Chapter 12
Developing a Speaking Diagnostic Tool for Teachers to Differentiate Instruction for Young Learners of Chinese

Xiaoxi Sun, Jinghua Fan and Chee-Kuen Chin

Abstract Chinese language (CL) education in Singapore is encountering increasing challenges as a possible result of the familial language shift from Chinese toward English in the last few decades. One such challenge is the diversity of students' level of Chinese proficiency. To address this challenge, the Singapore Ministry of Education recommended that CL teachers make instructional differentiation based on good knowledge about students' Chinese proficiency levels and different developmental trajectories. It was against such a background that the package of Chinese language Oral Proficiency Diagnostic Tool (OPDT) was developed by the Singapore Centre for Chinese Language for Primary 1 (Grade 1) CL teachers in Singapore. The OPDT was comprised of the Chinese Oral Proficiency Diagnostic Rubrics, the Diagnostic Activity Package, and the Differentiated Instruction Activity Package that aimed to assist CL teachers to diagnose the strengths and weaknesses of Primary 1 students' Chinese oral proficiency, to engage students in using Chinese so that teachers could have adequate student output for diagnosis purposes, and to enable teachers to differentiate their instruction with reference to the result of their diagnosis. This chapter reports on our development of the OPDT, its key features, and our effort to validate it as a useful tool for CL teachers. We hope the Singapore experience would shed light on Chinese as a second language education in other contexts and informs classroom-based formative assessment, instruction, as well as professional learning of CL teachers.

Keywords Chinese language education · Chinese language oral proficiency diagnostic tool · Language assessment · Classroom assessment · Assessment for learning

X. Sun (✉) · J. Fan · C.-K. Chin
Singapore Centre for Chinese Language, Singapore, Singapore
e-mail: sunxiaoxi1234@gmail.com

D. Zhang and C.-H. Lin (eds.), *Chinese as a Second Language Assessment*,
Chinese Language Learning Sciences, DOI 10.1007/978-981-10-4089-4_12

Introduction

Singapore is a multilingual society with a population comprised of three major ethnic groups, including Chinese, Malay, and Indian. The ethnic languages of the three groups (i.e., Chinese, Malay, and Tamil) as well as English (i.e., the medium of school instruction) are the four official languages of the country. Under the bilingual education system in Singapore, all school students are required to study English as well as their respective mother tongue or ethnic language (Pakir 2008; Shepherd 2005). For example, ethnic Chinese Singaporeans are required to be competent in both English and Chinese. However, becoming bilingual (and biliterate) in the country is becoming an increasingly challenging task. A recent survey revealed that 59% of Chinese children spoke English as their dominant home language as a possible result of the importance ascribed to English as the medium of instruction in schools (Ministry of Education 2011). The familial language shift from Chinese toward English in the last few decades has brought about great changes to the sociolinguistic milieu in Singapore, and more importantly, it has created a lot of challenges for school education in the country, one being the diversity of students' level of oral proficiency in Chinese (Zhao and Liu 2010; Zhao et al. 2007). While some children have developed some level of oral competence, albeit not always satisfactorily, in Chinese at the commencement of primary schooling, others have to learn Chinese as a completely new language in primary school.

To address the aforementioned challenges, policies concerning mother tongue education, including Chinese language education, have been constantly reviewed and curriculum refined in Singapore. A notable, recent policy, as recommended in the 2010 report of the Mother Tongue Language (MTL) Review Committee (Ministry of Education 2011), is that MTL teaching should "first build the oracy foundation before learning reading and writing" (p. 15) and "recognise different starting points and apply appropriate methods for different learners" (p. 13). With this policy recommendation, an important focus of Chinese education in early primary grades, especially Primary 1 (P1) (Grade 1), is on the development of children's oral competence in Chinese. Teachers are thus advised to adapt their instruction to meet the learning needs of students from different home language backgrounds and with diverse levels of oral proficiency in Chinese.

However, instructional adaptation or differentiation is arguably not easy for Chinese language (CL) teachers as it is conditional upon their clear awareness of individual differences among children and good knowledge about their proficiency levels and different developmental trajectories. In other words, CL teachers first need to be able to appropriately assess students and diagnose their specific needs so as to inform their instructional planning and differentiation. There are, however, several challenges in practice. Firstly, current assessments commonly administered in early primary grades in Singapore do not align with the curricular focus on oral proficiency. For example, school examinations largely focus on vocabulary, grammar, and reading with speaking only taking up 10–20% of students' grades.

Consequently, test results provide very limited information for CL teachers to implement effective practice in oracy-focused instruction in their classrooms. Secondly, students' oral competence is commonly tested by means of out-loud-reading and picture-describing. Such ways of testing measures students' prepared "performance" rather than their abilities to use Chinese interactively in authentic or real communicative contexts. Thirdly, and most importantly, the assessments commonly adopted in schools are summative in nature (e.g., end-of-semester examination). As a result, CL teachers know little about their students' specific strengths and weaknesses in oral language development. Such a lack of knowledge about students' learning needs apparently hinders teachers' practice of differentiated teaching. Thus, it is very important that teachers know how to appropriately and effectively use performance-based, formative assessments in their everyday classroom instruction to diagnose the performance of their students so that differentiated instruction could be possible. Yet, most CL teachers in Singapore have not been prepared or supported to conduct such assessments.

It was against such a backdrop that, in 2010, the Singapore Ministry of Education (MOE) commissioned the Singapore Centre for Chinese Language (SCCL) to undertake a two-year project to develop a scientific, effective, and user-friendly assessment tool to support CL teachers' use of formative assessment and differentiated instruction to promote students' oral proficiency development. Specifically, the objectives of the project were threefold: (1) to develop a scale that would allow Singapore CL teachers to diagnose the strengths and weaknesses of P1 students' Chinese oral proficiency; (2) to develop a set of diagnostic oral activities that could engage students in using Chinese so that teachers could have adequate student output for diagnosis purposes; and (3) to develop a set of exemplar classroom activities that would enable teachers to differentiate their instruction with reference to the result of their diagnosis. The end product of the project was the P1 Chinese language Oral Proficiency Diagnostic Tool (hereafter, the OPDT), which is a package comprised of three components that serve different purposes, including the P1 Chinese Oral Proficiency Diagnostic Rubrics (hereafter, the Diagnostic Rubrics); the Diagnostic Activity Package (DAP); and the Differentiated Instruction Activity Package (DIAP).

