexts, and a larger majority of them were conducted from the perspective of students (e.g.,

Aryadoust, 2016; Baleghizadeh & Masoun, 2014; Cheng & Warren, 2005; Hung et al., 2016). In Chinese language education in the United States, despite recent calls for school teachers to use multiple assessments, including SA and PA (ACTFL, 2013; NEALRC & CLASS, 2007), little research has been conducted from the teacher's perspective, such as how teachers view SA and PA, whether they implement SA and PA in their classrooms, and what factors may be associated with their use or nonuse of these two forms of student-centric assessment.

Research Questions

To address the aforementioned gap in research on Chinese language assessment, this paper explored the perceptions of SA and PA among K–12 Chinese language teachers in US schools and the factors behind the implementation of SA and PA in their classrooms. The following two questions guided our exploration.

1. How do K–12 Chinese language teachers in the United States understand SA and PA?
2. What factors, if any, are associated with those teachers' use, or nonuse, of SA and PA?

Method

Research Design

This is an exploratory study with a qualitative methodology. We interviewed ten teachers teaching Chinese as a second language (CSL) in US K–12 schools to understand their perceptions and practices of SA and PA. Through this study we intended to have an initial look into the SA and PA and establish a ground for future research in SA and PA in K–12 Chinese language classrooms in US schools.

Participants

The ten participating teachers had diverse backgrounds and experiences of teaching L2 Chinese in US schools and their contexts of teaching also showed notable differences (see Table 1). They were selected and contacted and later voluntarily participated in this study because of their connections with the authors. While this sample does not allow us – and it is not our purpose, either – to generalize the findings to all K-12 Chinese language teachers, the diverse backgrounds and contexts of teaching of the participants are expected to collectively provide some shared voices about SA and PA among Chinese language teachers; more importantly, these individual cases also offer an insight into possible impact of personal and contextual variations on the implementation of SA and PA in K-12 Chinese language classrooms.

Table 1 provides an overview of the backgrounds and contexts of the teaching of the ten teachers. All participants were native Chinese speakers and all were female. Seven of them were teaching in public schools and three in private schools. The age ranges of the participants varied, but most fell in 26-30. The grades of the students they taught varied, from kindergarteners to high school students. Overall elementary school teachers reported that their students' Chinese proficiency level was novice low, whereas those who taught in middle or high schools reported that their students' proficiency varied, depending on the level of classes the students took. For example, a few high school teachers (e.g., participants F, H, and J) taught both beginning-level learners and the learners studying Advanced Placement Chinese. The ten participants also had varied experiences of teaching, including years of teaching Chinese in US schools (3-9 years). Four of them (i.e., participants A, D, F, and G) were previous teachers of Chinese or English in mainland China or Taiwan; one (i.e., participant J) had also taught Chinese to college students in the US; the rest did not have additional experiences of teaching. As shown in Table 1, all participants received a bachelor's degree – not necessarily in language education, though – in mainland China or Taiwan; some also received a master's degree in China. All except two (i.e., participants F and G) received a master's degree from an American university in curriculum and teaching. Finally, all teachers except one (i.e., participant J, who taught in a private high school) were certified Chinese language teachers; one

(i.e., participant C, who taught in a Chinese immersion program) was also certified in elementary school education (K-6).

Note. All teachers were female. Age refers to the age range a participant indicated during the interview. Grades refer to the grade(s) that a teacher was teaching at the time of the study. State refers to the state where a participant was teaching. Years of teaching refer to the total number of years a teacher had in teaching K-12 schools in the United States; the number included years in the current school(s) as well as previous teaching in other US schools (but not years of other teaching experiences, including those in non-US contexts). Certification refers to the type of teaching certificate that a participant received from an educational authority in the United States (e.g., State Department of Education).

