Chapter 3

Evaluation model for bilingual education programs

This chapter presents an evaluation model adapting Bhola's total literacy system for bilingual education programs. This thesis focuses on evaluating the selected bilingual education programs in the aspect of how much they are supportive to the students' CALP development. A checklist with a series of questions is designed based on the total literacy system and the CALP theory. In the checklist, there are eleven categories. Each category is a necessary part for a bilingual education program to be successful. There are two to five questions in each category, which guides the researcher to find out if the selected programs follow the concept of the basic principles of CALP development. At the end, the results of the evaluation are scaled by five levels: promotive; permissive; non-discriminative; tolerative; and prohibitive.

3.1 Program evaluation

In order to define this study as a program evaluation, it is necessary to explore the definition of program evaluation. Definitions of program evaluation have been discussed by many scholars. However, here are some examples of the definitions contributing to this study.

Firstly, a program evaluation determines whether the program is successful or not. Malone (2006) defines a program evaluation as a means for measuring a program against its initial objectives/goals. She suggests that "[a program evaluation] tells us: if the objectives [of the program] are being met and how well they are being met; which parts of the program are working and why they are working; which parts of the program are not working and why they are not working; if the program has actually helped the learners and the community as a whole" (2006:73). According to her statement, a program evaluation is a process of judging the program on whether 'the objectives are being met' (successful) or 'the objectives are not being met' (not successful) at each level. Therefore, her definition focuses on what answers, 'successful' or 'not successful', are brought by the program evaluation. Similarly, this study aims to determine whether the selected bilingual education programs are successful or not, based on whether the program builds up a CALP supportive program or not.

Secondly, a program evaluation can contribute to the program stakeholders in their decision making and documentations. A program evaluation here is a part of program management. Bhola (1990:9) mentions that "the essential objective of doing professional evaluation is to generate information that can be used in the planning and implementation of programs to improve the quality of life." He considers that a program evaluation is meant to contribute decision-making processes to the program. In terms of contribution to decision-making, this study, as it is an academic research, does not make any decisions or changes to the programs directly. However, it presents a model for program evaluation based on CALP theory.

The third definition of a program evaluation is that of a formal (defensible) process. Worthen, Sanders and Fitzpatrick (2004:5) define that "[a program evaluation is] the identification, certification, and application of defensible criteria to determine an evaluation object's value (worth or merit) in relation to those criteria." Their definition is based on technical processes. (Malone and Bhola's definitions focus on the outcomes and purposes of a program evaluation.) The formal process definition emphasizes the importance of identifying and certifying 'defensible criteria' in a program evaluation. The 'defensible criteria' are a set of standards to judge the quality of the program. Because of the defensible criteria, the evaluation results provide a clear reason why 'worth or merit' of the program is chosen. Following the definition, this study first sets up its 'defensible criteria'. Fundamental to these defensible criteria is whether the development of CALP has taken place.

Table 9 below summarizes the discussions in this section.

Table 9: Desirable features of a program evaluation and the characteristics of the evaluation model used in this study

Program evaluation (desirable)	Evaluation model used in this study
(1) Determining whether the program is successful or not (2) Contributing to the stakeholders (3) Identification, certification, and application of 'defensible criteria'	(1) Determining whether the selected bilingual education programs are supportive or not supportive of the students' CALP development (2) Proposing a theory-based bilingual education planning to the selected programs (3) The defensible criteria are whether the development of CALP has take place

3.1.1 Defensible criteria

In our daily life, we conduct many evaluations. For example, when someone chooses to buy a product from several choices, he/she compares prices, quality, designs and brands, before picking up the one that they think is the best. Sometimes, none of it is worth buying. That is also an evaluation. However, a program evaluation needs clearer reasons why the evaluator chooses a product that is deemed as 'worthy', rather than random 'likes or dislikes'. The clearer reasons are given by its 'defensible criteria'. The criteria also must be widely acceptable to all the readers and stakeholders. Therefore, the 'defensible criteria' are recognized as objective credibility for the program evaluation.

Malone's (2006) definition of a program evaluation suggests that the criteria are 'original objectives of the program'. The original objectives refer to expected outcomes of the program before it was planned. They are often associated with the needs of students, parents, community, and the local government. This approach is called an objectives oriented approach.[13] However, since the original objectives can be changed in each program's context, it is questionable if they are defensible in cases where the evaluation needs to compare more than two programs whose original objectives are provided by different stakeholders. Otherwise, when the two programs have the same outcomes, one program which has higher standards (expectations) is evaluated 'not successful' while the other one that has lower standards is evaluated 'successful'. Therefore, the original objectives are not very suitable for the defensible criteria of this study which examines two bilingual education programs in different locations. Actually, the original objectives themselves are one of the objects to be evaluated in this study (see Section 3.3.1).

---

[13] The objectives oriented approach is also introduced in Worthen, Sanders and Fitzpatrick (2004:71-87)

This study understands that the general goal of all bilingual education programs is helping the minority students to study effectively. In order to reach this goal, even a transitional bilingual education should support the development of the students' CALP. That is because the CALP development of students is claimed to be one of the most important indicators of a successful bilingual education program (Baker 2006; Cummins 2000, 2001). Malone (2006) presents a school plan, following the principles of CALP theory (see Table 10). This is an example of a CALP supportive transitional bilingual education program.