In the following sections of this chapter, we report on the process of developing the OPDT and the different purposes of its three components. To show the effectiveness of the OPDT for enhancing students' oral proficiency development, we also report an intervention study in which P1 teachers were trained to use the package for in-class assessment and differentiated teaching. Although this project was conducted with young learners in Singapore, it sheds light on Chinese language education in other contexts with respect to using assessment to inform instruction and enhance students' CL learning.

The Diagnostic Rubrics

As a crucial component of the OPDT, the Diagnostic Rubrics are a formative assessment tool designed by the project team for CL teachers to diagnostically assess P1 students' oral performance in Chinese when they are participating in classroom oral activities. Teachers use the Diagnostic Rubrics to assess students' oral competence in Chinese with both a holistic score that represents the students' overall level and a set of analytic scores to demonstrate their strengths and weaknesses in different areas. The Diagnostic Rubrics cover four levels defined as four developmental stages, and include indicators of vocabulary, grammar and sentence structure, pronunciation and intonation, and interpersonal interaction and expression. Within each level and for each indicator, there are specific descriptors that represent how strong or weak a student's performance is.

Construction of the Diagnostic Rubrics

The development of the Diagnostic Rubrics began with developing a prototype, which was then validated and calibrated with some iterative processes. The prototype of the Diagnostic Rubrics took its form on the basis of empirical data of one-to-one interviews with 184 P1 students sampled from different modules (i.e., Bridging, Core, and Enrichment) in four primary schools.[1] The final composition of students from the three modules was 59, 70, and 55, respectively. Each interview lasted for 25–30 min, with three main topics on food, toys, and festivals. The children were first asked to describe or compare pictures shown to them with guiding questions like "What do you see in these pictures?" and "Do you see any difference between the two pictures?" They were then engaged in interactions about the pictures with the interviewer, who asked open questions according to a child's previous descriptions of the pictures. The interviews were both audio- and video-taped.

The audio-taped data were transcribed and coded on four categories, namely, lexical, sentential, pronunciation and intonation, and discourse. With the coded speech data as the baseline, four levels were created to represent the participating students' Chinese oral proficiency. In addition, key characteristics of the speeches

[1]All of the participating schools in the project were selected with recommendation from CPDD (Curriculum Planning and Development Division), MOE, and they consisted of an equal number of mission and neighborhood schools, with comparable student composition ratios. Students attending mission schools in Singapore generally come from English-speaking homes, whereas those in neighborhood schools show a variety of home language use patterns, but typically with a significant proportion using Chinese as the major home language. It is beyond the scope of this paper to provide details of the three different modules. The modular approach is an essential characteristic of the (differentiated) Chinese curriculum in Singapore. Interested readers can refer to Li et al. (2012) for an introduction to the modular curriculum.

of those students within each of the four proficiency levels were constructed into four indicators that belonged to two main aspects of oral language proficiency, including linguistic (i.e., lexical, sentential, and pronunciation and intonation) and communicative competence (i.e., discourse).

The prototype Diagnostic Rubrics were first used by some members of the research team as well as master teachers affiliated with SCCL for trial assessment of some P1 students with the speech data previously collected. In addition, we also consulted a panel of experts of Chinese language education and assessment to solicit their feedback and professional input. The panel was comprised of a former member of the Singapore Examinations and Assessment Board, an assessment expert from National Institute of Education (a teacher training college in Singapore), and a visiting professor of Chinese language education at SCCL.

After incorporating the experiences from trial use and the experts' input, a full version of the Diagnostic Rubrics (with a user's manual and speech samples) was prepared. In the Diagnostic Rubrics, there were four distinguishing levels, which were metaphorically named as emerging stage (萌芽期), developing stage (展叶期), blooming stage (开花期), and accomplishing stage (幼果期) to symbolize the developmental levels of students' Chinese oral competence at the beginning of primary schooling. Across all four levels, the Diagnostic Rubrics described two main areas of oral proficiency, namely, linguistic competence and communicative competence. Linguistic competence was presented with indicators of vocabulary, grammar and sentence structure, and pronunciation and intonation. Communicative competence was indicated as interaction and expression, which looked into actual language use in interpersonal interaction. Each indicator included qualitative descriptors drawing on ascending quantifiers that represented degrees to which students demonstrate evidence of linguistic and communicative abilities across the four stages. For example, vocabulary at the emerging stage is described as "very poor or limited vocabulary; vocabulary restricted to simple responses and incomplete meaning expressions;" and communicative competence at the accomplishing stage is described as "sufficient responses that often include details; natural pausing and smooth turn-taking; frequent initiations through questioning; chunks of coherent discourse on a given topic." Finally, in addition to these four indicators for analytic rating, the Diagnostic Rubrics also included a holistic rating option to assign an overall performance level/stage of students.

Validation of the Diagnostic Rubrics

Unlike speaking scales used primarily in a high-stakes testing context, our Diagnostic Rubrics were designed to be used in classrooms by CL teachers as a formative assessment tool that could guide instructional planning for differentiation. To examine whether the Diagnostic Rubrics could be effectively used by teachers to diagnose their students' oral Chinese proficiency, a validation study with iterative processes was conducted.

A group of 20 P1 teachers participated in the study. They came from four different schools that did not participate in the study described earlier, where some P1 students were sampled and interviewed to provide the baseline data for developing the prototype Diagnostic Rubrics. After training through a workshop, those teachers independently used the Diagnostic Rubrics for three weeks, and focus group discussions were also conducted with them at this time. The tool was then calibrated based on the feedback from the teachers. Afterward, members of the research team were assigned to visit the classes of the 20 teachers for two weeks, during which time, the teachers and members of the research team (or co-raters) referred to the calibrated Diagnostic Rubrics independently and assessed the same group of students identified for this study in each teacher's class.[2] A total number of 196 students across 20 P1 classes were diagnosed. Table 12.1 shows the distributions of different ranges of intra-class correlation coefficients between the ratings of a teacher and his/her co-rater. The degree of rating consistency indicates the inter-rater reliability of the Diagnostic Rubrics.