Data Collection

The data came from a larger study that investigated the assessment literacy and classroom-based assessment practices of K–12 World Language teachers in the United States. They were collected through semi-structured interviews, each of which lasted 1.5-2 hours. The interviews were all conducted online in Chinese through Skype or Zoom and at the same time audio-recorded. They consisted of three major parts. The first part focused on the teachers' backgrounds, such as the information reported in Table 1. The second part focused on their perceptions and practices of various types of classroom-based assessment. The last part elicited the teachers' report on their professional development needs in language assessment. As an essential component of the second part of the interview, the teachers were asked questions on their understandings about SA and PA and use of these two forms of student-centered assessment in their classrooms and were then probed to explain their use or nonuse.

Data Analysis

The audio-recordings of the interviews were transcribed in Chinese for subsequent analyses (excerpts, however, are presented below in English translation). The first step of analysis for the current study was open coding (Strauss & Corbin, 1998), in which meaningful segments of the scripts about SA and PA were marked and labeled. The labels were named by key words or short phrases based on the research questions. Build on this initial coding, we explored themes and patterns related to the teachers' understandings and perceptions of SA and PA, and their reported use or nonuse of SA and PA and the reasons for such use or nonuse. All authors read the interview transcripts and then discussed codes and excerpts together to ensure the credibility of the coding process.

Findings

Teachers' Understandings and Reported Use of SA and PA

At the interviews, all participants reported that both SA and PA could increase students' learning motivation and help enhance their self-regulated learning in addition to providing opportunities for them to practice Chinese language skills, particularly the affordances of PA for students to work together in a cooperative learning environment.

Table 1.

Backgrounds and Contexts of Teaching of the Ten Participants

Participant	Age	Grades	State	Years of Teaching	Degrees	Certification
A	46-50	6-8; mixed levels	Illinois	9 (additional middle school teaching experiences in Taiwan)	Bachelor's degree in English Literature (Taiwan); Master's degree in Curriculum and Instruction (US).	K-12 Chinese Teaching Certificate
B	26-30	3-5; novice	Michigan	5 (no additional experience of teaching)	Bachelor's degree in Public Relations (mainland China); Master's degree in Curriculum and Instruction (US)	K-12 Chinese Teaching Certificate
C	41-45	K-1; novice	Michigan	9 (no additional experience of teaching)	Bachelor's degree in International Business (mainland China); Master's degree in Curriculum and Instruction (US)	Chinese Teaching Certificate (K-12), and Elementary School Teaching Certificate (K-6)
D	46-50	K-5; novice low	Texas	5 (additional experiences of teaching English to college students in mainland China)	Bachelor's degree in education (mainland China); Master's degree in English language and literature (mainland China); Master's degree in English literature (US)	Chinese Teaching Certificate (K-12)
E	26-30	K-1; novice low	Wisconsin	5 (no additional experience of teaching)	Bachelor's degree in law (mainland China); Master's degree in Curriculum and Instruction (US)	Chinese Teaching Certificate (K-12)
F	26-30	7-12; mixed levels	Arkansas	3 (additional experience of teaching Chinese to college students in mainland China)	Bachelor's degree in English Language and Literature; Master's degree in Teaching Chinese as a Second Language. Both from mainland China.	Chinese Teaching Certificate (7-12)
G	26-30	9-12; mixed levels	Arkansas	3 (additional years of teaching Chinese in mainland China)	Bachelor's degree in marketing; master's degree in Chinese philology; and PhD in Teaching Chinese as a Foreign Language. All degrees from mainland China.	Chinese Teaching Certificate (7-12)
H	26-30	9-12; mixed levels	Pennsylvania	3 (no additional experiences of teaching Chinese)	Bachelor's degree in Chinese language and literature (mainland China); Master's degree in Linguistics (mainland China); Master's degree in Foreign Language Education (US)	Chinese Teaching Certificate (K-12)
I	26-30	1; novice low	California	7 (no additional experience of teaching)	Bachelor's degree in British and American literature (mainland China); Master's degree in Curriculum and Instruction (US).	Chinese Teaching Certificate (K-12)
J	41-45	9-12; mixed levels	Pennsylvania	7 (additional years of teaching Chinese to college students in US)	Bachelor's degree in English Language and Literature (mainland China); Master's degree in Public management (US).	None

Increasing students' motivation. All participants reported that SA and PA, if used properly, can help increase students' motivation for learning Chinese. Four teachers (participants A, B, D, and H) reported that SA provides students with opportunities to develop a sense of ownership for their own learning, and could thus encourage them to be more intrinsically motivated to learn Chinese and take initiatives to learn it. Participant B, for example, mentioned, "if students could evaluate themselves through self-assessment, they would feel that they have the capability of controlling their own learning… and then…love to learn…" Two teachers (participants D and H) also reflected that students might feel peer pressure during pair work for PA, which they believed PA would motivate students to put more efforts on reviewing what they have learned for a better accomplishment. As alluded to by Teacher H in the excerpt below, PA can help students know each other's work better and motivate them to learn from each other.