Table 10: An example of progression plan (Malone 2006:15)

K1	K2	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Develop oral L1 L1 as LOI	Develop oral L1 Introduce written L1 L1 as LOI – all subjects	Develop oral & written L1 Introduce oral L2 L1 as LOI – all subjects	Develop oral & written L1, oral L2 Introduce written L2 L1 as LOI – all subjects	Develop oral & written L1, L2 L1 as LOI for most subjects L2, with help from L1, for some subjects	Develop oral & written L1, L2 L1 as LOI for some subjects L2, with help from L1, for some subjects	Develop oral & written L1, L2 L1 as LOI for some subjects L2, with help from L1, for some subjects	Develop oral & written L1, L2 L1 as LOI for some subjects L2, with help from L1, as LOI for some subjects
100% L1	90% L1 10% L2	80% L1 20% L2	70% L1 30% L2	60% L1 40% L2	50% L1 50% L2	40% L1 60% L2	40% L1 60% L2

K:kindergarten, L1:first language, L2:second language, LOI:language of instruction

The defensible criteria of this study are the basic principles of CALP development: 1) Instruction in both languages; and 2) Long term transfer process. These basic principles are necessary for effective development of students' CALP in a bilingual education program. They are rooted in CALP theory and are therefore generally applicable to any bilingual program rather than specific to one program. The CALP theory is also proved by many empirical studies and supported by a number of international organizations, such as UNESCO (2005) and the New Zealand Ministry of Education (May et al. 2004). In other words, CALP theory is widely accepted in this field. Therefore, this study uses the basic principles as defensible criteria. It also evaluates the selected programs according to whether the evaluation object of each program follows the principles or not.

3.1.2 Evaluation objects

A list of the evaluation objects for literacy programs is presented by Malone (2006:73-74).[14] These are: program plan; curriculum/teaching method; personnel; training; materials; learner's progress; program growth; program's cost effectiveness; and long-term impact of the program on the community. The list shows that the evaluation objects are a combination between two types of objects. One is resources and the other is phenomena. Resources are visible objects, including human resources and materials. On the other hand, phenomena are invisible objects, such as influence and impacts.

Another list of the evaluation objects is found in Baker's (2006:312-316) discussions of key topics in effective bilingual education programs. He first points out that "dual language policies, provision and practices are a keystone of such [effective bilingual] schools." He, then, lists other key topics in effective bilingual schools: intake of students and language balance; staffing; shared vision, mission and goals among staff; staff professional development and training; leadership; curriculum; supporting ethos and environment; high expectations; and parents. His list is also a combination between resources and phenomena.

In light of Malone and Baker above, this study lists these evaluation objects: Chinese national/regional constitutions and laws related to language and education; school policies; documents and records of the programs; personnel (teachers, students, parents and professionals); training; curriculum; materials (textbooks and exercise books); extended reading materials and multimedia; lessons (contents, teaching languages and methods); displays in classroom; ideologies; motivations; language balance (in and out of classroom); supporting organizations. As it was discussed in the previous chapter, the listed evaluation objects can be categorized into Bhola's eleven subsystems.

3.2 Questionnaire

A questionnaire was designed with a list of questions based on the defensible criteria and evaluation objects (see Appendix 2 and 3). The translated Chinese version (Appendix 2) was used in the interviews with the local teachers. The English version