Table 12.1 shows that the correlations between the ratings given by most teacher-and-co-rater pairs were significant, which means that the teachers used the Diagnostic Rubrics in a fairly consistent way as did the members of the research team who developed the tool. The proportions of significant correlation coefficients in vocabulary, grammar, pronunciation, communication, and overall score were, respectively, 65, 70, 70, 80, and 80%. Take communication as an example; the ratings of most of the teachers (N = 12 or 60%) were strongly correlated with those given by the research team (with intra-class correlation coefficients being 0.800 or higher); and there were only four teachers (20%) whose ratings did not show significant correlations with those of the research team. The inter-rater reliability for overall performance was particularly high, as all of the 16 significant cases displayed significant correlations greater than 0.700. The high agreement between teacher-and-researcher pairs indicates that the Diagnostic Rubrics are a fairly reliable tool for CL teachers to use to diagnose the speaking ability of P1 students in the classroom context.

On the other hand, as shown in Table 12.1, there were still a small number of teacher-and-co-rater pairs that did not seem to have given consistent ratings. The major disagreement seemed to pertain to vocabulary and pronunciation, as there were, respectively, 7 (35%) and 6 (30%) cases in which the inter-rater reliability was not significant. Even within the significant cases, the inter-rater reliability of two teacher-and-co-rater pairs was below 0.600 for vocabulary. Taken together, the findings suggested that the indicators related to linguistic form, unlike communication and overall performance, would require further calibration to ensure consistent understanding. At the completion of this first round of the validation study,

[2]Because the limited size of the research team, each member of the team visited multiple teachers' classes and rated multiple groups of students. Previously, during the process of developing the Diagnostic Rubrics, the inter-reliability among project team members was 0.849 (vocabulary), 0.862 (grammar), 0.820 (pronunciation), 0.868 (communication), and 0.882 (overall). All were intra-class correlation coefficients.

Table 12.1 Distribution of ranges of intra-class correlation coefficients

	Non-sig. cases no. (%)	Sig. cases no. (%)					Total no. (%)
		<0.600	0.600–0.699	0.700–0.799	0.800–0.899	0.900–1	
Vocabulary	7 (35%)	2 (10%)	0 (0%)	4 (20%)	2 (10%)	5 (25%)	20 (100%)
Grammar	6 (30%)	0 (0%)	3 (15%)	1 (5%)	4 (20%)	6 (30%)	20 (100%)
Pronunciation	6 (30%)	2 (10%)	0 (0%)	5 (25%)	4 (20%)	3 (15%)	20 (100%)
Communication	4 (20%)	1 (5%)	1 (5%)	2 (10%)	7 (35%)	5 (25%)	20 (100%)
Overall	4 (20%)	0 (0%)	0 (0%)	2 (10%)	9 (45%)	5 (25%)	20 (100%)

an in-depth discussion was conducted between the participating teachers and the research team to find out possible reasons behind the disagreement (especially for vocabulary and grammar). Based on the feedback during the discussion, the descriptors for each indicator or diagnosing category at each level were further refined and finalized.

Diagnostic Activity Package (DAP)

Teachers' appropriate and effective use of the Diagnostic Rubrics arguably depends on students' classroom discursive participation. In other words, only when students make full use of their knowledge and skills in actual communication can their oral performance reflect their true level of oral proficiency and can this oral proficiency be reliably assessed by their teacher. To maximize students' classroom oral output and support teachers' classroom use of the Diagnostic Rubrics, we also developed the DAP with 33 curriculum-aligned interactive activities. Of the 33 oral diagnostic activities, the first three are about self-introduction, toys, and food, which P1 students should feel most comfortable talking about at the beginning of their first school year and CL teachers are supposed to focus their instruction on *Hanyu Pinyin*. The remaining 30 activities were designed to align with the topics of the 15 units of the P1 Chinese textbook (two activities for each topic/unit). All of the DAP activities were initially used by the 20 teachers who participated in the aforementioned validation study; their feedback was solicited to refine the activities.

Activity Outline

Table 12.2 shows the outline or major components of a DAP activity. It is the first of the two activities designed to be aligned with Unit 1 (*Going to School*) of the children's P1 Chinese textbook. As Table 12.2 shows, in addition to *Topic*, a DAP activity outline is comprised of *Activity Type*, *Content*, *Objective*, *Prerequisite Knowledge*, *Suggested Class Period*, *Length of Activity*, *Organization and Procedure*, *Resources*, and *Assessment*.

Table 12.2 Example of outline of a DAP activity

1.	Unit/topic 课文	Unit 1 going to school (diagnostic activity 1) 第一课《上学校》(口语诊断活动一)
2.	Activity type 活动类型	Elaboration and group interaction 话题展开、小组对话
3.	Content 活动内容	Locations in school that you like and do not like 喜欢和不喜欢的学校地点
4.	Objective 活动目的	Diagnose students' knowledge of the following words, sentence structures, and expressions, and their ability to use them for communication Location nouns: campus, canteen, library, etc. Verbs: eat, play, watch/see, etc. Adjectives: beautiful, yummy, sweet, good (in smell), interesting, etc. Sentence structures: clauses that begin with *because*; coordination sentences Expressions/speech acts: explaining, questioning, and giving opinions Pronunciations and tones 诊断学生使用下列词语、句型、表达方式及进行沟通的交流能力。 处所名词: 校园、食堂、图书馆等 动词: 吃、玩、看等 形容词: 美丽、好吃、甜、香、好玩等 语法句型: "因为"起头单句、并列句 交际表达: 解释、提问、表达意见 语音语调
5.	Prerequisite knowledge 活动所需先备知识	Nouns for different locations of the school 学校各个地点的名词
6.	Suggested class period 建议活动进行时间	As part of the aural/oral class period 配合听听说说课进行
7.	Length of activity 活动时间	30 min 30分钟
8.	Organization/procedures 活动安排/组织	Suggested grouping: three students per group [a]Procedures: sharing and questioning-answering in group; group presentation 建议人数:每组三名学生 [a]流程:小组分享与问答、呈现
9.	Recourses 活动资源	PowerPoint, question cards, checklist, pictures of different school locations 简报、提问卡、检查表、学校不同地点的图片
10.	Assessment 活动评价	Diagnostic rubrics 口语诊断量表