> *Excerpt 1* (Participant H; PA; motivation)
> A benefit (of peer assessment) is that you can get to know what your classmates are doing and how they are doing … actually many students would make comparisons (of themselves with others), like, my work is not as good as that of my classmates. In this way, students could get motivated.

Enhancing students' self-regulated learning. SA and PA, if used appropriately, can also help enhance students' self-regulated learning (Zimmerman, 1990). For example, SA, if used in the form of a checklist, can help a student quickly assess what he/she has achieved and what has yet to be accomplished (Oscarson, 2014). With this learning goal set in mind, the student could then regulate his/her learning accordingly. In this sense, SA could serve as a learning guideline for students. As reported by Participant A, "…if a student knows his/her goal ahead, the teacher would no longer need to push hard for him/her…" She further expressed, as shown in the following excerpt, that students should be guided (through activities like SA and PA) on taking ownership of and setting goals for their own learning; if they are unclear about their goals, they would not be motivated to learn.

> *Excerpt 2* (Participant A; SA and PA; self-regulation)
> If you always tell students that you should do what and what, students might not be able to understand (your expectations) and thus lack motivation (for learning) … I believe if students … have goals set for themselves, they would be happy to make efforts to achieve those goals.

Opportunities for students to practice language skills. Some teachers in this study also reported that SA and PA tasks provide good opportunities for students to practice specific language skills over and beyond their assessment purposes. Participant D, for example, described that "I used pair work as a way of peer assessment ... I ask one student [to] read aloud Chinese words shown on flashcards and the other student to translate the words into English. This activity helped students check whether they had understood (the meanings of) the target Chinese words … It also gave the students an opportunity to further practice those words." Some teachers also commented that PA activities afford students opportunities to practice their listening skills by assessing others' oral performance in Chinese such as in a dialogue. Participant D, for example, indicated that those opportunities could help her direct her K–1 students' attention to others' pronunciations and help those students correct their own mispronunciations. Teachers' attention to the benefit of reflexive learning of PA is particularly evident in the following sharing by Participant A on her use of PA to get her middle-school learners to evaluate each other's oral presentation.

> *Excerpt 3* (Participant A; PA; oral presentation)
> My purpose is to let them reflect, like, if they thought the student who just performed did not speak well, how could he/she do better? This is like self-reflection …

In summary, the participants reported several major benefits of SA and PA in their contexts of teaching. Not only do SA and PA tasks provide opportunities for students to practice and improve language skills but SA and PA can guide students through the learning process and help them take charge of their own learning (e.g., goal setting), which eventually can promote their motivation for Chinese language learning.

Factors Associated with Teachers' Use (or Nonuse) of SA and PA

The findings reported in the foregoing section suggest that all the participants seemed to have understood the importance of SA and PA and their roles in classroom instruction and student learning.

However, it seemed that these teachers' SA and PA practices were not as good as expected, although they seemed to understand the importance of SA and PA in instruction and student learning. When asked the extent to which they used SA and PA in their classroom assessment practices, only two of the participants (i.e., Participants A and C) reported fairly frequent use of them; and for most of the others, SA and PA were barely used. Three major factors that emerged from the probed responses of the teachers at the interviews seemed to explain this variation. These factors included concerns about students and the accuracy and reliability of SA and PA, lack of sufficient time and expertise for conducting SA and PA, and availability of school support. In addition, it was also inferred that the variation in teaching experiences might be an additional factor.