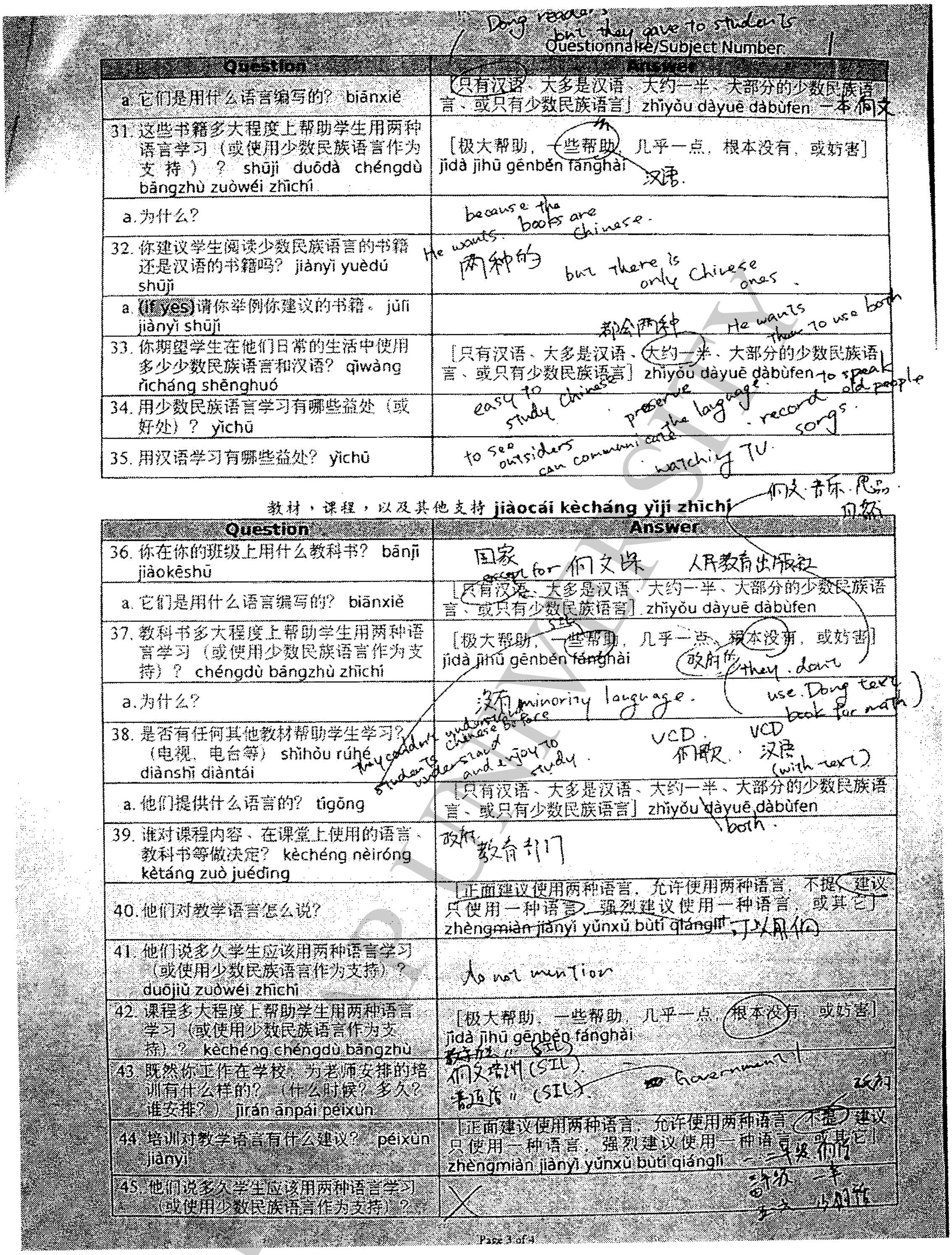
---

[14] Malone (2006) uses the term of 'literacy programs', including literacy programs for children and bilingual education programs.

(Appendix 3) was printed for the interpreters' reference. The interview method was chosen because when the teachers were unfamiliar with those questions, it was possible for an interviewer to clarify immediately. This helped the teachers to answer the questions candidly. Any additional information acquired during the interviews was also noted and used as needed for the purposes of the study.

The questionnaire contained 53 questions under three sections: in class, outside the class, and other support. The first 16 questions covered the background of the interviewees, such as age, nationality, teaching grade, speaking languages, and so on. This questionnaire provided adequate information for the interviews to determine the following: 1) which evaluation object supported the use of two languages for instruction, and 2) which evaluation object supported long term transfer process.

The school was asked to pick ten teachers to be interviewed according to the teachers' class schedule but as random as possible. Since one of the selected schools had only nine teachers, all of them were interviewed. One interpreter (who interpreted between English-Dehong Dai or Chinese-Dong) also attended the interviews. A second interpreter, an English teacher in a Dehong Dai school (English-Chinese), also helped to interpret occasionally. During interviews, the researcher read out the questions in Chinese. In order to avoid miscommunication, the teachers also were handed a copy of the questionnaire. The teachers answered either in Chinese or in their own language. When the teachers answered in their language, the interpreter translated to English or Chinese immediately. All conversations in the interviews were recorded by a voice recorder. The researcher and the interpreters also took notes on the answer sheet (see Figure 11).



Dong readers but they gave to students

Questionnaire/Subject Number 1

Question	Answer
a. 它们是用什么语言编写的？ biānxiě	[[✓ 只有汉语]、大多是汉语、大约一半、大部分的少数民族语言、或只有少数民族语言] zhǐyǒu dàyuē dàbùfen 一本侗文
31. 这些书籍多大程度上帮助学生用两种语言学习（或使用少数民族语言作为支持）？ shūjí duōdà chéngdù bāngzhù zuòwéi zhīchí	[极大帮助，[✓ 一些帮助]，几乎一点，根本没有，或妨害] jídà jīhū gēnběn fánghài 汉语.
a. 为什么？	because the books are Chinese.
32. 你建议学生阅读少数民族语言的书籍还是汉语的书籍吗？ jiànyì yuèdú shūjí	He wants. 两种的 but there is only Chinese ones.
a. (If yes)请你举例你建议的书籍。 jǔlì jiànyì shūjí	
33. 你期望学生在他们日常的生活中使用多少少数民族语言和汉语？ qīwàng rìcháng shēnghuó	都会两种 He wants them to use both [只有汉语、大多是汉语、[✓ 大约一半]、大部分的少数民族语言、或只有少数民族语言] zhǐyǒu dàyuē dàbùfen to speak
34. 用少数民族语言学习有哪些益处（或好处）？ yìchù	easy to study Chinese. preserve the language. record old people songs.
35. 用汉语学习有哪些益处？ yìchù	to see outsiders. can communicate. watching TV.