Note [a]Procedures are further specified in the DAP (see Activity Procedures below). All three components of the OPDT, including the DAP, are in Chinese for teachers. The examples here and below are presented in both Chinese and English for non-Chinese speaking readers of this paper

Activity Type informs teachers about what their students are supposed to do for an activity. There are altogether seven activity types in DAP. In addition to elaboration (话题展开) and group interaction (小组对话) shown in Table 12.2,

there are situational interaction (情境会话), picture-based interaction/talk (看图对话/说话), decision making (决策活动), game (游戏), and role play (角色扮演). They are all student-centered and communication-oriented activities to involve students in interactional use of the Chinese language. In the present example, students are expected to elaborate on a topic based on group interaction. They may express ideas, ask questions, give clarifications, etc. *Objective* provides guidance to teachers on what to focus on in the process of diagnosing based on their students' Chinese oral output. In this example, teachers are encouraged to focus on how competent their students are at interacting with each other by use of specific words and sentence structures; whether they are capable of performing particular speech acts; and how good their pronunciation and intonation are. *Prerequisite Knowledge* helps teachers figure out how much the activity fits their students' Chinese proficiency level, and what their students need to be warmed up for. Other information like *Activity Length*, *Suggested Class Period* to use the activity, and group size helps teachers evaluate the feasibility of the activity for a particular class, and if needed, make reasonable adjustments to the organization of the activity. To provide teachers with clear guidance on activity organization, the procedure of each activity is carefully designed and specified in the DAP with step-by-step directions (see Activity Procedures below for details). Activity resources including PPTs, pictures, worksheets, checklists, etc. are also prepared and attached to each activity. All these efforts are to make it possible for teachers to use the Diagnostic Rubrics effectively to diagnose students' performance while they are participating in the activity.

Activity Procedures

As indicated above, specific procedures with step-by-step directions are provided for each activity. Table 12.3 shows an example of how the DAP supports teachers to conduct an activity procedurally. It is the second of the two activities designed to align with Unit 2 (*Schoolbag Says…*) of the children's textbook. It shows how procedurally a decision-making activity is to be conducted on what stationery students would like to bring to school and why. As this example shows, each diagnostic activity typically begins with a warm-up session, in which students are exposed to a set of pictures to get prepared for participating in the group activity, such as words and basic sentence structures to be used. The class is then divided into groups; and tasks are then assigned to each group with clear task instructions prepared in the DAP for the teacher. The warm-up session is followed by teacher and/or student demonstration, so that students can get deeper insight into the nature of the activity, what they are supposed to do, and how to apply the required knowledge they have learned in the warm-up session.

Students then participate in the group activity during which they interact with each other in Chinese, and the teacher diagnoses target students' Chinese abilities against the Diagnostic Rubrics. In this example, students as a group need to make a decision on ranking the popularity of different stationeries through discussion.

Table 12.3 Example of DAP activity procedures

1.	Teacher presenting pictures and asking a few students to say the names and utilities of the stationeries in the pictures 教师展示图片并请几个学生轮流说出文具的名称和用途。
2.	Teacher distributing to each group a set of picture cards, a table for ranking stationaries, and a checklist 教师给每组学生一套图卡、文具排行榜表格和检查表。
3.	Teacher giving detailed task instructions Form groups of four with members numbered 1, 2, 3, and 4, respectively Each member to share for 2 min with reference to the following questions: If you are allowed to bring only one stationery to school, which one would you choose? Why? (Note: if a student does not like the stationaries on the picture cards, s/he is allowed to choose any other stationeries that s/he likes.) Why would you not choose other stationeries (choose one to give your explanation)? (Note: more capable students can be asked to additionally explain what they could possibly do if these unchosen stationaries turn out to be needed later in school.) Group discussion to decide which stationery is the best and the most useful, and then rank all the stationeries based on their levels of usefulness When one student is sharing, another student takes note on the checklist of the presenting student's choice and the number of reasons s/he explained 教师交代任务: 你们四人一组,一个是1号,一个是2号,一个是3号,另一个是4号。 每个人有2分钟的时间分享: 如果你只能带一样文具来学校,你会选哪一样?为什么?(注:如果学生不喜欢这些文具,可以选自己喜欢的其他文具来代替。) 为什么不选其他文具?(选一个文具来说)(注:程度较好的学生,要求他们解释不选其他文具的原因,还有需要用到它们的时候怎么办。) 过后小组讨论哪一样文具最好,最有用,并给文具排名。 说的同学要请另一个同学在检查表写明自己的选择、说了几个原因。
4.	Teacher demonstration: I would choose a pencil, because I can use it to do classwork I would not bring a ruler, because my pencil can be used as a ruler to help me draw a line I think a pencil is the best choice, because I can use it to both write and draw 教师示范如下: 我选择带铅笔。因为要做功课的时候,我可以用铅笔做功课。 我不选尺,因为我会用铅笔画直线。 我觉得铅笔最好,因为我可以用铅笔写字和画画。
5.	Teacher asking a student to demonstrate 教师请一个学生出来示范。
6.	Teacher diagnosing a group of students or circulating around the classroom 教师诊断一组学生或四处巡视。

(continued)

Table 12.3 (continued)

7.	Group presentation Each group member first tells his/her choice, and then provides reasons for the choice and describes the chosen stationery Group explains how the stationeries that their members chose were ranked Teacher invites other students to ask questions or comment on the presenting group's ranking result Teacher collects the checklist Teacher gives remarks on students' performance in this activity 教师请小组出来呈现。 各组员先说出自己的选择,并说明理由、形容所选择的文具的样子。 小组解释自己怎样为这些文具排名。 过后,教师请其他学生提问或评论该组的排名。 教师回收检查表。 教师总结活动。

To complete the task, they need to hold an argument, express opinions, and give explanations, which provide output for the teacher to diagnose their linguistic and communication competence in Chinese. It is notable that notes are also embedded in procedural descriptions to guide the teacher in adapting the difficulty level of one or more steps of the tasks to cater to the needs of individual students. As shown in this example, if a student is perceived to have demonstrated a stronger performance than others or be learning faster than others, the teacher can ask the student to explain why other stationeries were not chosen (over and beyond explaining why one was chosen, which is expected of every student) and how the student would handle the situation when those stationeries turn out to be needed in school.

