The Fine-tuning Medium-of-instruction Policy in Hong Kong: A Case Study of the Changing School-based Test Papers in Science Subjects

Jim Yee-Him CHAN

Division of English Language Education, Faculty of Education
The University of Hong Kong

This study examines changes in the school-based Chinese vis-à-vis English test papers in two content-area science subjects (i.e., integrated science and computer and information technology) in a secondary school in Hong Kong during the implementation of the new medium-of-instruction (MOI) initiative, in that schools were shifting to adopting content-based English-medium instruction (EMI) teaching at the junior secondary level. The analysis of the assessments was via two main stages. First, a descriptive approach was adopted to investigate any observable changes in the assessment designs and formats. Second, all the test items in the test papers were evaluated in the dimensions of cognitive and knowledge categories using the revised Bloom's taxonomy as the analytical framework. The findings reveal that there was a reallocation of question types and pedagogical foci as well as a decrease in score proportions and varieties of questions assessing higher-level cognitive processes and knowledge in the revised test papers. This article concludes by highlighting the possible effects brought about by the revised school-based assessment and the need to continue to investigate the influence of the new policy.

Keywords: assessment; content-based teaching; medium of instruction; fine-tuning policy; revised Bloom's taxonomy; Hong Kong

Introduction

Since the era of globalization, the adoption of English-medium instruction (EMI) in content-area subjects has become increasingly popular worldwide due to the social aspiration that English immersion could lead to learners' higher English proficiency (e.g.,

Ferguson, 2006; Hamid, Nguyen, & Baldauf, 2013). While English has traditionally been the language for international communication especially in scientific fields (Ammon, 2010), the globalization of university education has further favored EMI as it enables academic exchanges and attracts foreign students (Doiz, Lasagabaster, & Sierra, 2013). Even though there is a recent trend that mother-tongue education has been promoted at the primary level in most countries around the world such as in Africa, Asia and Europe due to pedagogical reasons (Okebukola, Owolabi, & Okebukola, 2013), the choice of the medium of instruction (MOI) at the transitional secondary (and sometimes even senior primary) level has remained a critical task for policy makers considering the perceived economic power and pragmatic values of English as an international language (see Crystal, 2003).

Regardless of the policy implementation in diverse sociolinguistic contexts, one major challenge emerging in the literature on bilingual education has been the question of how to enhance students' English learning by adopting content-based EMI education while, at the same time, ensure their learning outcomes in academic subjects (Ferguson, 2006). In these subjects, the nurturing of one's higher-order thinking skills and academic knowledge tend to be most crucial. This dilemma of achieving the dual goal of enhancing students' English and academic capability is particularly pronounced in science education as it entails both the learning of the concept of science inquiry and the acquisition of an academic language that diverges from students' everyday language (Rollnick, 2000; see also Gee, 2008). Rollnick (2000) describes such process of studying science through a second language as being "initiated into two social practices at once" (p. 100). In many cases, learners confront the demands of academic learning via a language that they have not yet fully mastered while teachers lack the knowledge and institutional supports required to address students' educational needs (Lee, 2005). From this perspective, research has revealed that positive teaching/learning outcomes are largely subject to conditions such as students' age, cognitive levels and curricular expectation, teachers' expertise and experience in teaching, and the available resources and teaching time (Stoller, 2008).

Apart from the aforementioned factors that affect the outcomes of bilingual education, significant scholarly work has underlined the profound impact of assessment designs on the accomplishment of educational goals particularly via washback, that is, how examination or testing influences the teaching and learning process, at both micro (e.g., students, teachers, parents, test developers, curricula, institutions) and macro levels (e.g., teaching methodologies by teachers, learning strategies by learners) (e.g., Ahmad & Rao, 2012; Cheng, 1997; Gardner, 2006). The effects of washback can be both positive and negative depending on the

assessment forms. For instance, summative assessment (or "assessment of learning," see Birenbaum et al., 2006) is often argued to be associated with surface-learning approaches involving learners' passive acceptance of information rather than true understanding of its meaning whereas formative assessment (also "assessment for learning," see Berry, 2008; Black, Harrison, Lee, Marshall, & Wiliam, 2003) tends to promote deep or meaningful learning (Brooks, 2002; Gardner, 2006).[1] Although the latter has generally been recognized as an effective means to achieve more desirable learning outcomes (e.g., Berry, 2008; Black & Wiliam, 1998; Wiliam, Lee, Harrison, & Black, 2004), the former has traditionally been, and still is, a major component in the education system serving as end-of-key-stage tests for evaluative, selective and screening purposes (Brooks, 2002; Birenbaum et al., 2006). Under these circumstances, the influence of the assessments on teaching and learning practices is likely to be more prominent especially if they are more high-stakes external or nation-wide public examinations (see Brooks, 2002). Given the multiple teaching/learning objectives in bilingual science education, it is the purpose of the present study to investigate how these diverse educational goals are reflected in teaching and learning practices, particularly regarding summative assessments, at the secondary school level in Hong Kong, where the effectiveness of EMI bilingual education, like other postcolonial countries (e.g., Malaysia, see Hashim, 2009; Singapore, see Dixon, 2009) or regions that adopt bilingual immersion programs for foreign language acquisition (e.g., Canadian French immersion in Canada, see Johnson & Swain, 1997; dual language education in the U.S., see Lindholm-Leary, 2001), has been a controversial issue.

The Changing MOI Policy in Hong Kong

In the international city Hong Kong, the tension between the societal demand for better English education and pedagogical considerations has played a crucial role in MOI policy development at junior secondary level (Secondary 1–3) over the past few decades (Poon, 2013). Whereas the colonial government's laissez-faire policy[2] allowed secondary schools to decide on their own school-based MOI arrangement before Hong Kong's return to Chinese sovereignty in 1997, the Hong Kong Special Administrative Region (HKSAR) government implemented a compulsory mother-tongue streaming policy after the handover (1998–2009), which resulted in only around a quarter of the secondary schools (114 of 421) that remained EMI schools (Evans, 2013). The majority of the schools were compelled to switch to Chinese-medium instruction (CMI) for teaching content-area subjects at junior

secondary level. According to the HKSAR government, this change in the MOI policy was mainly on the basis of educational grounds, namely, the perceived undesirable classroom practices (e.g., code-mixing/switching) and teaching and learning outcomes when EMI teaching was applied for all students who had diverse language proficiencies and learning ability (Education Department 1997; see also Education Commission, 1990). Despite this educational intention, this clear-cut EMI/CMI divisive policy was subsequently found to be unpopular with the general public as well as educational practitioners as it has apparently created a labelling effect in that CMI schools were generally perceived to be inferior to the EMI counterparts (Choi, 2003).

Since the 2010–2011 academic year, Hong Kong's education system has entered a new phase in which the Education Bureau (2010) has implemented the fine-tuning MOI policy that sought to remove the "labels" of EMI and CMI secondary schools. More specifically, this major initiative offers greater flexibility to individual schools to determine their own MOI arrangements based on the language competencies of the students and teachers and the support measures from the schools (Education Bureau, 2010; Education Commission, 2005). Subsequently, school administrators could decide on the proportions and modes of EMI or CMI teaching in a particular subject and class based on their own pedagogical and practical needs.[3] Given the high demand for bilingual education in Hong Kong, one consequence revealed in Kan, Lai, Kirkpatrick, and Law's (2011) recent survey is the increase in proportions of EMI teaching in former CMI schools, particularly in subjects such as Mathematics and Science probably because studying them in English is considered by the school administrators to be more beneficial to higher education and possess fewer linguistic challenges (see J. Y. H. Chan, 2013). However, the major concern remains whether studying a subject in a second language might undermine learners' academic performance (see Lo & Lo, 2014).