教材，课程，以及其他支持 jiàocái kècháng yǐjí zhīchí 侗文·音乐·民族团结

Question	Answer
36. 你在你的班级上用什么教科书？ bānjí jiàokēshū	国家 except for 侗文课 人民教育出版社
a. 它们是用什么语言编写的？ biānxiě	[[✓ 只有汉语]、大多是汉语、大约一半、大部分的少数民族语言、或只有少数民族语言] zhǐyǒu dàyuē dàbùfen
37. 教科书多大程度上帮助学生用两种语言学习（或使用少数民族语言作为支持）？ chéngdù bāngzhù zhīchí	[极大帮助，[✓ 一些帮助]，几乎一点，[✓ 根本没有]，或妨害] jídà jīhū gēnběn fánghài SIL 政府的 they don't use Dong text book for math
a. 为什么？	没有 minority language. they couldn't understand Chinese before. students understand and enjoy to study.
38. 是否有任何其他教材帮助学生学习？（电视、电台等） shìhòu rúhé diànshì diàntái	VCD 侗歌. VCD 汉语 (with text)
a. 他们提供什么语言的？ tígōng	[只有汉语、大多是汉语、大约一半、大部分的少数民族语言、或只有少数民族语言] zhǐyǒu dàyuē dàbùfen both.
39. 谁对课程内容、在课堂上使用的语言、教科书等做决定？ kèchéng nèiróng kètáng zuò juédìng	政府 教育部门
40. 他们对教学语言怎么说？	[正面建议使用两种语言，允许使用两种语言，不提，[✓ 建议只使用一种语言]，强烈建议使用一种语言，或其它] zhèngmiàn jiànyì yǔnxǔ bùtí qiánglì 可以用侗
41. 他们说多久学生应该用两种语言学习（或使用少数民族语言作为支持）？ duōjiǔ zuòwéi zhīchí	do not mention
42. 课程多大程度上帮助学生用两种语言学习（或使用少数民族语言作为支持）？ kèchéng chéngdù bāngzhù	[极大帮助，一些帮助，几乎一点，[✓ 根本没有]，或妨害] jídà jīhū gēnběn fánghài
43. 既然你工作在学校，为老师安排的培训有什么样的？（什么时候？多久？谁安排？） jìrán ānpái péixùn	[illegible] 〃 (SIL) 侗文培训 (SIL). 普通话 〃 (SIL). Government 政府
44. 培训对教学语言有什么建议？ péixùn jiànyì	[正面建议使用两种语言，允许使用两种语言，[✓ 不提]，建议只使用一种语言，强烈建议使用一种语言，或其它] zhèngmiàn jiànyì yǔnxǔ bùtí qiánglì 一二年级 [illegible]
45. 他们说多久学生应该用两种语言学习（或使用少数民族语言作为支持）？	X [illegible]

Page 3 of 4

Figure 11: An example of notes from the interviews

3.3 Customizing subsystems

As discussed in the previous chapter, Bhola's total literacy system is a useful tool for examining bilingual education programs. However, it still needs to be adapted to evaluate a bilingual education program in detail. This involves applying CALP theory as a criterion in the process of evaluating each subsystem of the total literacy system. Each subsystem was examined in a specific way. The evaluation checklist reflects this process (see Appendix 4).

It is noted that the eleven subsystems are interdependent. Though this process categorizes the evaluation objects into each subsystem, it is just for convenience. So, one evaluation object may be influenced by several subsystems.

3.3.1 Ideological subsystem

The ideological subsystem is often invisible as it represents human mind, beliefs, and expectations. Therefore, Bhola (1994) expresses the ideological subsystem as 'the soul of a literacy program' and 'a set of values to use in making choices'. Hilgendolf, Locnikar and Nichols (1996) consider that this ideological subsystem is "the basic (or underlying) purpose of a literacy program." They explain that the ideologies are cultural and economic ideas. Such ideas are related to the politics of the country, including spiritual goals. In cases where the people (or community) have any religious reasons why they want to (or want the others to) learn literacy skills, the religious ideas are also involved in the ideologies. The questions in this category ask about the underlying purpose and goal of the program, the national goals for literacy, and clearness of the goals.

Bhola (1994:158) also states that "the ideology of a literacy program is not always congruent with the ideology of the nation-state... teachers and others at the grassroots can make ideology come to life at the field level. On the other hand, they can subvert the program's ideology." This means that ideologies do not only come from national level (top-down), but also from the community level which involves school heads, teachers, parents, and possibly students themselves. The community level actors may have different ideologies from the national-level actors.

This study understands the ideological subsystem in two ways, which are 'the underlying purpose of the bilingual education program' and 'what people think how things should be'. First, this study researched the official documents that declare their purposes of the bilingual education programs to see if they clearly support the basic principles of CALP theory. Secondly, the interviews asked the local teachers questions such as "What languages do you think should be used in class?" and "How long do you think the students should have any special support by using minority languages?" If the teachers had CALP development supportive ideologies, this would be shown in their answers.