Teaching Resources

In addition to specifying procedures for teachers to organize a DAP activity, teaching resources related to that activity are also provided. They include pictures, cards, worksheets, PowerPoints, checklists, etc. that can be directly used by teachers to generate student output for diagnostic purposes. Figure 12.1 shows an example of teaching resources provided for a DAP activity. This picture shows the typical schedule of a student; each section includes a clock face with the time indicated. Students are expected to work in pairs or small groups to ask and answer questions about the picture using appropriate words (e.g., 踢足球 play soccer; 做功课 do homework) and sentence structures (e.g., …每天几点… /when does … [do

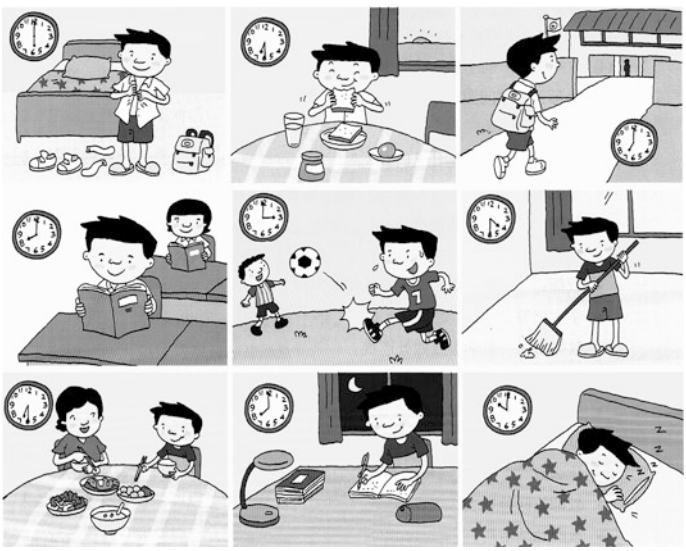

Fig. 12.1 Picture as a DAP teaching resource for teachers

something] every day?) to describe when the student does what every day. While this picture-based activity was designed for diagnostic purposes, it could also be used by the teacher as a pedagogical task to scaffold students to work on subsequent communicative activities in which students may ask each other authentic questions about what they themselves typically do at what time every day.

To sum up, with clear instructional procedures and resources, the DAP enables teachers to conduct in-class diagnosis of their students' oral performance as a natural or normal component of their everyday classroom teaching. As these activities are curriculum-aligned, they can easily fit in the classroom without interfering normal teaching practice. Teachers do not need to pull their students out for individual assessments, which would be very time-consuming and is often logistically challenging for CL teachers in Singapore.

Differentiated Instruction Activity Package (DIAP)

The OPDT project did not stop at helping teachers diagnose their students' CL performance. An important objective of the project was to help teachers make use of the diagnostic results of individual students so that they could benefit from differentiated learning opportunities for enhanced development of Chinese oral proficiency. To help CL teachers address students' weaknesses as informed by the diagnostic information, the DIAP, which includes a set of exemplar oral interactive activities, was developed with joint collaboration among the project team, experienced CL teachers, as well as master teachers affiliated with SCCL. A key characteristic of the DIAP is thus that all the activities are diagnosis-informed. More specifically, when we developed the DIAP, we referred carefully to the indicators (e.g., vocabulary and communication) and performance levels or developmental stages in the Diagnostic Rubrics. We then used our expert knowledge to develop exemplar activities that could address the challenges that were expected to be faced by some students during their learning of the 15 curricular units.

There were altogether 30 DIAP activities, with two for each of the curricular units. These DIAP activities fall into two types, with one more form-focused to highlight linguistic knowledge (hereafter, Type-I) and the other more oriented for interactive competence in communicative contexts (hereafter, Type-II). A rationale behind such a design is the enhancement of input (Wong 2005), which seeks to integrate traditional form-focused exercises into communicative contexts. For each of the 15 units from the P1 Chinese textbook, activities of the two types are paired, with Type-I allowing for essential practice of language forms and getting students ready for communicative use of them in Type-II activities.

Table 12.4 shows the two types of ability focus and their corresponding indicators in the Diagnostic Rubrics for the 30 DIAP activities across 15 curricular units. As the table shows, all the DIAP activities not only cover all key indicators but also consider the four developmental stages of competence as reflected in the Diagnostic Rubrics. In addition, when we designed the 30 activities, we also

Table 12.4 Ability focus of DIAP activities and corresponding indicators in the diagnostic rubrics

Unit	Type-I activity ability focus	Corresponding indicators in diagnostic rubrics	Type-II activity ability focus	Corresponding indicators in diagnostic rubrics
1.	Simple sentence	Grammar and sentence	Turning-taking and the turn of roles in a dialogue	Response and initiation
2.	Statement and simple question conversion	Sentence pattern	Enhancing the ability to initiate by asking questions	Initiation and sustaining a dialogue
3.	Continuous questioning and information supplement	Response	Situational question and answer	Sustaining a dialogue
4.	Comparative sentence	Sentence and function	Sustained output based on comparison	Sentence function and continuous utterance
5.	Using different question sentence structure to ask for one questions	Sentence functions	Questions and description	Sustaining a conversation with continuous utterance
6.	Asking questions	Sentence functions	Situational conversation with sentence pattern changes	Sentence function and sustained conversation
7.	Word order and sentence expansion	Grammar	Description with a chunk of meaning unit	Extended output
8.	Sentence expansion and word order	Vocabulary and grammar	Explaining with a chunk of meaning unit	Extended output
9.	Word order and sentence pattern change	Grammar and sentence	Report and quoting others	Extended output
10.	Different sentence structures for the same function	Sentence function	Appropriateness in communication	Naturalness and fluency
11.	Word order and compound sentence	Grammar and sentence function	Question, ask for clarification and sustained conversation	Topic expansion and extended output
12.	Choice of sentence structure	Sentence function	Group discussion	Negotiation of meaning and expansion of topic
13.	Word order and collocation	Vocabulary and grammar	Asking for clarification and questioning	Turning-taking and negotiation of meaning
14.	Word class and collocation	Vocabulary	Descriptive chunk	Order of meaning units
15.	Variety of sentence pattern	Sentence function	Group discussion and coherency	Quality and quantity in an extended output

consulted the curricular objectives of the 15 units. This means that all of the activities for linguistic forms and functions or communication are aligned with the objective designations in the P1 CL curriculum in Singapore. Naturally, the complexity of language forms and use in these DIAP activities increases across the 15 units. In other words, the DAP activities can be regarded as representing a general progression of growing competence (e.g., linguistic sophistication and complexity of language functions) across the 15 units in an academic year.