To date, there is little empirical research into the effectiveness and outcomes of the fine-tuning policy, although investigations regarding the previous EMI/CMI divisive policy tend to suggest that EMI teaching causes more emphasis on memorization and copying (e.g., Lo & Macaro, 2012; Ng, 2007; Yip, Coyle, & Tsang, 2007) and students' lower self-concept in science subjects than the CMI counterparts (Lau & Yuen, 2011; Yip & Tsang, 2007; Yip, Tsang, & Cheung, 2003; see also Fung & Yip, 2014). While there is a sign that these results have resurfaced after the introduction of the new policy based on findings derived from typical research means such as questionnaire surveys and interviews (e.g., J. Y. H. Chan, 2014; Poon, Lau, & Chu, 2013), it seems to be also important to explore other aspects of

school practices in response to this policy shift. One example is Lin (2013), who has illustrated some teachers' innovative work (e.g., teaching materials) using plurilingual pedagogies.

Another less-investigated but perhaps even more pertinent area in school practices is the school-based assessment in Hong Kong, which exerts great impact on students' everyday learning. As the fine-tuning MOI initiative is implemented at the junior secondary level, it tends to have less direct influence on the high-stakes public examination in the end of secondary education because schools have their own choice of MOI in content subjects at the senior secondary level (see J. Y. H. Chan, 2013). Although the washback effect of internal assessments at the junior level are expected to be less salient than the external counterparts, these assessments are often summative for the purpose of selection and grade promotion rather than supportive means for learning (Berry, 2011). According to Berry (2011), even though the importance of assessment for learning (or formative assessment) has been highly emphasized by the government, there seems to be few changes in the schools' assessment practices due to Hong Kong's deep-rooted examination-oriented culture. Nonetheless, not only is junior secondary education important for establishing students' academic foundation, but the internal test papers also play a significant role in bridging students' achievement gap between the junior and senior levels and have great impact on their learning motivation in specific subjects (J. Y. H. Chan, 2013; Fung & Yip, 2014).

Therefore, the present study centers on the changes of a school's tailor-made test papers in their chosen EMI science subjects before and after the fine-tuning. Aiming for a more systematic analysis, the investigation adopted the revised Bloom's taxonomy, which elicited the underlying assessed categories of cognitive processes and knowledge in the test papers. These two dimensions are particularly relevant to the learning objectives in science subjects. For instance, the syllabus for science education suggests that junior secondary students should acquire basic scientific knowledge and develop skills to enquire and to solve problems (Curriculum Development Council, 1998). Many of these knowledge (e.g., phenomena, facts and concepts in science; scientific vocabulary and terminology; applications of science) and skills (e.g., classifying, inferring, predicting, interpreting data) largely correspond to the evaluation criteria in the revised Bloom's taxonomy.

The Revised Bloom's Taxonomy

The original *Taxonomy of Educational Objectives* was published by Bloom and his

colleagues in 1956 (Bloom, 1956), serving as "a framework for classifying statements of what we expect or intend students to learn as a result of instruction" (Krathwohl, 2002, p. 212). It provides carefully developed definitions for six cognitive categories, that is, "knowledge," "comprehension," "application," "analysis," "synthesis," and evaluation," in which the latter ones are considered the more desirable educational goals (Krathwohl, 2002; see also Mayer, 2002). Since its introduction, this hierarchical framework has been widely applied to designing (e.g., Crowe, Dirks, & Wenderoth, 2008) and/or evaluating the range of (e.g., Swart, 2010) curricular objectives and test items regardless of subject matters and grade levels. Forty-five years after the development of the original taxonomy, a revised version was devised to incorporate a dual perspective on learning and cognition instead of adopting the original single dimension approach (Anderson & Krathwohl, 2001). More specifically, the revised taxonomy allows two aspects, a noun and verb, to form the basis for the dimensions of cognitive processes (i.e., "remember," "understand," "apply," "analyze," "evaluate," "create") and knowledge (i.e., "factual," "conceptual," "procedural," "meta-cognitive") respectively (Krathwohl, 2002). Each aspect of these two dimensions includes a corresponding subset of thinking processes or subtypes into which specific educational objectives can be categorized. For example, "recognizing" and "recalling" are two cognitive processes under the category of "remember" (see Table 1) and "knowledge of terminology" and "knowledge of specific details and elements" are two subtypes of "factual knowledge" (see Table 2). Some of the distinct advantages of this design over the original one as well as some others such as the Structure of the Observed Learning Outcomes (SOLO) taxonomy (see C. C. Chan, Tsui, Chan, & Hong, 2002) are its higher level of clarity for analysis and its refocused educational needs and incorporated new knowledge and thoughts that serve better contemporary society (Anderson & Krathwohl, 2001).

Table 1:   The Cognitive Process Dimension

Category	Remember	Understand	Apply	Analyze	Evaluate	Create
Cognitive process	• Recognizing • Recalling	• Interpreting • Exemplifying • Classifying • Summarizing • Inferring • Comparing • Explaining	• Executing • Implementing	• Differentiating • Organizing • Attributing	• Checking • Critiquing	• Generating • Planning • Producing

Sources: Adapted from Anderson and Krathwohl (2001).

Table 2:   The Knowledge Dimension

Type	Subtype
Factual knowledge	• Knowledge of terminology
	• Knowledge of specific details and elements
Conceptual knowledge	• Knowledge of classifications and categories
	• Knowledge of principles and generalizations
	• Knowledge of theories, models, and structures
Procedural knowledge	• Knowledge of subject-specific skills and algorithms
	• Knowledge of subject-specific techniques and methods
	• Knowledge of criteria for determining when to use appropriate procedures
Metacognitive knowledge	• Strategic knowledge
	• Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge
	• Self-knowledge

Sources: Adapted from Anderson and Krathwohl (2001).

The Study

Research Design

The main objective of this study was to explore the changes in the design and pedagogical foci of junior secondary test papers during the implementation of Hong Kong's fine-tuning MOI policy in a typical former CMI secondary school as a case for investigation. As higher-order thinking skills in addition to academic knowledge are central to science education where EMI teaching might adversely influence some students' learning (Fung & Yip, 2014; Lau & Yuen, 2011; Yip & Tsang, 2007; Yip, Tsang, & Cheung, 2003), it compared the S1 school-based test papers in two EMI science subjects, namely integrated science (IS) and computer and information technology (CIT), in the selected school right before (i.e., the 2009−2010 academic year) and immediately after the fine-tuning (the 2010−2011 academic year). Given that all the content-area subjects in this school were taught in Chinese previously, it is expected that the designs of these papers reflect some of the teachers' initial concerns regarding how to adapt to this abrupt shift of language use, although the direct correspondence between the changes in the policy and test papers might not be easily identified merely by observing the changes in their content and designs. For this reason, the analysis and discussion of the test papers are supplemented by findings derived from interviews with the corresponding IS and CIT teachers.