The listed questions in Table 11 were employed for examining this subsystem. Selected bilingual education programs in this research should answer all the questions from the list.

Table 11: Checklist (ideological subsystem)

1. Does the program state its underlying purpose of student's CALP development?
2. What languages do teachers think should be used in class?
3. How long do the teachers think the students should have any special support by using both languages of L1 and Chinese?
4. Others

3.3.2 Policy and planning subsystem

Policy and planning subsystem is different from the ideological subsystem. Ideology is behind (underlying) the policy, but policy and planning are outcomes of the ideologies. Bhola (1994:161) says, "The essential objective of a policy and planning system is, of course, to convert literacy ideologies into literacy policies. In turn, these policies must evolve into plans within a particular time-frame, using particular resources." The policy and planning subsystem moves a program in more direct ways than ideological subsystem does. It is often written or announced officially in constitutions, documents and meeting reports. Therefore, for the most part this study can use the library research on such official documents for this subsystem.

Bhola (1994:161) states that "yet too often the policy subsystem tends to be a weak subsystem. Policies are not always clearly defined; politics take over and planning becomes merely management of crises." He points out that the clarity of the policy is also an issue in this subsystem. However, it is difficult to say if it is a policy problem or organizational problem. This subsystem also overlaps many other subsystems. That is why the subsystem could affect literacy promotion (mobilization), organization and institutional issues, and publications.

To avoid confusion, this study focuses on the government policies and school policies which are officially announced on paper. It examines whether they clearly cover these two following points: 1) Do they support the use of two languages for instruction?; 2) How long (until which grade) do they support the program to use the two languages in education? Therefore, the questions in the checklist are also the combinations of the government/school policies and the two points (see Table 12).

Table 12: Checklist (policy and planning subsystem)

1. Do the government policies state the use of two languages for instruction?
2. Do the school policies state the use of two languages for instruction? (if it is different)
3. How long do the government policies support the use of two languages in education?
4. How long do the school policies support the use of two languages in education? (if it is different)
5. Others

3.3.3 Institution building and organizational subsystem

The institution building and organizational subsystem is one of the most difficult concepts to adapt to this study. That is because there is no such institutional structure that is more or less supportive of two languages of instruction and a long term transfer process. If this study focuses on the attitude of a particular organization, it will be just someone's ideologies and motivations. Therefore, this study considers this subsystem in a different way from the other subsystems. It is viewed as a linkage and consistency between the other subsystems.

This study presents a general overview of the whole program structure through library research and interviews. For example, even though the government tries to promote bilingual education, often the local schools do not understand what they should do in practice. There is obviously a miscommunication between the government and the local schools. It shows weak institutional structure. The quality of the linkage is the matter here.

At the same time, a series of questions pertaining to issues of whether 'there is or there is not' of each subsystem were also asked. That is because if a certain subsystem is missing in a program, the program will lose the linkage between the subsystems. Since bilingual education is often conducted as formal education, it may not be necessary to establish new institutions for the program such as supervising committees and material development committees in the local community. However, it needs to make clear who would take the responsibility of each role such as making decisions, providing professional support, conducting trainings, and so on. Therefore, these questions were asked in interviews: "Do you have any professional support?"; "Do you have the support of any institutions (government offices, universities, publishers, etc)?"; "Do you know any government policies talking about teaching languages?"; and "Who makes decisions on curriculum, languages used in class, textbooks, and so on?"

The checklist corresponds to the discussions in this section (see Table 13).

Table 13: Checklist (institution building and organizational subsystem)

1. Consistency of the use of two languages for instruction through the whole system.
2. Consistency of how long two languages should be used through the whole system.
3. Does the program have all the 11 subsystems?
4. Others

3.3.4 Mobilizational subsystem

Motivations are often indirect productions of other resources, such as language policies, teaching materials for teachers, bilingual reading materials, and so on. Teachers may be motivated to use two languages because their teaching materials are written in both. The students may be encouraged to read books in two languages when they find books available in both languages. That makes the mobilizational subsystem a wide ranging object and it is often related to the other subsystems. The motivations may also come from the community's encouraging environment to teach children in two languages such as their reading culture, economic needs, and so on.

The mobilizational subsystem could be both visible (examples such as a teacher actually suggesting to students to read books in both languages) and invisible (somehow teachers and students think it is not right to use L1 in class). Library research would help to find some efforts that the government or local schools have taken to encourage teachers and students to use two languages. That is also something visible. Observations and interviews examine the subsystem in both visible and invisible ways.

Interviews focus on how the program encourages teachers and students to use two languages. These questions were asked: "Are the students allowed to use both languages in class?"; "Do you suggest that students should read books in both languages (minority languages and Chinese)?" More direct questions asked for motivations to learn in the languages: "In what languages are the students interested in studying?"; "Are there any benefits (or motivations) of studying in minority languages?"; "Are there any benefits of studying in Chinese?" Observations were intended to support the answers from the interviews. In the observations, language use of teachers and students, the displays on the walls in classroom were observed.