Because they are closely curriculum-aligned, the DIAP activities can be easily integrated into daily classroom teaching while highlighting particular language points or skills that students may be diagnosed to be underdeveloped. On the other hand, it is noted that intended to be exemplars, the two activities in each unit necessarily could not address the differential learning needs of all students, and thus could and should not be used for all students without any adaptations. When we developed the DIAP, we were clearly aware that there was no way that we could capture all (unique) needs of students across all P1 classrooms in Singapore. Consequently, while each exemplar activity provides some space for adaptation (e.g., different degrees of intensity and various forms of practice with the targeted learning points) by teachers, they are expected to develop their own interventional activities to expand the capacity of the DIAP for differentiated teaching and meet the unique situation or set of diagnosis results among their students.

Table 12.5 shows an exemplar DIAP activity that addresses Type-I ability (i.e., language forms). It illustrates how an activity targets a specific diagnosed error in students' use of place adverbials in Chinese. In English, place adverbials typically follow a verb, such as in the sentence *I study in the library*. In Chinese, however, a place adverbial usually appears between the subject and a verb, such as 我 (*I*) 在 (preposition *in/at*) 食堂 (*canteen*) 吃饭 (*eat*). Likely due to negative transfer from English, Singaporean students who come from an English-speaking family often make an error of putting the prepositional phrase "在 (in/at)…," which indicates where an action happens, after a verb or a verbal phrase.

To address this predicted challenge in Chinese learning, we designed this form-focused language game as a DIAP activity. As shown in Table 12.5, the game is separated into several steps. The whole class is first divided into groups of three students. Then, a set of flash cards with words of a different category are given to each group member. As shown in the table, Student 1 would have words of animated nouns such as people or animals; the cards taken by Student 2 would show words of actions such as *read*, *sing*, and *eat*; and Student 3 would receive a number of words indicating places such as *canteen* and *library*. Finally, the three group members interact with each other in the modeled format with the help of the keywords on the cards in their hands. This game, through constructing funny sentences, thus facilitates students' learning of the correct order of the subject (animated nouns; e.g., *the schoolmaster*), place adverbials (e.g., *in the toilet*), and the action verb or verbal phrase (e.g., *made a handstand*).

Table 12.5 Exemplar DIAP activity

Grouping and activity instructions 分组与活动说明	Teacher forms groups of three students, and informs the class that they are going to play a game called "*Our Sentence is the Funniest*." Each group member gets a set of cards with words of a particular category (i.e., *something/somebody, somewhere, and doing something*), and asks questions following the format below Student 1 (word in hand about *something/somebody* indicated by X): Where is X? Student 2 (word about *somewhere* indicated by Y): X is at/in Y Student 1 (to Student 3): What is X doing? Student 3 (word in hand about *doing something* indicated by Z): X is doing Z at/in Y After everyone makes at least one full sentence with words of all three types (like what Student 3 gives), the group discusses which sentence is the funniest 学生分三人一组,安排为1、2、3,说明要玩一个游戏:"我的句子最有趣"。每人分得一组词卡,学生可以就自己手中词卡上的词汇(某人/物、{在}某处、{做}某事)进行提问,形式如下: 学生1(手中的是主语):X在哪里? 学生2(地点词):X在Y(地点)。 学生1(对学生3): X在做什么? 学生3:X在Y做Z(事)。 每人都至少造一个句子后,就讨论决定哪个句子最有趣。
Demonstration 示范	Teacher invites a group to demonstrate the game with three rounds in which the three sets of word cards rotate among the three members. In each round, each member works with a word of a different category. After three rounds, the group decides which of the three sentences is the funniest 教师请一组学生出来示范,确定学生1、2、3。在三轮示范中,他们轮流抽取不同范畴的词语。问答示范结束后,请学生讨论并决定哪个句子最有趣。
Tips 小贴士	1. Design of word cards In order for students to construct funny sentences, the words for *somebody/something* can initially be from the textbook and then include other things or people that students are familiar with. The words for *somewhere* and *doing something* can be new words annotated with *pinyin*, such as canteen, bus station, change clothes, swim, etc. 2. Differentiation A slip of paper with the sentence frame (i.e., [who] in/at [where] [do what]) can be provided for students with low oral proficiency; they can produce a sentence by filling in the slots with appropriate words. Students with strong oral proficiency can follow the procedure of the game and give their sentences, and then be encouraged to construct additional sentences on paper with different but reasonable words using either characters or *pinyin*. They are encouraged to add modifiers to construct expanded or more complex sentences, such as *a beautiful fairy*, *beside a quiet pond*, *eating yummy chicken rice*, etc. 1. 词卡的设计 为了让学生设计出有趣的句子,开始时的"某人/某物"词卡可以是书上的人物,过后可以用身边的人物;"某处"和"某事"的词可以是注上汉语拼音的生词,例如食堂、车站、换衣服、游泳等。 2. 学习的差异性 如果是口语能力较弱的学生,可以提供印好句型的纸条[(什么人)在(哪里)(做什么)],让他们跟着句型来填入词语后说话。如果是口语能力较强的学生,先要求他们按规划的要求说几个句子,然后鼓励他们自己在白纸上填入不同的、合理的词语(或汉语拼音),还加入修饰词来进行造句游戏,例如加入"美丽的仙女""安静的水池边""吃香喷喷的鸡饭"等。

It is noted that while students of all ability levels could engage in a DIAP activity, we also considered the capacity of an activity for differentiation for students with different levels of proficiency. In the game presented in Table 12.5, for example, students with stronger oral competence could be encouraged to discuss and decide on one or two sentences that they agree to be the funniest, while less capable students could just focus on practicing the basic sentence structure targeting the order of place adverbials. This activity also has a capacity to allow more capable students to practice expanded or more complex sentences on the target structure. For example, an adverb(s) could be added to modify a target verb, and an adjective (s) could be included to modify a target animated noun. There is also a possibility to get students to use conjunctives (e.g., 可是 but/however) to connect different parts of a compound sentence. If writing Chinese characters is a learning objective, students could also be asked to write out the sentences they and their peers produced while playing the game. In addition, the design of the flash cards also allows teachers to make adaptations. For example, different sets of cards can be prepared for the same set of words, with diverse combinations of Chinese characters, *pinyin*, and pictures, and then be given to different students with varied levels of word recognition ability.