The School-based MOI Policy and Assessments

The majority of students in the selected government-aided secondary school were categorized into band 2 (out of the 3 bands), which indicates that they have a medium level of academic ability in Hong Kong. Like many other similar schools, IS and CIT were chosen as two EMI content-area subjects (Kan et al., 2011), in which English was supposed to be the medium for written communication in teaching materials and assessments and verbal communication as a classroom language. In addition, as the school had been preparing for the new MOI initiative, some proportions of English were used in the teaching of the S1 IS and CIT syllabi under the school-based policy in the 2009–2010 academic year. In order to cater for learning diversity, these proportions of English elements differed for classes involving students with higher and lower academic capability, so were in the corresponding parts of the teaching materials and test items in the assessments. There were subsequently two versions of IS test papers for all the assessments, having more and less English questions (yet with the same meaning with the Chinese ones) for the two student groups respectively. Nevertheless, this design was not applied to the CIT assessments as it was a subject with much less lesson time and, therefore, less content knowledge to be assessed. In each academic year, there were two IS uniform tests and examinations respectively but there were only two CIT examinations, all of which were analyzed for the two academic years.[4] Since the implementation of the fine-tuning initiative in the 2010–2011 academic year, IS and CIT have become two EMI non-language subjects in the target school and all teaching materials and assessments have been written in English. Though the number of uniform tests and examinations has remained unchanged, the two-version policy for the IS assessments has been cancelled. Instead, students' academic diversity was addressed by incorporating a few long questions with two difficulty levels, one being more academically difficult but consisting of a higher full score and vice versa (see further discussion below).

Data Analysis

The analytical procedure of the test or examination papers was via two phases. First, the general formats and designs of all these papers were identified in each of the two academic years for comparison. These involved the changes in the types of questions (e.g., multiple choice, fill-in-the-blank, matching) and their weighting. The question types were mainly defined on the basis of the name of subsections appearing on the papers (see below).

The second phase highlighted the changes in the underlying assessed categories of cognitive processes and knowledge in all the test papers according to the revised Bloom's taxonomy (see Anderson & Krathwohl, 2001, for the criterion of classification). More specifically, each test item, together with their assigned scores, was evaluated against the taxonomy table, yielding percentages of scores for each of the hierarchical levels regarding the two dimensions (see Appendix 1). This procedure was first individually conducted by two researchers, who then compared for instances of disagreement followed by cross-checking and further negotiation to reach a mutual agreement. Table 3 shows the total number of test items in the IS and CIT papers respectively before and after the fine-tuning. In addition, descriptive interpretation based on the changes in question types, use of language and pedagogical foci was also included. Further explanation for the changes is supplemented by findings derived from semi-structured interviews with 3 IS and 2 CIT teachers (including the heads of departments). All the teachers were asked in Cantonese (their first language) to express their views on the fine-tuning MOI policy, school-based MOI arrangements, difficulties facing themselves and their students as well as any corresponding solutions. Prior approval for the data collection procedure was obtained from the corresponding teachers and the principal.

Table 3:   Number of Test Items in IS and CIT Papers Before and After the Fine-tuning

	Number of test items in the 2009−2010 academic year	Number of test items in the 2010−2011 academic year
IS papers	240	200
CIT papers	135	103

Findings and Discussion

## *The Changes in the Format and Design*

The first phase of the analysis aimed to uncover how content-area teachers designed test papers differently before and after the fine-tuning in order to suit both their purposes of assessment and learning of students. Focusing on the two science subjects (i.e., IS and CIT) that were transiting from CMI to EMI, this section discusses the changing format (e.g., question types) and design (score distribution, pedagogical foci) in the internal summative assessments.

Integrated Science

According to the subject teachers, the school-based junior secondary IS syllabus has undergone several stages of development some years before the implementation of the fine-tuning policy so as to ensure a smooth transition. As it was expected that students would encounter additional linguistic challenges in EMI teaching, the syllabus has been modified, reorganized and reduced. Some of the content that would appear in the senior secondary science subjects has been omitted (teachers' interview responses). In the S1 IS syllabus, though six main topic areas, namely "introducing science," "living things," "cells and human reproduction," "energy," "water — the wonderful solvent" and "matter as particles," were included both before and after the policy change, there was clearly less assessed (and possibly taught) knowledge involved in each area in the latter as can be observed in the design of the test papers (see discussion below).

The format of test papers tends to be consistent in the same academic year but some changes regarding the incorporated types of questions have been made after the fine-tuning. Figure 1 illustrates the proportions of scores for each question type accumulatively in the test papers in the year before and after the change of MOI policy. The main difference lay in the reduction of marks in questions requiring answers in full sentences or phrases such as long (difference in score percentage: −7.1%) and short questions[5] (−10.6%) and, in turn, an increase in score proportions for sections such as multiple choice (+5.7%), fill-in-the-blank (+9.4%) and, particularly, true or false (+5.7%) which was not included in the papers before

Figure 1:   Score Percentages of Question Types in IS Test Papers Before and After the Fine-tuning
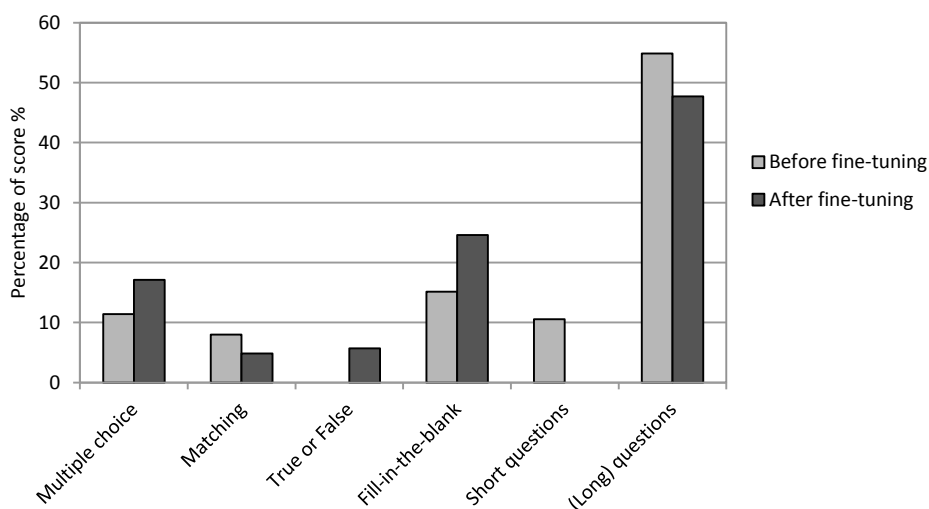
the 2010–2011 academic year. There was, nevertheless, a decrease in the overall score percentage for the section of matching (−3.1%) in various IS papers after the fine-tuning.

According to the IS teachers, the school-based syllabus had to be reorganized, simplified and reduced after the fine-tuning because they found it difficult to "teach everything" using English owing to time limitations. In terms of assessments, they were also concerned about students not being able to "interpret the questions" and, therefore, suggested that "knowing question words" and "memorization" were, in fact, important examination techniques. As can be seen in the IS test papers after the fine-tuning, despite the higher score percentage of multiple-choice questions, some of them in fact carried more marks per question (2 marks) than those in the former version (1 mark), resulting in the reduced amount of assessed content and more time for answering. Similarly, this design was also applied to some of the fill-in-the-blank questions particularly when there were no options of vocabulary provided. Other measures that apparently helped students cater for this language-versus-academic-knowledge dilemma in the new test papers comprised providing more pictures in the questions to aid understanding, giving Chinese translations to difficult/new vocabulary (e.g., "exhaust gas 廢氣," "Shenzhou 7 spacecraft 神舟七號"), and requiring diagram drawing or picture matching (instead of writing in English) as answers. In some questions, it is likely that students could respond simply based on the pictures provided.

What seems to be more interesting is the unique design of some final (long) questions in the new test papers that aim to assess both students' academic knowledge and language skills (i.e., both understanding and writing in English). In these questions, the students were given additional language support for answering in complete sentences, where compulsory scores were assigned. They were also guided to gradually write in complete sentences via the provision of some sample paragraphs in the form of fill-in-the-blank structure (see Exemplar 1 in Appendix 2). In addition, students of diverse academic ability were given the option to answer either level 1 or level 2 of the same question, in which the former carried fewer scores. In general, the rationale for this change in design, as the teachers acknowledged, was to cater for students' learning diversity and their language barrier brought about by the fine-tuning MOI policy. Nevertheless, the language focus of the new test papers, at least as revealed in these "long questions" and the increased score percentages for "fill-in-the-blank" questions, tended to be at the word rather than sentence level because students were less required to write complete sentences. It was also reported by the teachers

that dictations had become common practices to help students remember key words, phrases or sometimes sentences in the content subjects.