For example, in China, often there is a sign, saying, “Speak in Standard Chinese.” It means “do not speak a local dialect of Chinese” or “use Standard Chinese correctly.” However, the students may misunderstand it as “do not speak minority languages.” This gives a negative result for the mobilizational subsystem.

The listed questions in Table 14 are a checklist of this subsystem. If any efforts (and obstructs) other than the list are found in the programs, they are counted in the 'others'.

Table 14: Checklist (mobilizational subsystem)

1. Are the students allowed to use both languages in class?
2. Do the teachers encourage the students to read books in both languages?
3. Do the teachers find any benefits for the students' CALP development from learning in both languages?
4. In which languages are the students interested in learning and how does it change in the long term?
5. Others

3.3.5 Professional support subsystem

Professional support can be considered in two ways. One is as human resources (professionals) and the other is as knowledge (theories). Bhola (1994:174) considers it as knowledge, saying, “We know today that professional 'knowledge' is needed for all the various subsystems of the total literacy system. Theory and research are necessary for policy and planning, institution building, mobilization, curriculum development, teaching, training, and evaluation.” However, this professional knowledge still needs someone professional who is able to integrate it into the program.

On the other hand, Hilgendolf, Locnikar and Nichols (1996) consider it as human resources. It is presented in the questions that they used in their research projects: “Are other professionals involved in this program? How are they involved?”; “Is it necessary to have a professional network you can contact?”; and “What inputs have other professionals given to your program?”

This study examines the professional-support subsystem in the aspect of whether the professionals (either human resource or knowledge) of the program follow the basic principles of CALP development. These following questions are asked in interviews: “Who provides you with the professional support (professors, government officers)?

How do they support?"; "What do the professionals say about teaching languages?"; and "How long do the professionals say the students should learn in both languages (or using minority language as support)?"

The questions in Table 15 are a checklist of this subsystem. The term 'professionals' includes both human resource and his/her knowledge.

Table 15: Checklist (professional support subsystem)

1. What languages do the professionals tell the teachers to use for instruction?
2. How long do the professionals tell the teachers to use both languages of L1 and Chinese?
3. Others

3.3.6 Curriculum development and programming development subsystem

Curriculum development and programming development subsystem put into practice what is decided in the policy and planning subsystem. While the policy and planning subsystem decides the direction of the programs, the curriculum development and programming development subsystem decides more specific processes. Bhola (1994:177) says, "[The curriculum development and programming development subsystem] chooses what is taught and determines the context of the teaching-learning process."

This study focuses on whether the curriculum of each program supports instruction in two languages and a long term transfer process. Baker (2006:315-316) also indicates that an effective curriculum planning includes "language and literacy development across the curriculum" and "smooth language transitions between grades." For example, bilingual education in China is often misunderstood as a minority language class being added onto the national curriculum as a separate subject. Therefore L1 is not used in other subjects, such as Chinese, Mathematics and Science. It is not supportive of the CALP development, because such curriculum does not support bilingual education across the curriculum.

This study examines the curriculum through library research and interviews. The library research handles the curriculum itself. The interviews asked the teachers the effects of the curriculum through the question: "How much does the curriculum help the students learning in both languages (or using minority languages as support)?" The listed questions in Table 16 are a checklist of this subsystem.

Table 16: Checklist (curriculum development and programming development subsystem)

1. What languages does the curriculum tell the teachers to use for instruction?
2. How long does the curriculum support using both languages for?
3. How much do the teachers think the curriculum helps the students to learn in both languages?
4. Others

3.3.7 Media and materials subsystem

Bhola (1994:178) says, "[The media and materials subsystem] aims to procure from outside or to produce within the system materials of all kinds for functional literacy programs and for teachers, learners and other facilitators for the pre-literacy, literacy and post-literacy stages of literacy work." His understanding of media and materials is much wider than textbooks. He adds (1994:179-180), "The media and materials subsystems must work with the media of newspapers, radio and television to support the literacy effort of the country."

However, the media and materials subsystem has to be differentiated from the post literacy subsystem. In this study, the subsystem focuses on textbooks and other materials used in class (Bhola calls them primer and follow-up books), including extra class readings (books that are provided for homework). It examines those materials in two aspects: what languages are used and until which grades are they provided. The checklist of this subsystem includes a question which asks the teachers' opinions of the textbooks (see Table 17).

Table 17: Checklist (media and materials subsystem)

1. In what languages are the textbooks written?
2. In what languages are other class materials written? (if there are)
3. Until which grade are the bilingual textbooks provided?
4. Until which grade are the other bilingual class materials provided? (if there are)
5. How beneficial are the textbooks in helping the students to learn in both languages according to the teachers?
6. Others

3.3.8 Orientation and training subsystem

The orientation and training subsystem is intended for teachers in most cases. Baker (2006:315) says, "Staff professional development can be designed to help all staff effectively serve language minority students. For example, staff development programs can sensitize teachers to students' language and cultural backgrounds, increase their knowledge of second language acquisition and help develop effective curriculum approaches in teaching language minority students." Training influences the teachers in the whole program. The teachers in China usually do not know about bilingual education until they are assigned to the schools in minority communities. Thus, it is important whether the training provides (or does not provide) the teachers with the principles of CALP development.