Students and reliability concerns. Most of the participants taught young children with very low proficiency (see Table 1). Thus, it did not seem a surprise that many reported concerns of using SA and PA from the perspective of students. Participant I, a private school teacher who taught first graders in California, for example, reported that her students could barely do SA (reliably), because it was very difficult to make sure that they understood a rubric (e.g., criteria like grammatical appropriateness and fluency of presentation and the distinction between different levels of performance) and her expectations.

> *Excerpt 4* (Participant I; SA; reliability)
> I think a challenge is to set up a clear goal for students. These young learners are not clear what the ratings (in a rubric) mean – which presentation is worth 1 point and which ones 2, 3, or 4 points – and therefore, a challenge to the teacher is how to make sure the rubric and expectations are clearly understood by them.

Over and beyond the challenge of making sure the goal or expectations are clearly communicated to students is an additional concern about students' ability to follow the expectations (reliably) to conduct SA and PA. Participant I, for example, further shared in the excerpt below that PA in her class of first graders could easily become an activity of looking for each other's faults, which would result in no one feeling the PA activity as a motivating and rewarding experience. Participant A's emphasis in the excerpt on student training for conducting SA was also echoed by Participants B and D, who also taught young children.

> *Excerpt 5* (Participant I; PA; reliability)
> PA among young children is like looking for each other's faults. Everyone would then feel uncomfortable. Without appropriate training, young children would only look for others' faults …

Some teachers also mentioned the undesirable impact of students' peer relationships (e.g., friendship) on the accuracy and reliability of conducting PA. Participant E, expressed her concern that students might give a very high score to their close friends and a very low one to those whom they perceived as unfriendly to them, despite the fact that the rubric and expectations might have been clearly communicated to them by the teacher.

Participant A's concern added to Participant E's in that students' ratings of their peers might also depend on their mood during the assessment activity, which would result in a legitimate concern about the low intra-rater reliability of PA. In addition, Participant A echoed a concern previously reported about SA (e.g., AlFallay, 2004) in that some students' overconfidence could inflate their ratings and consequently undermine the reliability, and validity, of SA. In an example she gave at the interview, Participant A felt the language of her students' essays was too general and would need elaborations and details; yet her students' SA results were all very positive, which led her to believe that it would not be desirable to give much weight on SA results if these results were to be included as part of the students' final grades. This concern of Participant A's also seemed to explain why only one teacher in the study (i.e., Participant H, who taught in a private high school) indicated that she would

consider incorporating PA and SA results as part of her students' final grades; and why none of the elementary school teachers reported an inclination to include SA and/or PA results in their students' report cards.

Time constraints and lack of expertise. In addition to the aforementioned concerns from the perspective of students, time constraint and lack of expertise were also reported by many teachers as discouraging their implementation of SA and PA. The teachers shared that designing and conducting SA and PA and interpreting the assessment results were all very time-consuming. Participant E, who taught young children (kindergarteners and first graders), for instance, framed SA for her students as the teacher sitting together with each one to guide them on assessing their own language performance, possibly due to her concern that those students would not otherwise be able to understand her expectations clearly and subsequently follow the expectations accurately and reliably to assess themselves. Thus, it does not seem surprising that Participant E was not positive on the possible use of SA in her K-1 classrooms, as shown in the excerpt below.

> *Excerpt 6* (Participant E; SA; time)
> I think conducting SA is very time consuming to the teacher, because you need to understand that a (one-to-one) conference would very likely be needed. You (as the teacher) would need to make sure that each student is clearly aware of his/her learning goals and where he/she is now; you would also need to help him/her remember how to achieve those goals.

In addition to time considerations, the lack of expertise for conducting SA and PA was reported as another concern. Without a clear understanding of the specific procedures for conducting SA or PA, it would reasonably be hard for teachers to carry out these types of assessment in their classrooms. As an example, Participant G reflected in the following excerpt that she did not have a deep understanding of the key concepts and procedures underlying SA and PA (and other forms of alternative assessment).