Computer and Information Technology

Notwithstanding its shorter length and less assigned lesson time, the S1 school-based CIT syllabus consisted of a much wider range of academic content, and more chapters, which was mainly divided into two components, namely theories (e.g., the computer system, hardware, software) and practices (e.g., computer applications, programming language "Logo"). The changes in score proportions of questions types in the CIT assessments before and after the fine-tuning are shown in Figure 2, indicating an increase in the percentages of multiple choice (+5.0%), true or false (+5.0%) and short questions (+15.0%) but a decrease in those of matching (–7.5%), programming (–10.0%) and (long) questions (–7.5%). Nonetheless, as can be observed in the following more detailed analysis, the significantly reduced proportion of programming-type questions might become a key concern to the CIT subject, where application should play an equally, or even more, vital role.

Figure 2:   Score Percentages of Question Types in CIT Test Papers Before and After the Fine-tuning
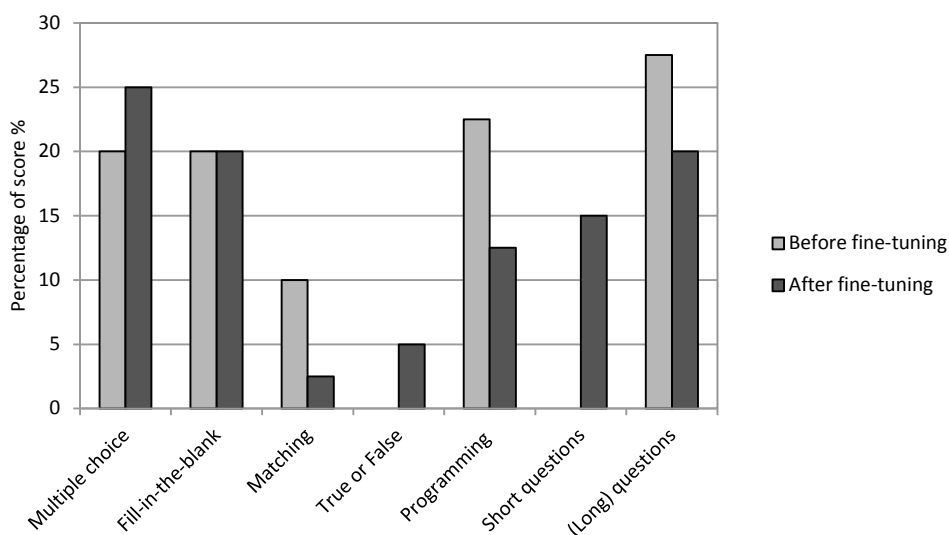
Resembling the changes in the IS papers, more scores (2 instead of 1 per question) were also assigned to individual questions in the multiple choice and true or false sections in the revised CIT papers. This, again, reduced the quantity of the assessed subject content

while potentially allowing students with more thinking time to complete the papers. Other than including more pictures in some of the questions, there were two parts of the fill-in-the-blank questions, the first providing vocabulary options whereas the second incorporating the first letter of the target words as clues. As suggested by one of the CIT teachers, this design was deliberately adopted in the examination papers in the 2010–2011 academic year to address students' language difficulties. As for the "long questions" section, 5% was taken to account for the language accuracy for writing in complete sentences. On top of these, though 100% was traditionally the maximum score for each CIT examination paper, there were 20% bonus scores added to the new version in an extra section called "extended questions," 5% of which, again, contributed to writing accurate complete sentences. Without affecting the total mark of the whole paper, it could be expected that these bonus scores could slightly increase the students' final result and, possibly, provide them with a more positive impression about their performance and, hence, enhance their motivation.[6]

The Changes in Assessed Academic Competence

By employing the revised Bloom's taxonomy as the analytical framework, the second phase of the investigation determined the extent to which the underneath educational objectives have been shifted in test items during the MOI policy transition in two dimensions: (a) cognitive processes (consisting of 6 hierarchical levels) and (b) knowledge (consisting of 4 levels) (see the taxonomy table in Appendix 1). This section discusses each of these aspects regarding the changes in the IS and CIT papers.

The assessed cognitive processes

Generally speaking, the first class of the cognitive processes, "remember," tends to correspond to the degree to which a student has learnt some subject content and retained it over some period of time, and it is often associated with rote learning (Anderson & Krathwohl, 2001). In terms of teaching and assessing, this is, however, a less desirable goal than meaningful or constructive learning that is most related to the thinking processes in other categories (e.g., "understand," "apply," "analyze," "evaluate") because they promote transfer (rather than retention), that is, the ability to solve new problems, answer new questions or facilitate learning new subject matters based on the acquired knowledge (see also Mayer, 2002). Nevertheless, at the junior secondary level, it could be anticipated that assessment questions with lower-level cognitive processes (e.g., the categories of

"remember," "understand," "apply") are more prominent than those with a higher level (e.g., the categories of "analyze," "evaluate," "create") because the establishment of junior students' foundation in academic knowledge is a prerequisite before they acquire other higher-order thinking skills.

1.   *Integrated Science*

The evaluation of the changes in test items in the IS test papers revealed a clear tendency of a decreasing number of questions involving higher-level cognitive processes. As depicted in Figure 3, although questions associated with the category of "remember" accounted for the majority of score percentages in the IS papers both before and after the fine-tuning, the proportions of scores for these questions have clearly increased by 14.0% since the policy change. In contrast, those of other test items involving the categories of "understand" (−6.9%), "apply" (−6.6%) and "evaluate" (−0.6%) have all been decreased. There were, nevertheless, no instances of questions assessing the learners' abilities in the categories of "analyze" and "create." This pattern was partly due to the increased score percentages for questions such as true or false, fill-in-the-blank and multiple choice as they tended to rely on memorization, which, according to Anderson and Krathwohl (2001), mainly require learners to "recognize" or "recall" the relevant knowledge from their long-term memory. Even though some better designed multiple-choice questions could also

Figure 3:   Score Percentages of Cognitive Process Categories in IS Test Papers Before and After the Fine-tuning
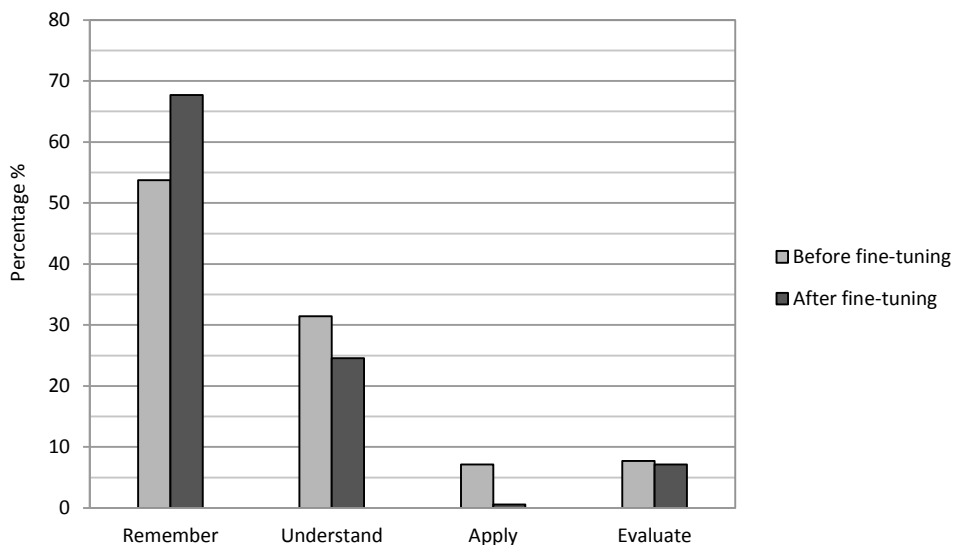
be somewhat challenging to students (e.g., if they require students to apply their knowledge to new scenarios), Black and Wiliam (1998) suggest that, unlike open questions that allows students to provide explanatory answers, close questions tend to induce lower-order thinking skills. Similarly, Stanger-Hall (2012) particularly argues that multiple-choice questions indeed hinder learners' critical thinking skills in introductory science classes as their restricted format does not encourage learners to modify their study strategies.