Training is not limited to externally conducted lectures. It also includes individualized self-learning by the teachers themselves. Bhola (1994) says, "The best a teacher of adult literacy can do at the field level in terms of training is to take responsibility for his or her own self-training and continuous growth."

Library research explored documents and reports recording the teacher training. The interviews also asked questions related to what kind of training the programs have done and if the training follow the principles of CALP theory. These are the questions for interviews: "Since you have worked in the school, what kind of training was arranged for the teachers?"; "What did the training say about teaching languages?"; and "What did they say about how long the students should study in both languages (or using minority language as support)?" The questions in Table 18 are a checklist of this subsystem.

Table 18: Checklist (orientation and training subsystem)

1. What languages do the training tell the teachers to use for instruction?
2. According to the training, how much time should teachers use both languages of L1 and Chinese?
3. Others

3.3.9 Teaching-learning subsystem

The teaching-learning subsystem is the core of a bilingual education program. The other subsystems all create supportive environments to the program, but it is the teachers who decide what they teach in class after all. Moreover, if the students do not appreciate the program and take advantage of it in their learning activities, the program can not be successful.

This subsystem is about teaching methods, lesson plans, teaching languages (oral and written), students' attitude in the class, students' language use, and so on. They are represented in these interview questions: "When you teach, which languages (and how much) do you use?"; "How do you use the languages in different ways according to the students' grades?"; "Do the students actually use minority languages in class?"; "How long do you actually give the students special support in minority languages?"; and "If the students ask you a question in a minority language, in what language do you respond?" If they use both Chinese and a minority language until Grade 6, it means the program is supportive to CALP development. Observations are also used to support the answers.

Table 19 is a checklist of this subsystem. The collected data from interviews and observations is summarized in each question.

Table 19: Checklist (teaching-learning subsystem)

1. What languages do the teachers use for instruction?
2. What languages do the students use in class?
3. How long do the teachers give the students any special support by using both languages?
4. Others

3.3.10 Post-literacy subsystem

The post-literacy subsystem represents the issues of post-literacy literatures for the students' further studies. Even though the students learn how to write and read in both languages in school, their CALP does not develop effectively unless they have enough opportunities to continue practicing their new skills. Thus, there is a great need for materials to help bridge the gap between basic textbooks and the world of general written matter. It must be noted that learning the skill of reading is differentiated from having enough materials to read and cultivating the reading habit. Without progress towards post-literacy materials, the reading skill is likely to be quickly lost. This is especially if the skill has been only half-acquired because it was presented through a national education system (Wendell 1982). Literature after basic literacy acquisition inspire the students to learn literacy skills and empower them through providing new knowledge.

This study understands the post-literacy subsystem as reading materials and other media beyond the lessons and homework. They are the books in school/community library where the students can read in their free time, including newspaper, magazines, and TV/radio programs which the students watch and listen at home. The study uses library research, interviews and observations in this subsystem. The questions are, "In public or school library in the village (or town), how many books are there for further studies?", "In what languages are they written?", and "How much do the books help the students learning in both languages (or using minority languages as support)?" The questions in Table 20 are a checklist of this subsystem.

Table 20: Checklist (post-literacy subsystem)

1. In what languages are the books in school or public libraries written?

2. In what languages are other materials, such as newspapers and TV programs written/made? (if there are)

3. How much do the teachers think the books help the students learning in both languages?

4. Others

3.3.11 Evaluation subsystem

Evaluation is not only a tool to summarize the results of a program. It also provides the information of needs in the community (initial evaluation) and encourages the staff to improve the programs (in service). In these ways, the evaluation subsystem makes the program healthy and effective. At the same time, it also reminds the program staff what the program values as important. Therefore, evaluating which languages of instruction are used in class and the length of the transfer process encourages the local teachers to keep following the principles of the students' CALP development. It is also important in what languages the evaluation is conducted. For example, the students are encouraged to learn minority languages when they have exams in both languages.

Library research examines what kinds of evaluations have been done and whether the students are evaluated in both languages. Interviews also help to know how the teachers evaluate the students and how the teachers themselves are evaluated. As Bhola (1994) shows that teachers are one of the most important evaluators, it is important to find what factors the teachers evaluate in the classroom. The questions are, "Who (students, teachers, supervisors, funders, etc) has evaluated your class and school?", "Which aspects of the class have been evaluated?", and "Have there been any evaluation is performed concerning teaching languages?"

The questions from the checklist of this subsystem (see Table 21) simply examine these two points: 1) Does the evaluation concern the languages of instruction?; and 2) In what languages is the evaluation performed?

Table 21: Checklist (evaluation subsystem)

1. Have there been any evaluation is performed concerning the languages for instruction?
2. In what languages are the evaluations/exams for the students performed?
3. Others

3.4 Scales (analysis and interpretation)

Another issue that should be discussed here is how to measure the collected data.