> *Excerpt 7* (Participant G; SA/PA; knowledge)
> I'm not familiar with the theoretical underpinnings of these forms of assessments (i.e., self and peer assessment). I possibly got some ideas from colleagues or workshops I previously attended … But I am not very clear how to describe them appropriately …

Support from school. Different from the above "barriers" that hindered the use of SA and PA, a few teachers also shared factors that seemed to have facilitated their use of the two forms of assessment in their classrooms, such as support from schools. Participant G, for example, made a specific reference to a school-wide initiative for SA across subject areas. More specifically, her school provided a Socratic observation and evaluation form for all teachers to conduct PA and SA; and as a result, students were familiar with the format of the assessment, and did not require a lot of training for using it in her Chinese language classes because of the wide use of the assessment (in English-medium classes).

> *Excerpt 9* (Participant G; SA; support)
> … (we) use a Socratic form, which is used by dividing students into two circles. Students sitting in the inner circle discuss issues, whereas those in the outer circle observe the discussion and fill in a peer assessment form of the discussion. Students in the inner circle also fill in a self-assessment form. This Socratic form is provided by the school district, and this form of assessment is promulgated in the whole district. Students (are thus very familiar with the form and) do not need to be trained (in my class), as they have already been trained in their English-medium classes…

Teaching experience. Besides the factors discussed above that all emerged from the teachers' direct report at the interviews, it also appears that teaching experiences tended to have an impact on the participants' classroom implementation, or the lack thereof, of SA and PA. By comparing the teachers who reported a fair amount of use of SA and/or PA and those who rarely used them, we found that the former tended to be those who had had much more teaching experiences. Overall, those who reported a lot of concerns and challenges for conducting SA and PA were those who appeared to be in the early stage of teaching (e.g., within the first three years of teaching Chinese in US schools). Participant C, for example, reported that she was very open to both SA and PA in her classrooms even though her students were young (i.e., K–1) and novice learners. Additionally, we noticed

that those teachers with more experiences in teaching in US schools tended to be more flexible and skillful in using multiple ways of assessment to support student learning. For example, Participant A, who had taught in her current middle school in Chicago for about 9 years, reported her use of checklists, such as ACTFL Can-Do Statements, to let students self-assess at the beginning of learning a unit and then ask them to self-assess again at the end of the unit so that her students could compare the results of the two self-assessments to check any progress they may, or may not, have made for the unit of study.

Discussion and Conclusion

The ten cases of teachers from diverse backgrounds and contexts of teaching in US schools generated some important insights into Chinese language teachers' perceptions and practices of SA and PA. The most important one was perhaps the gap between the recognition of the benefits of SA and PA (e.g., enhancing students' motivation and self-regulated learning) among all the teachers and the rare use of these forms of student-centered assessment in most teachers' classrooms. In what follows, we discuss this gap by drawing upon the factors that emerged from the participants' interview responses as well as those reported in the literature but not necessarily revealing in the teachers' responses. We also discuss the implications of the findings for classroom assessment and teacher professional development in language assessment.

Making Sense of the Gap between Perceived Importance and Classroom Practice

A major concern from the participants about implementing SA and PA pertained to their students, who were mostly very young and low in proficiency in the present study and were thus perceived to be unable to assess themselves as well as others accurately, fairly, and reliably. Reasonably, if students have difficulty understanding the procedures and expectations of SA and/or PA or could not follow the expectations for fair and reliable assessment, why would teachers have any incentive to implement them? As a matter of fact, many of the concerns voiced by the teachers in the present study from the perspective of students, such as friendship, had been previously reported in the literature on SA and/or PA (Brown & Harris, 2013; Cho et al., 2006; Edwards, 2014; Harris & Brown, 2013; Oscarson, 2014; Sluijsmans & Prins, 2006).

Of course, one way to address the challenges discussed above would be to provide sufficient training for students so that they could understand and later follow the expectations for accurate and reliable assessment, which has actually been argued as an integral component of implementing SA and PA (Brown & Abeywickrama, 2010; Brown & Harris, 2013; Edwards, 2014; Oscarson, 2014). Yet, a subsequent concern would be the time constraint for making this happen, particular among very young learners. As some teachers reported in this study, conducting SA and PA is very time-consuming. In a class with a large number of students, it would be particularly challenging to ensure that detailed guidance and feedback, which are believed to be essential for successful practices of SA and PA (Blanche & Merino, 1989; Edwards, 2014; Oscarson, 2014), are provided contingently for all students. Thus, given the typically small number of class hours for Chinese language programs in K–12 schools, it seems reasonable that the teachers in this study generally did not find SA and PA to be "cost-effective" for implementation in their classrooms.