Furthermore, the reduced proportions for questions requiring high-order thinking processes were also reflected in the differing designs of the short/long questions. For instance, even though "understand" was found to be the second most frequently occurring cognitive category in most of the IS test papers, there was a wider range of assessed processes in the test items before than after the fine-tuning. Based on Anderson and Krathwohl's (2001) definition, to "understand" requires learners to construct meaning from instructional messages and, in the Chinese questions in the previous IS papers, some typical skills in this regard comprised "explaining," "classifying," "exemplifying," "inferring" and "comparing." Subsequently, when EMI teaching was adopted, the test items evaluating students' cognitive category of "understand" tended to be limited to a few thinking processes such as "explaining" (Exemplar 2 in Appendix 2) and "inferring" (Exemplar 3). One clue for this change is the tendency that some of these questions were in the same or similar format and focused on particular academic knowledge, the most frequently assessed of which was the topic "energy" which required students to describe the energy conversion taken place in some familiar or unfamiliar situations (Exemplar 3). This concentration on specific content knowledge and design of questions could be perhaps attributed to the pedagogical foci of the subject teachers in the classroom, who, in the interviews, also acknowledged that they regarded these as most essential education outcomes. Whereas this arrangement has inevitably led to the reduced amount of learning content, it was perceived by the teachers as a feasible way to strike a balance between students' language challenges and learning effectiveness (see also J. Y. H. Chan, 2014).

Notably the most distinctly reduced cognitive type in the test papers after the fine-tuning was "apply," that is, carrying out or using a procedure in a given situation (Anderson & Krathwohl, 2001). In the IS subject, as this aspect is mainly taught or evaluated via scientific experiments, the assessment of this category in the written paper before the policy change was mainly in science-related calculations (Exemplar 4). Despite the lower language requirement for this kind of questions, they were rarely found in the revised papers in English probably because these science-related calculations were deemed to be less crucial

than some other aspects of knowledge at this grade level and were hence downplayed in the assessment. Although the high-order cognitive processes involved in "evaluate" is apparently difficult to be taught and assessed especially at junior secondary level, we could perhaps identify attempts in both the Chinese and English test papers to incorporate some of corresponding thinking processes such as "recommending" and "judging" (Anderson & Krathwohl, 2001), though some of these recommendations might have already been previously mentioned (or even taught deliberately) in the classroom and possibly led to students' memorization (see Exemplar 5). Given that test items involving "judgment" were found to be quite common in experiment-type questions in both the two groups of IS papers, it could be deduced that this skill was also perceived to be vital in the science subject.

2.   *Computer and Information Technology*

With respect to the CIT subject in S1, the score percentages involving the cognitive process type of "remember" also greatly outnumbered other categories followed by "understand," "apply" and "evaluate" in the test papers in the two academic years (Figure 4). Similar to the design of IS test papers, cognitive processes involved in the category of "remember" was mainly assessed in the forms of multiple choice, fill-in-the-blank, true or false and matching questions, in addition to some rather straightforward long/short questions that solely elicited students' memorized knowledge (e.g., the definition of "data" and "input"). More importantly, a more detailed comparison again revealed a similar pattern to that of the IS subject in that more lower-order thinking questions (i.e., "remember": +6.0%), together with the additional 5% language score contributing to language accuracy in the "long questions" section, replaced the higher-order counterparts (i.e., "understand": −0.5%; "apply": −2.5%; "evaluate": −8.0%) after the fine-tuning. In the test papers in both academic years, these test items assessing higher-level cognitive processes were mainly included in the sections of programming and short/long questions. Furthermore, the categories of "analyze" and "create" were also not applied to the test papers for this S1 CIT subject.

Indeed, these decreasing score proportions of assessments on higher-level processes could be accounted for mainly by the simpler question designs in the English papers after the fine-tuning. For example, in the aspect of "understanding," while there were more sophisticatedly devised questions in the previous Chinese papers that required students to process some given data and practice various thinking skills such as "classifying," "explaining," "exemplifying," "inferring" and "comparing," those in the English papers tended to be rather short in length and limited in the range of question variety. Even if the

Figure 4: **Score Percentages of Cognitive Process Categories in CIT Test Papers Before and After the Fine-tuning**
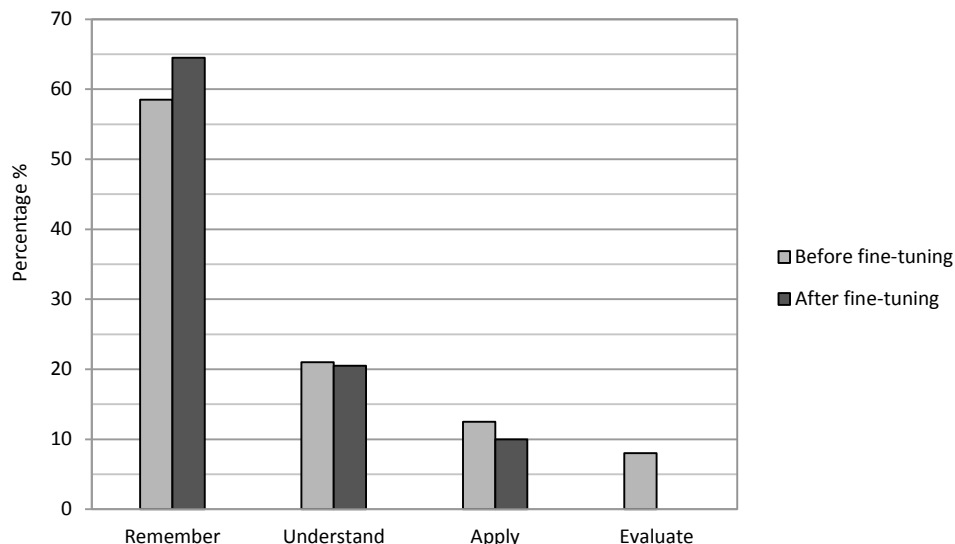
* As 5% of the scores in the CIT test papers after the fine-tuning was taken to account for students' language accuracy, the total percentage for the assessed academic content became 95%.

question aimed to evaluate students' "inferring" skills (Exemplar 6), the data provided were relatively easier to comprehend in comparison to those in the former Chinese papers (cf. Exemplar 7), which perhaps have a higher language requirement if English had been used.

The application aspect of the CIT subject lies not only in the daily computer-related issues but, more technologically, also in programming which is part of the CIT curriculum. In this respect, the section of programming in the written paper is probably unique to the CIT subject owing to the difficulties of assessing students' ability to apply their knowledge. As revealed in the study findings, the language issue in this kind of assessment items was apparently less prominent as fewer words were required in the questions and, in essence, they tended to ask the students to undergo the cognitive process of "implementing," that is, applying a procedure to an unfamiliar task (Anderson & Krathwohl, 2001). Therefore, the difference in the score proportion of questions in this category of "apply" was relatively small, the major divergence of which was found to be the more covered syllabus content in the Chinese test papers than the English counterparts. Specifically, before the policy change, two areas of academic knowledge were included, namely the simplified Cangjie (Chinese) Input method and "Logo" commands but when EMI was adopted, the former was omitted. Like the findings in the IS papers, the category of "evaluate" also only appeared in the

assessments when Chinese was the MOI possibly because a higher language ability was needed to understand and respond to the questions, which required students to provide recommendations based on some given information.