Baker (2006:285) points out the difficulties and problems of the former evaluations of bilingual education, saying, “A particular problem is that measures of success [in a bilingual education program] tend to be restricted to what is measurable. Quantitative outcomes (e.g. test scores) are used; qualitative evidence has rarely been gathered.” His question of ‘measurable’ data is involved in the differences of quantitative data and qualitative data. The quantitative data is numerical, which are represented in ‘yes-no’ and ‘how many’ questions. They are often directly and visually measurable. On the other hand, the qualitative data is not numerical, which is represented in ‘how well’ questions. Baker's statement insists that the qualitative data in a program evaluation is as important as the quantitative data.

As stated earlier in the chapter, this study seeks to determine whether the selected bilingual education programs are supportive or not supportive of the students' CALP development'. In order to measure the success of the programs, this study needs to answer not only 'yes-no' questions, but also 'how well' questions. Therefore, this study collects both quantitative and qualitative data.

The qualitative data, as well as quantitative data, can be measured by using dimensions to scale the level of 'how well'. For example, in his discussion of ideological subsystem, Bhola (1994:161) says, “It is not absolutely necessary to have a clear ideology to support literacy or even to have only one ideological justification for literacy promotion. All that matters is that literacy work be permitted and promoted.” He considers that the results can be scaled by the two dimensions: 'permitted/promoted' and 'not permitted/not promoted'. Another example is Hilgendolf et al. (1996) using three dimensions: fully implemented in program; partially implemented in program; and not implemented in program. Those are the examples that make qualitative data measurable.

This study expands on the scales used in the examples of Bhola and Hilgendolf et al. This is because bilingual education in China is often institutionalized as formal education, which means that all the programs are promoted at some levels. The two dimensions and three dimensions appraisal do not give clear differences. Therefore, in this study, each subsystem is scaled by five levels (1-5 points): promotive (5), permissive (4), non-discriminative (3), tolerative (2) and prohibitive (1). The idea of five levels is adopted from Zhou (2004).[15] He uses the scales to identify the levels of national language policies. However, it can be reflected in all the subsystems. That is why some interview questionnaires are set with multiple choices in five levels: strongly supportive/strongly suggested; supportive/suggested; neutral/permitted/a little; not at all; and obstructive. The score of each subsystem and the average of all the scores are presented as the answer of 'how well' the programs support the students' CALP development.

Table 22 is a rubric which was used in this study. It was designed based on the principles outlined in this section.

Table 22: Summary of the evaluation scales

	Promotive (5)	**Permissive (4)**	**Non-discriminative (3)**	**Tolerative (2)**	**Prohibitive (1)**
In which languages?	Both in L1 and Chinese	Both in L1 and Chinese, but one is more than the other	Both languages can be used, but one is only for trifle issues / Not mentioned (positive to using both languages)	Both languages can be used only when it is necessary / Not mentioned (negative to using both languages)	Only L1 or only Chinese
How long?	More than 5 years / Until Grade 6	3-5 years / Until Grade 3-5	1-2 years / Until Grade 1-2 / Not mentioned (positive to long term)	Less than 1 year / Supplemental / Not mentioned (negative to long term)	None
How much?	Strongly supportive / Strongly suggested / Very much	Supportive / Suggested / Positive	Neutral / Permitted /A little	Not at all	Obstructive

---

[15] Zhou (2004) also uses another dimension, 'overt' and 'covert', as an axis of coordinates. However, this study only applies the five scales.

3.5 Summary

Figure 12 is an image of the evaluation model that was used in this study. It consisted of three dimensions to determine whether the bilingual education programs are CALP supportive or not. The first dimension is based on two principles of CALP development. The second dimension is from Bhola's eleven subsystems which were used as categories that could be checked through each program. The last dimension refers to the five scales (promotive, permissive, non-discriminative, tolerative, and prohibitive). This allowed the study to evaluate not only supportive or not supportive, but also 'how much' each program is supportive.

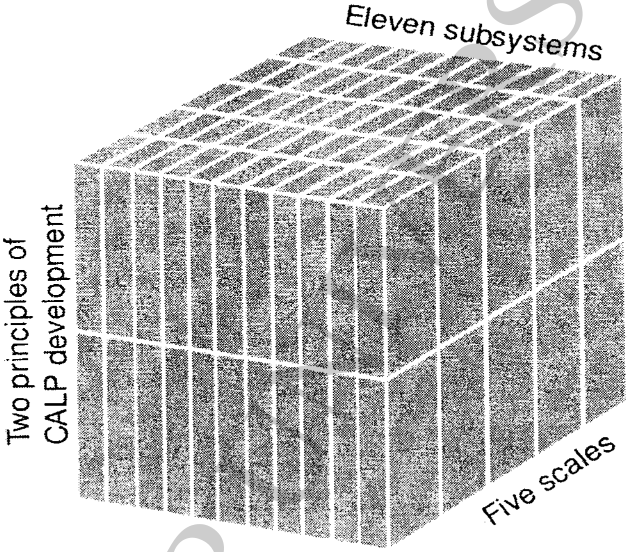


Figure 12: An image of the evaluation model