Effect of Using the OPDT on Students' Oral Proficiency Development

To examine if using the OPDT, including the DIAP, could achieve its expected positive effect on enhancing students' oral proficiency development, an intervention study with an interval of about three months was conducted subsequently in four primary schools with 16 P1 teachers who had not participated in any previous stages of the OPDT project. Sixteen teachers and their P1 classes from four other schools served as the control group in which no one had been exposed to the OPDT. In both the intervention and the control groups, two of the four schools were mission schools where, in the Singapore context, students tend to largely come from an English-speaking home; whereas, the other two schools were neighborhood schools where typically a significant proportion of students use Chinese as their major home language. Given the centralized education system in Singapore, all schools followed the same curriculum developed by the MOE. The teachers participating in the intervention were first trained to use the Diagnostic Rubrics, and their ratings were analyzed for inter-rater reliability. They were then trained to use the DAP for in-class diagnosis of students' oral Chinese competence against the levels and indicators in the Diagnostic Rubrics; and the DIAP to differentiate classroom instruction based on the assessment results. The teachers in the control group followed their regular instructional arrangement (i.e., business-as-usual group).

Table 12.6 Distribution of ranges of intra-class correlation coefficients

	Sig. cases no. (%)					Total no. (%)
	<0.600	0.600–0.699	0.700–0.799	0.800–0.899	0.900–1	
Vocabulary	0 (0%)	4 (25%)	4 (25%)	6 (37.5%)	2 (12.5%)	16 (100%)
Grammar	2 (12.5%)	2 (12.5%)	4 (25%)	8 (50%)	0 (0%)	16 (100%)
Pronunciation	4 (25%)	6 (37.5%)	3 (18.8%)	3 (18.8%)	0 (0%)	16 (100%)
Communication	0 (0%)	1 (6.3%)	5 (31.3%)	8 (50%)	2 (12.5%)	16 (100%)
Overall	0 (0%)	2 (12.5%)	3 (18.8%)	8 (50%)	3 (18.8%)	16 (100%)

Teacher Training in Using the Diagnostic Rubrics

After being trained in a workshop, the 16 P1 teachers participating in the intervention were asked to use the Diagnostic Rubrics to assess 14 randomly selected student speech samples from the pool of interviews collected at the initial stage of developing this tool. The coefficients of intra-class correlations among the 16 teachers were all statistically significant with a range of 0.705–0.752 across the diagnosing indicators. It means the teachers were able to assign highly consistent ratings among themselves for different indicators in the Diagnostic Rubrics. In addition, their ratings were also compared with those of the research team who developed the Diagnostic Rubrics for further inter-rater reliability analysis, following the same procedure that we used to calibrate the tool.

Like Table 12.1, Table 12.6 presents the distribution of different ranges of intra-class correlation coefficients for each diagnosis category. All of the correlation coefficients were significant. Specifically, the correlation coefficients varied from 0.600 to 0.931 for vocabulary, from 0.542 to 0.876 for grammar, from 0.479 to 0.819 for pronunciation, from 0.640 to 0.957 for communication, and from 0.669 to 0.942 for overall performance. The inter-rater reliability of all sixteen teachers' ratings was above 0.600 for vocabulary, and among them, inter-rater reliability of eight teachers' ratings was above 0.800. A similar pattern was also found for communication and overall performance. On the other hand, a few teachers seemed to have had difficulty in rating for grammar and pronunciation, especially the latter. The intra-class correlation coefficients of two teachers' ratings were as low as 0.580 and 0.542 for grammar. Similarly, in the case of pronunciation, the inter-rater reliability of four teachers' ratings was below 0.600. This result seemed to echo the findings of some previous studies, albeit on languages other than Chinese and in different contexts. For example, Hendricks and colleagues (1980; cf. Fulcher 2003) found in their study that the inter-rater reliability for pronunciation was as low as 0.43. In the Singapore context, teachers' relatively less prevalent agreement with the research team on grammar and pronunciation did not seem to be a big surprise, given that people may have different understandings about what constitutes correct grammar and pronunciation, and thus may have different levels of tolerance (and ratings) of grammatical use and pronunciation that seemed to "deviate" from the standard of the research team (Shang and Zhao, this volume). To reconcile any

Table 12.7 ANCOVA comparing post-assessment ratings between the control and the intervention groups with pre-assessment ratings as the covariate

	Control (N = 133)	Intervention (N = 139)	F
Vocabulary	8.93	8.96	122.66***
Grammar	8.72	8.80	118.64***
Pronunciation	10.08	9.85	115.37***
Communication	8.94	9.06	135.26***
Overall	8.94	9.06	152.85***

***$p < 0.001$

discrepancy that the intervention teachers appeared to have with the research team, a group discussion session was organized before the teachers were trained to use the DAP to elicit student output for classroom diagnosis purposes and the DIAP for differentiating instruction.

Pre- and Post-assessments

Experienced research assistants from the project team observed all 32 classes in the intervention and the control groups, and use the Diagnosis Rubrics to assess the oral language proficiency of a selected sample of students in each class. Altogether 139 students were sampled from the 16 classes in the intervention group with an average of about 8.7 students from each class; about 133 students were sampled from the 16 classes in the control group with an average of about 8.3 students from each class. For each diagnosis indicator (i.e., vocabulary, grammar, pronunciation, communication, and overall performance), the score ranged from 1 to 12.[3] Two rounds of assessments were conducted, one prior to the intervention and the other after the intervention, with an interval of about three months. In between these two assessments, the research assistants continued to observe all of the classes and record focal students' oral language output for subsequent qualitative analyses. Due to the limited space of this chapter, we only report the result of the quantitative analyses that compared the two groups. Table 12.7 shows the two groups' post-assessment ratings and the ANCOVA results with the pre-assessment ratings as the covariate. As shown in the table, after adjusting for the difference between the two groups prior to the intervention, the post-assessment ratings of the intervention group were significantly greater than those of the control group (all ps < 0.001).