The above observed changes in the test paper design generally accord with the interviewed CIT teachers' concern about their limited classroom time given to teach both the subject-related language and content knowledge and, sometimes, they might have to "focus more" on the former. One of the teachers also acknowledged that there had been "more memorization" (e.g., "fill-in-the-blank questions" with options provided) in the revised examination papers and, because of time constraint, they might have to alter their teaching strategies to "focusing on key terms in specific topics," "helping students to revise specifically for examinations" and "incorporating examination content in the form of worksheets as main teaching materials" (cf. Brooks, 2002, about the negative potential of summative assessment).

The changing assessed knowledge

We now shift the attention to changes of the choice of test items with reference to the dimension of knowledge in the revised taxonomy. Among the four knowledge types, only three of them at the first three levels (i.e., "factual," "conceptual," "procedural") were identified in the assessment items in all the IS and CIT test papers whereas questions involving metacognitve knowledge (at the highest level) was absent as it refers to one's awareness of and knowledge about their cognition, which are likely to be too demanding and difficult to be implemented in assessments at junior secondary level (Anderson & Krathwohl, 2001).

1. Integrated Science

In general, while questions assessing "factual," "conceptual" and "procedural" knowledge were quite evenly distributed in the Chinese IS test papers before the fine-tuning, the score proportions have shifted toward the first two types (i.e., "factual": +4.6%; "conceptual": +15.1%) in the papers after the policy change, resulting in decreasing score percentages in questions focusing on "procedural" knowledge (–19.7%) (Figure 5). Quite dissimilar to the dimension of cognitive assessment in which the pedagogical foci are mainly reflected in the question types (e.g., multiple choice, fill-in-the-blank) and the implicit "verb" embedded (e.g., "recall," "explain"), the choice of knowledge type depends on subject matter knowledge and is less affected by the assessment format. From this viewpoint, the

findings suggest that the three types of knowledge appeared in all the differing sections in all the test papers but they tended to be subject to the assessed areas of subject matter content. Therefore, a more detailed investigation into the topic areas involved in the test papers is needed in order to provide an explanation for this trend in the changing assessed knowledge types.

Figure 5: Score Percentages of Knowledge Types in IS Test Papers Before and After the Fine-tuning
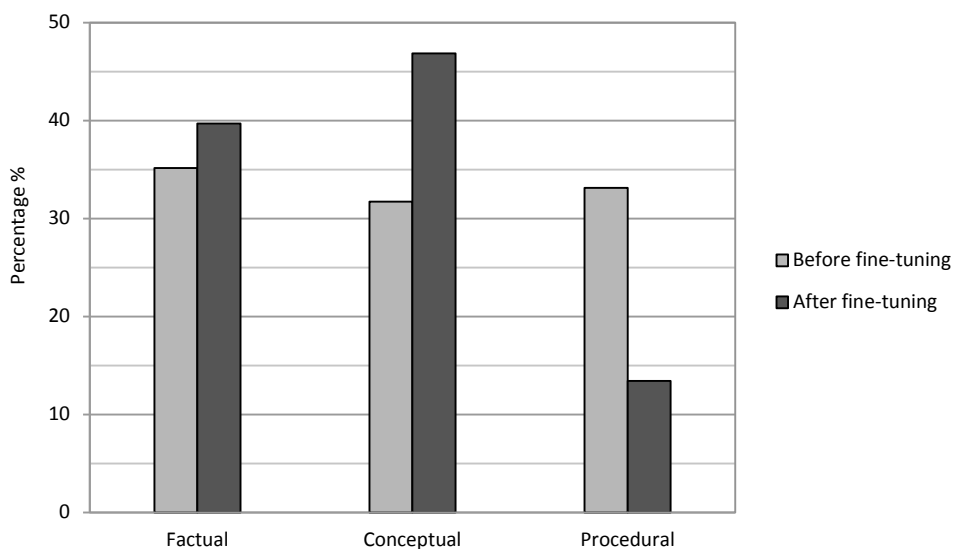
Anderson and Krathwohl (2001) define "factual" knowledge as the "basic elements students must know to be acquainted with a discipline or solve problems in it" and "conceptual" knowledge corresponds to "the interrelationships among the basic elements within a larger structure that enable them to function together" (p. 47) (see also Table 2). Both these two categories of knowledge are likely to be integrated into the assessment regardless of the question formats as well as subject matter areas, especially if they are combined with the various kinds of cognitive process to form diversifying test items. However, in the IS subject, the assessment of "procedural" knowledge is mostly associated with scientific investigation, which is only central to specific content areas such as "introducing science" (e.g., basic scientific methods, equipment, techniques and precautions), "cells and human reproduction" (e.g., use of microscope) and "water — the wonderful solvent" (e.g., specific techniques: filtration, distillation), among the six included in the S1 syllabus. From this perspective, the changes in the score distribution of assessed

topic areas in the test papers before and after the fine-tuning have suggested a reduction in the score percentages in the first two aforementioned topics ("introducing science": –13.4%; "cells and human reproduction": –17.4%) but an increase in those for the third one ("water — the wonderful solvent": +28.3%) (Figure 6). On top of these, it should be stressed that the first topic, among all three, contained the most elements relevant to scientific investigation and, thus, had the greatest potential to produce test items focusing on procedural knowledge. As mentioned above, the noticeable increase in score proportions under the topic of "energy" (+11.7%) also to some extent accounted for the increased percentages of questions carrying "conceptual" knowledge (e.g., the classification of energy types) in the IS paper after the fine-tuning. Furthermore, there was also a decrease in the assessed content in topics such as "living things" (–9.7%) and "cells and human reproduction" (–17.4%), possibly because of their greater linguistic challenges as information in these topics, as can be observed in the Chinese vis-à-vis English test papers, tended to require more descriptive writing (e.g., describing symptoms of pregnancy, writing the functions of specific human organs) and involve more technical vocabulary (e.g., different parts of genital organs and cells). As reported by the IS teachers, having considered students' language challenges and the limited teaching time, they "reorganized" and "simplified" the junior secondary IS syllabus after the fine-tuning according to the language difficulty level of the topics and their connection with the science syllabi at the senior level

Figure 6:   Percentages of Subject Matter Areas in IS Test Papers Before and After the Fine-tuning
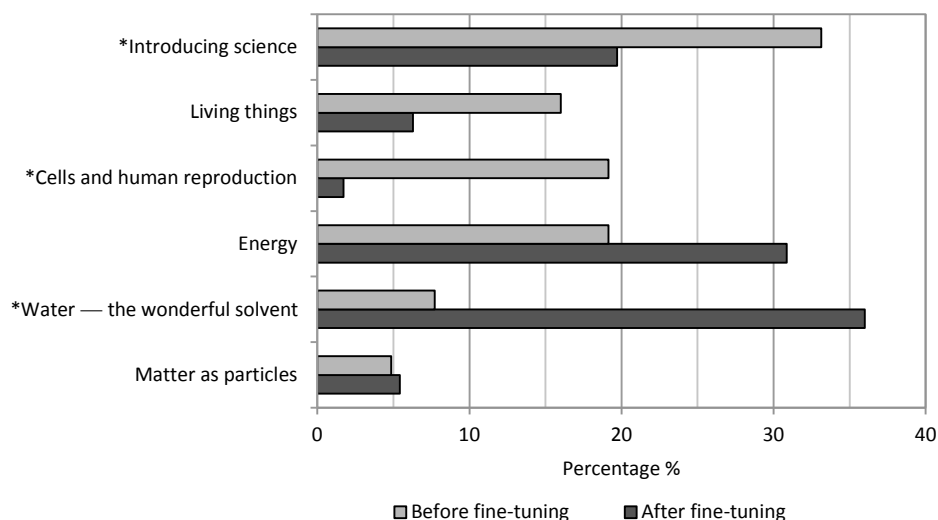

* Topics that have a particular focus on "procedural" knowledge.