Another factor for the rare use of PA and SA prevailing in the participants might be their lack of expertise – a possible result of insufficient training – in the specific procedures for conducting PA and SA, even though few directly voiced this out at the interviews. Conducting SA and PA effectively is no easy task, even if sufficient time could be granted. In practice, students would need to be provided with detailed guidance or be trained with specific examples to become familiar with a rubric (e.g., criteria, descriptors, and ratings), expectations, and procedures for conducting SA and PA (Brown & Harris, 2013; Edwards, 2014; Oscarson, 2014). As Edwards (2014) argued, "training before the PA task commences is probably the most important component of the overall PA training cycle, though training during and after the PA should not be neglected" (p. 18). This essentially means that teachers would need to have received training in SA and PA and be skilled with the procedures of implementing them so that they could train their students adequately for SA and PA tasks and ensure the tasks are completed effectively and efficiently.

Finally, we speculate that the gap between the perceived benefits of SA and PA and the rare implementation of them might be attributed to the teachers' cultural background. In the present study, all participants were native Chinese speakers educated in a culture or educational system (i.e., mainland China or Taiwan) where the teacher is regarded as the authority, classroom teaching is predominantly teacher-centered, and assessment often takes a single form and is primarily for summative purposes (e.g., a final exam). As indicated by Wang and Higgins (2008), pedagogical controversy for teachers across borders might arise from differences between cultures of instruction. It would thus seem reasonable that the participants, who were teaching in an international, cross-cultural context, may be oriented toward teacher-centric assessment practices (as opposed to student-centric ones, such as SA and PA). In the present study, even though most of the teachers received a master's degree in education from a US university and were certified to teach Chinese in the US, the influence from their primary educational experiences might have persisted. As Moore (1992) noted, the teaching conflicts and struggles of Chinese teachers working in an international context were often due to their deeply rooted philosophy of collectivist culture, which tends to make their transition to (teaching in) an individualist culture, such as the United States, an overwhelmingly intimidating experience.

Implications for Classroom Assessment and Teacher Professional Development

The findings of this study and the preceding discussion have some implications for assessment in Chinese language classrooms and professional development for Chinese language teachers. A notable one is that teachers need to develop a strong knowledge base about SA and PA and be trained in the specific procedures for conducting them, including training students, in the classroom context. This seems particularly important for those who work with young learners, who have been reported in the literature as well as the present study to have difficulty understanding and following expectations for fair and reliable assessment of themselves and their fellow students. In the present study, it was found that those who reported a fair amount of SA and PA implementation tended to be those who had a significant number of years of teaching Chinese in US schools (e.g., participants C and A; see Table 1). While there was no direct evidence to suggest that those teachers necessarily had stronger assessment literacy, it seems reasonable to assume that they were more experienced, and were thus more skilled, in multiple ways of assessing students.

Another implication is that school support is highly necessary. As described earlier in the Findings section, a teacher (i.e., participant G) reported the synergy that her school district built on using a Socratic form for SA and PA across subject areas had greatly facilitated her use of these forms of student assessment in her high school Chinese language classrooms. While there was no explicit information from the other teachers on any facilitating effect of having school support or concerns about not having that support, it seems clear from the experience of participant G that school support is an important condition for bridging any gap between teachers' perceptions of the benefits of SA and PA and their rare implementation of them in actual classroom practices.

Despite the fact that both SA and PA have multiple benefits for language learning and that they should be an integral component of language teachers' repertoire of classroom assessment (ACTFL, 2013; Brown & Abeywickrama, 2010; Butler, 2016; Edwards, 2014; Oscarson, 2014), there has been little attention to SA and PA in Chinese language education in the United States. Under this context, this study is important because it aimed at bridging the research gap and exploring how Chinese language teachers, who taught in US K–12 schools, understood SA and PA as well as the rationale associated with their perceptions and practices of SA and PA. This study provided some insight into Chinese language teachers' perceptions and practices of SA and PA in their classrooms.