[3]Within each level or stage (i.e., emerging, developing, blooming, and accomplishing), we further distinguished three sub-levels, including lower, middle, and upper. Thus, a score of 1 indicates lower emerging; that of 12 indicates upper accomplishing.

Conclusions

In this chapter, we reported a project conducted to address the challenges posed to CL teachers in Singapore where an increasing number of students come from English-speaking families and learn Chinese as a second language in school. In response to this change in the language background of CL learners, the Mother Tongue Languages Review Committee (Ministry of Education 2011) recommended that more pedagogical emphasis be put on oral proficiency development in early primary grades and that teachers adopt differentiated strategies to support oral language development of all students. Again this backdrop, the OPDT project was conducted to develop a set of tools that P1 teachers could use to conduct formative assessments of students' oral proficiency, diagnose individual students' learning needs in different areas, and then implement differentiated teaching. Specifically, the Diagnostic Rubrics allow teachers to conduct in-class diagnostic assessments of their students in regard to both linguistic and communicative competence at four levels. The DAP provides student-centered activities for teachers to engage students in communicative use of Chinese to generate output for reliable and effective diagnosis. Finally, the DIAP provides exemplar activities that teachers can directly adopt to enhance students' linguistic knowledge as well as communicative competence, or they can use these activities as references and their diagnostic assessment results as the basis to develop their own classroom activities for differentiated teaching. A subsequent intervention study with P1 teachers also confirmed that the OPDT was a useful and effective tool for CL teachers to use to inform instructional differentiation for Chinese learners with different learning needs.

While the project was conducted in Singapore with a focus on young learners, the implications for Chinese as a second language assessment and instruction are far beyond that particular learning context. As a matter of fact, the increasing diversity among learners of Chinese does not pertain to Singapore alone. Historically, the teaching and learning of Chinese in non-Chinese-speaking countries happened largely in tertiary institutions, and learners were primarily adult learners studying in a university-based Chinese program. However, with the increasing popularity of Chinese language and culture, particularly recently under the Chinese government's promotion of the Confucius Institutes and Confucius Classrooms (Zhao and Huang 2012; see also Teng, this volume), people from diverse backgrounds are learning Chinese globally. One notable characteristic of recent changes is the fast growth of Chinese programs in K-12 schools and an increasing number of young learners (Asia Society and The College Board 2008). Along with this increase of popularity of Chinese, however, are challenges for teachers like those faced by teachers in Singapore.

In the United States, for example, there is typically no centralized regulation on Chinese offerings at different grade levels across schools in a school district, not to mention across districts. Although ACTFL proficiency guidelines (ACTFL 2012) provide a framework for curriculum development, assessment, and instructional planning (in many states, there are state World Language standards, which can also

be a framework of reference), in practice, articulations on what should be learned at which stage are rare (Asia Society and The College Board 2008). Teachers usually develop their own curriculum and teach in their own way to meet the needs of a particular class of students. While this flexibility in curriculum and instruction is not contested, it often results in a conundrum when students move up the ladder in their Chinese learning in that a high-grade CL class could be comprised of students with very different learning histories and consequently different learning needs. The challenge for teachers is often doubled up with students from different ethnic or cultural backgrounds. It is not uncommon to see heritage language learners of Chinese, English Language Learners learning Chinese as their third language, as well as English-speaking students learning Chinese in the same classroom. To address these challenges, differentiation is arguably the key; and to make differentiation appropriate and effective, it is essential, as is evidenced in the OPDT project, that teachers know how to diagnose the strengths and weaknesses of individual learners and subsequently use assessment information to inform their pedagogical adaption or differentiation (Blaz 2006; Reese 2011). In this respect, the Singapore experience reported in this chapter certainly sheds light on Chinese as a second language education in other contexts and informs classroom-based formative assessment, instruction, as well as professional learning of CL teachers.

Acknowledgements The authors would like to thank the Singapore Centre for Chinese Language for granting the permission to reproduce in this chapter example activities from its copyrighted book entitled *Primary 1 Chinese Language Oral Proficiency Diagnostic Tool: Diagnositic Activity Package* (小学一年级华文口语能力诊断工具—诊断活动).

References

ACTFL. (2012). *ACTFL proficiency guidelines 2012*. Alexandria, VA: American Council on the Teaching of Foreign Languages.

Asia Society and The College Board. (2008). *Chinese in 2008: An expanding field*. Retrieved on April 2, 2016 from http://asiasociety.org/files/Chinesein2008.pdf

Blaz, D. (2006). *Differentiated instruction: A guide for foreign language teachers*. New York: Routledge.

Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Education.

Li, L., Zhao, S., & Yeung, A. S. (2012). Chinese language reform in Singapore: Teacher perceptions of instructional approaches and curriculum implementation. *International Journal of Bilingual Education and Bilingualism, 15,* 533–548.

Ministry of Education. (2011). *Nurturing active learners and proficiency users: 2010 mother tongue languages review committee report*. Singapore: Ministry of Education.

Pakir, A. (2008). Bilingual education in Singapore. In J. Cummins & N. H. Hornberger (Eds.), *Encyclopedia of language and education Volume 5: Bilingual education* (2nd ed., pp. 91–203). New York: Springer.

Reese, S. (2011, August). Differentiation in the language classroom. *The Language Educator*, 40–46.

Shepherd, J. (2005). *Striking a balance: The management of language in Singapore*. Frankfurt: Peter Lang.

Wong, W. (2005). *Input enhancement: From theory and research to the classroom*. New York: McGraw-Hill.

Zhao, H., & Huang, J. (2012). China's policy of Chinese as a foreign language and the use of overseas Confucius Institutes. *Educational Research for Policy and Practice, 9,* 127–142.

Zhao, S., & Liu, Y. (2010). Chinese education in Singapore: Constraints of bilingual policy from the perspectives of status and prestige planning. *Language Policy and Language Planning, 34,* 236–258.

Zhao, S. H., Liu, Y. B., & Hong, H. Q. (2007). Singaporean preschoolers' oral competence in Mandarin. *Language Policy, 6,* 73–94.