(e.g., Physics, Chemistry, Biology). As a result, some of the "overlapping" or "reappearing" content were omitted or reduced in the junior level IS syllabus and, hence, assessments.

2. Computer and Information Technology

In the CIT test papers, scores associated with "factual" and "procedural" knowledge generally outnumbered those with "conceptual" knowledge, but there was again a shift toward a higher score proportion assessing "factual" knowledge (+11.5%) and a lower proportion of score involving "procedural" knowledge (–10.0%) during the transition of the MOI policy (Figure 7). In addition, there was also a slight decrease in the score percentages for the evaluation of "conceptual" knowledge (–6.5%) in the revised paper in English. As discussed in the previous section, one observable change accounting for this score distribution was the inclusion of more definition-eliciting questions not only in multiple-choice, matching and fill-in-the-blank questions, but also in short/long questions probably because the teachers regarded terminologies (e.g., "software," "input," "data") as essential in junior secondary CIT education. Alternatively, "procedural" knowledge was mainly integrated in multiple-choice and programming questions in addition to a small proportion of fill-in-the-blank and short/long questions. Among the wide range of content topics in the CIT syllabus, this kind of knowledge was rather restricted in the application of "Logo" and

Figure 7: Score Percentages of Knowledge Types in CIT Test Papers Before and After the Fine-tuning
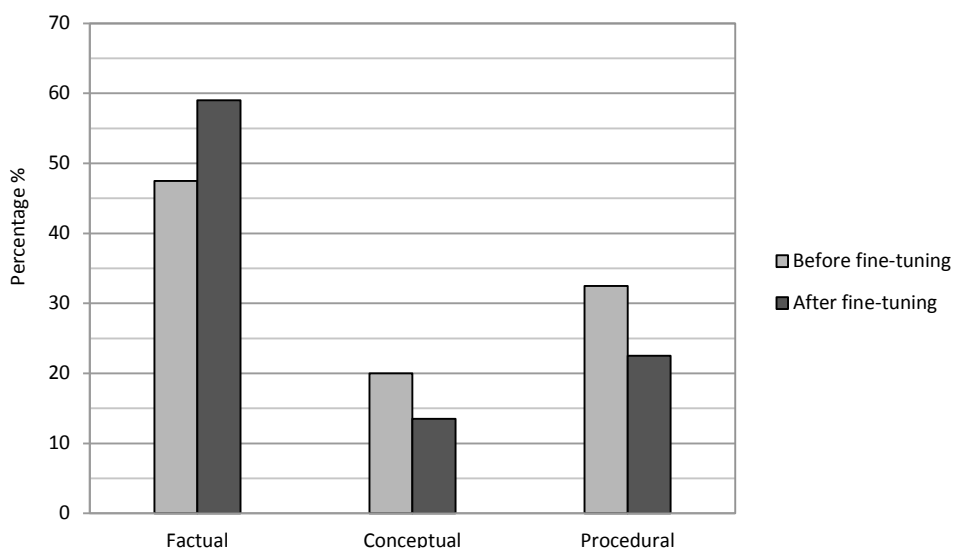
* As 5% of the scores in the CIT test papers after the fine-tuning was taken to account for students' language accuracy, the total percentage for the assessed academic content became 95%.

"simplified Chinese input method" (and a few other program applications such as "Windows" and "Microsoft Word"). One major difference between the test items before and after the policy change was the omission of the latter in the revised English syllabus, possibly leading to the decreasing amount of "procedural" knowledge.

Summary and Implications

This study has uncovered the changing designs of the S1 IS and CIT test papers in a typical former CMI secondary school before and after the implementation of the new fine-tuning MOI initiative. Apart from the descriptive approach to investigating the observable changes in the assessment designs and formats, a crucial perspective in the present analysis is the evaluation of the test items in the dimensions of the cognitive and knowledge types on the basis of the revised Bloom's taxonomy, yielding a systematic account of the shift in pedagogical foci of the test papers alongside the policy transition. The findings reveal that there was a tendency that more scores were reallocated to the types of questions requiring less understanding and writing in English in both the IS and CIT assessments (e.g., multiple choice, fill-in-the-blank, matching). Possibly with the intention of striking a balance between linguistic challenges and the greater time allocated to teaching, there were more concentrated foci on specific topic areas that were potentially the pedagogical foci in the classroom, which subsequently led to the reduction and refocusing of the assessed and/or teaching content in the syllabus.

Based on a more detailed analysis using the taxonomy framework, the major discovery is the decrease in the proportion of questions assessing both higher-level cognitive (e.g., "understand," "apply," "evaluate") and knowledge categories (e.g., "conceptual," "procedural") in the IS and CIT papers after the adoption of EMI teaching in the school, which potentially favored less desirable education outcomes of retention than transfer (Mayer, 2002). These higher-order perspectives, however, are crucial to science education, and perhaps all academic subjects, especially when students advance to the senior secondary and tertiary education. Nevertheless, this decreased proportion of questions assessing higher-order cognitive processes and knowledge tends to correspond to topics in specific subjects with greater linguistic demands (e.g., academic vocabulary, writing and reading subject-related descriptive language). In contrast, there seem to be fewer changes in topic areas such as programming in CIT because of their lower language requirements.

Although the washback effect tends to be more prominent toward the end of secondary

education where the public examination takes place, this test paper design, as probably influenced by the school's long-term curriculum planning (see J. Y. H. Chan, 2013), might also to some extent affect the daily teaching and learning practices. In this respect, the findings echo previous research into EMI vis-à-vis CMI education in that greater emphasis has been put on memorization and copy (e.g., J. Y. H. Chan, 2014; Lo & Macaro, 2012; Ng, 2007; Yip, Coyle, & Tsang, 2007), which, in turn, might result in students' lower self-concept in academic subjects (Fung & Yip, 2014; Lau & Yuen, 2011; Yip & Tsang, 2007; Yip, Tsang, & Cheung, 2003). The degree of influence is likely to be associated with an array of interrelated factors such as the school-based MOI arrangements, administrators and teachers' beliefs, teaching methodologies and their experience in EMI content-based teaching as well as students' language proficiency. The findings in this study have, therefore, provided a foundation for further investigation not only to monitor the impact of the fine-tuning MOI policy on a particular academic level or subject, but also the transformation of curricula and practices in the entire secondary education and, possibly, its extended influence on tertiary education. Throughout some years of policy implementation, it is possible that subject teachers gradually alter their perceptions of EMI teaching, teaching strategies and curriculum and test paper design based on their accumulated teaching experience. Nonetheless, even though the present study has only centered on the junior secondary school-based assessment in one Hong Kong secondary school, the changes in test papers reported in this case study may be common to those in similar school contexts especially if the teachers are also facing similar challenges when implementing EMI teaching. Furthermore, the methodological approach adopted in this study and insights from the findings might be equally applicable and valuable to wider educational contexts and perhaps with a broader scope, given the high popularity of bilingual education in the era of globalization (Baker, 2011).