This study showed that the Chinese language teachers had a good understanding about the role of SA and PA in classroom instruction and student learning, such as increasing students' motivation, enhancing students' self-regulated learning, and providing opportunities for students to practice their language skills. Their understandings of SA and PA, to a large extent, were consistent with benefits of SA and PA to student learning in previous literature (e.g., Brown & Abeywickrama, 2010; Cheng & Warren, 2005; Edwards, 2014; Oscarson, 2014; Ross, 2006). However, there seemed a variation between their recognition of the benefits of SA and PA and their classroom practices in SA and PA. The factors that could explain the variation included teachers'

concerns about their students (e.g., young age and low proficiency) and the accuracy and reliability of SA and PA, time constraints, lack of sufficient expertise, and school support. In addition, we speculated that the teachers' experiences of teaching in US schools and their prior educational experiences (in mainland China and Taiwan) might also have had an impact on their use or rare use of SA and PA. Findings in this study can contribute to SA and PA practices of Chinese language in K–12 classrooms and professional development for Chinese language teachers in US schools.

Our findings may be limited and cannot be generalized to all K–12 US school Chinese language teachers' SA and PA practices due to its relatively smaller body of participants. Another limitation is that the findings were all based on teacher self-report. Future research can investigate Chinese language teachers' perceptions and implementations of SA and PA by recruiting a larger body of participants and also observe teachers' classroom-based assessment practices. In addition, this study indicated that SA and PA not only benefited student language skills but also their metacognition skills such as self-regulation. Thus, future researchers can further investigate tools and the possibility of promoting students' metacognition development using student's' first language when necessary and possible in classroom SA and PA practices.

Acknowledgments

References

American Council on Teaching of Foreign Languages (ACTFL). (2013). *Program standards for the preparation of foreign language teachers*. Accessed on December 20, 2020 https://www.actfl.org/sites/default/files/caep/ACTFLCAEPStandards2013_v2015.pdf

AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System, 32*, 407–425.

Andrade, H. G., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research*, *97*(1), 21–30.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, *48*(1), 12–19.

Aryadoust, V. (2016). Gender and academic major bias in peer assessment of oral presentations. *Language Assessment Quarterly*, *13*(1), 1–24.

Baleghizadeh, S., & Masoun, A. (2014). The effect of self-assessment on EFL learners' self-efficacy. *TESL Canada Journal*, *31*(1), 42–58.

Baniabdelrahman, A. A. (2010). The effect of the use of self-assessment on EFL students' performance in reading comprehension in English. *TESL–EJ*, *14*(2). Accessed on January 30, 2020 from http://www.tesl-ej.org/wordpress/issues/volume14/ej54/ej54a2/

Blanche, P., & Merino, B. J. (1989). Self- assessment of foreign- language skills: Implications for teachers and researchers. *Language Learning, 39*(3), 313–338.

Blatchford, P. (1997). Students' self-assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison group. *Educational Psychology*, *17*(3), 345–359.

Borg, S. (2006). *Teacher cognition and teacher education: Research and practice*. London: Continuum.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.). *The Sage handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, CA: Sage.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). London: Pearson.

Brown, N. A., Dewey, D. P., & Cox, T. L. (2014). Assessing the validity of Can- Do statements in retrospective (then- now) self- assessment. *Foreign Language Annals*, *47*(2), 261–285.

Butler, Y. G. (2016). Self-assessment *of* and *for* young learners' foreign language learning. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 291–315). Dordrecht, The Netherlands: Springer.

Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5–31.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing, 22*(1), 93–121.

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891–901.

De Saint Léger, D. (2009), Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals, 42*, 158–178.

Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting Assessment for Learning: the beliefs practice nexus. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 365–379.

Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, *24*(3), 331–350.

Edwards, J. G. H. (2014). Peer assessment in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. II, Part 6) (pp. 730–750). Wiley.

Harris, L. R., & Brown, G. T. (2013). Opportunities and obstacles to consider when using peer-and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101–111.

Hung, Y. J., Samuelson, B. L., & Chen, S. (2016). The relationships between peer and self-assessment and teacher assessment of young EFL learners' oral presentations. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 317–338). Dordrecht, The Netherlands: Springer.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, *26*(1), 75–100.

Moore, S. J. (1992). *Introducing Chinese into high schools: The Dodge Initiative*. Washington DC: National Foreign Language Center.

National East Asian Languages Resource Center & Chinese Language Association of Secondary-elementary Schools (NEALRC & CLASS). (2007). *CLASS professional standards for K-12 Chinese teachers*. Accessed on January 30, 2020 from https://nealrc.osu.edu/sites/nealrc.osu.edu/files/teacher-k-12-class-teachers-standards.pdf

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*(1), 1–13.

Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education, Vol.7 Language testing and assessment* (pp. 175–187). Dordrecht, The Netherlands: Kluwer.

Oscarson, M. (2014). Self- assessment in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. II, Part 6) (pp. 712–729). Wiley.

Pajares, F. M. (1992). Teacher's beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332.

Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 365–383.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational psychologist*, *36*(2), 89–101.

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, *19*(2), 109–131.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation, 11*(10), 1–13.

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment & Evaluation in Higher Education*, *23*(4), 391–402.

Sluijsmans, D., & Prins, F. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation*, *32*(1), 6–22.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research.* Thousand Oaks, CA: Sage.

Topping, K. J. (2009). Peer assessment. *Theory into practice*, *48*(1), 20–27.

Wan-a-rom, U. (2010). Self-assessment of word knowledge with graded readers: A preliminary study. *Reading in a Foreign language, 22*(2), 323–338.

Wang, D. (2017). Self- and peer-assessment of oral presentation in Advanced Chinese classrooms: An exploratory study. In D. Zhang & C.-H. Lin (Eds.), *Chinese as a Second Language assessment* (pp. 271–286). Singapore: Springer.

Wang, L., & Higgins, L. (2008). Mandarin teaching in the UK in 2007: A brief report of teachers' and learners' views. *Language Learning Journal, 36*(1), 91–96.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3–17.

Xuexue Yang is a Ph.D. candidate in the Department of Teacher Education at Michigan State University. Her research interests include second and world language education, ELL/Bilingual Education, and Teacher Education. Xuexue has worked with Chinese language student teachers in the Chinese Teacher Certification Program in the United States. Her research has been published in peer-reviewed journals such as the *International Multilingual Research Journal* and Foreign Language Annals. She is an active member of the American Educational Research Association (AERA) and American Council on the Teaching of Foreign Language (ACTFL).

Yanjiang Teng holds a Ph.D. in Curriculum, Instruction, and Teacher Education. He is a researcher in the International Studies Program at Michigan State University. His research interests are second language teacher education and language services. He has published peer-reviewed articles on second language teacher preparation and intercultural crisis communications. Currently, he is the assistant editor of *Journal of Second Language Studies*, and a review board member of the *Chinese Language Teaching Methodology and Technology*.

Zheng Gu is a Ph.D. Candidate in the Department of Teacher Education at Michigan State University (MSU). Her research interests include second language pedagogy and transnational language teacher education. She teaches methods courses to pre-service teachers such as Literacy Instruction and Assessment, and works with world language teaching interns at MSU. She is currently working on a college-level Chinese textbook and the relevant teacher's handbook.

Dongbo Zhang is Professor of Language Education in the Graduate School of Education, University of Exeter. He previously held appointments at Michigan State University (where he directed the Chinese Teacher Certification Program) and Nanyang Technological University. His research interests include second language reading and language teacher education. He has been published in *Applied Linguistics*, *Language Learning*, *Modern Language Journal*, among many other applied linguistics journals. He currently serves on the editorial (advisory) board of *Language and Education*, *Research in Language and Education*, *TESOL Journal*, and *TESOL Quarterly*. He is a co-editor of *Chinese as a Second Language Assessment* (Springer).