Notes

1. In Taras's (2009) review of the two forms of assessment, summative assessment refers to a single process that makes a judgment on learners' performance according to criteria and standards while formative assessment helps learners to address these criteria or standards by providing appropriate feedback. Though the primary goal of summative assessment is often to measure/record one's learning outcomes, recent research has highlighted the importance of incorporating it as a positive part of the learning process (see Black et al., 2003; Taras, 2009).

2. Under the laissez-faire policy before 1997, over 90% of the secondary schools in Hong Kong claimed to have adopted EMI in content-area subjects, although in many cases code-mixing/switching was a common practice in the classroom accompanied by the use of English teaching materials (Education Department, 1997; Evans, 2013).

3. According to the Education Bureau (2010), three main types of MOI arrangements can be adopted: (a) mother-tongue teaching complemented with different modes of Extended Learning Activities (ELA) in English, (b) Chinese or English as MOI in specific subjects, and (c) Chinese or English as their MOI for all non-English subjects.

4. It is noted that there were some "extended questions" in all the IS test papers both before and after the MOI policy change, which allow students to attain bonus scores that were extra to the 100% full score. As these rather flexibly designed questions were quite different in their formats and scope of testing depending on the preference of individual teachers, they were not analyzed in the study.

5. The distinction between the two sections for short and long questions respectively disappeared after the fine-tuning and they were combined into a single section "questions," which contributed to a smaller score percentage.

6. As also discussed in note 4, this kind of extended questions also appeared in the IS papers and it is possible that they served a similar function. Another reason for such a design according to the IS teachers was to provide students with more opportunities to practice their high-order thinking skills.

References

Ahmad, S., & Rao, C. (2012). A review of the pedagogical implications of examination washback. *Research on Humanities and Social Sciences*, *2*(7), 11−20.

Ammon, U. (2010). World languages: Trends and futures. In N. Coupland (Ed.), *The handbook of language and globalization* (pp. 101−122). Malden, MA: Wiley-Blackwell.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman.

Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Bristol, England: Multilingual Matters.

Berry, R. (2008). *Assessment for learning*. Hong Kong, China: Hong Kong University Press.

Berry, R. (2011). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy and Practice*, *18*(2), 199−211. doi: 10.1080/0969594X.2010.527701

Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., … Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review*, *1*(1), 61−67. doi: 10.1016/j.edurev.2006.01.001

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, England: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, *5*(1), 7−74. doi: 10.1080/0969595980050102

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, NY: David McKay.

Brooks, V. (2002). *Assessment in secondary schools: The new teacher's guide to monitoring, assessment, recording, reporting and accountability*. Buckingham, England: Open University Press.

Chan, C. C., Tsui, M. S., Chan, M. Y. C., & Hong, J. H. (2002). Applying the Structure of the Observed Learning Outcomes (SOLO) taxonomy on student's learning outcomes: An empirical study. *Assessment and Evaluation in Higher Education*, *27*(6), 511−527. doi: 10.1080/0260293022000020282

Chan, J. Y. H. (2013). A week in the life of a "finely tuned" secondary school in Hong Kong. *Journal of Multilingual and Multicultural Development*, *34*(5), 411−430. doi: 10.1080/ 01434632.2013.770518

Chan, J. Y. H. (2014). Fine-tuning language policy in Hong Kong education: Stakeholders' perceptions, practices and challenges. *Language and Education*, *28*(5), 459−476. doi: 10.1080/09500782.2014.904872

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, *11*(1), 38−54. doi: 10.1080/09500789708666717

Choi, P. K. (2003). "The best students will learn English": Ultra-utilitarianism and linguistic imperialism in education in post-1997 Hong Kong. *Journal of Education Policy*, *18*(6), 673−694. doi: 10.1080/0268093032000145917

Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sciences Education*, *7*(4), 368−381. doi: 10.1187/cbe.08-05-0024

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge, England: Cambridge University Press.

Curriculum Development Council. (1998). *Syllabuses for secondary schools: Science (Secondary 1–3)*. Hong Kong, China: Government Printer.

Dixon, L. Q. (2009). Assumptions behind Singapore's language-in-education policy: Implications for language planning and second language acquisition. *Language Policy*, *8*(2), 117−137. doi: 10.1007/s10993-009-9124-0

Doiz, A., Lasagabaster, D., & Sierra, J. M. (Eds.). (2013). *English-medium instruction at universities: Global challenges*. Bristol, England: Multilingual Matters.

Education Bureau. (2010). *Enriching our language environment, realising our vision: Fine-tuning of medium of instruction for secondary schools*. Hong Kong, China: Author.

Education Commission. (1990). *Education Commission report no. 4: The curriculum and behavioural problems in schools*. Hong Kong: Government Printer.

Education Commission. (2005). *Report on review of medium of instruction for secondary schools and secondary school places allocation*. Hong Kong, China: Government Logistics Department.

Education Department. (1997). *Medium of instruction guidance for secondary schools*. Hong Kong, China: Author.

Evans, S. (2013). The long march to biliteracy and trilingualism: Language policy in Hong Kong education since the handover. *Annual Review of Applied Linguistics*, *33*, 302−324. doi: 10.1017/S0267190513000019

Ferguson, G. (2006). *Language planning and education*. Edinburgh, Scotland: Edinburgh University Press.

Fung, D., & Yip, V. (2014). The effects of the medium of instruction in certificate-level physics on achievement and motivation to learn. *Journal of Research in Science Teaching*, *51*(10), 1219−1245. doi: 10.1002/tea.21174

Gardner, J. (Ed.). (2006). *Assessment and learning*. London, England: Sage.

Gee, J. P. (2008). What is academic language? In A. S. Rosebery & B. Warren (Eds.), *Teaching science to English language learners: Building on students' strengths* (pp. 57−69). Arlington, VA: National Science Teachers Association.

Hamid, M. O., Nguyen, H. T. M., & Baldauf, R. B. (2013). Medium of instruction in Asia: Context, processes and outcomes. *Current Issues in Language Planning*, *14*(1), 1−15. doi: 10.1080/14664208.2013.792130

Hashim, A. (2009). Not plain sailing: Malaysia's language choice in policy and education. *AILA Review*, *22*(1), 36−51. doi: 10.1075/aila.22.04has

Johnson, R. K., & Swain, M. (Eds.). (1997). *Immersion education: International perspectives*. Cambridge, England: Cambridge University Press.

Kan, V., Lai, K. C., Kirkpatrick, A., & Law, A. (2011). *Fine-tuning Hong Kong's medium of instruction policy*. Hong Kong, China: The Hong Kong Institute of Education.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212−218. doi: 10.1207/s15430421tip4104_2

Lau, W. W. F., & Yuen, A. H. K. (2011). The impact of the medium of instruction: The case of teaching and learning of computer programming. *Education and Information Technologies*, *16*(2), 183−201. doi: 10.1007/s10639-009-9118-8

Lee, O. (2005). Science education with English language learners: Synthesis and research agenda. *Review of Educational Research*, *75*(4), 491−530. doi: 10.3102/00346543075004491

Lin, A. (2013). Toward paradigmatic change in TESOL methodologies: Building plurilingual pedagogies from the ground up. *TESOL Quarterly*, *47*(3), 521−545. doi: 10.1002/tesq.113

Lindholm-Leary, K. J. (2001). *Dual language education*. Clevedon, England: Multilingual Matters.

Lo, Y. Y., & Lo, E. S. C. (2014). A meta-analysis of the effectiveness of English-medium education in Hong Kong. *Review of Educational Research*, *84*(1), 47–73. doi: 10.3102/0034654313499615

Lo, Y. Y., & Macaro, E. (2012). The medium of instruction and classroom interaction: Evidence from Hong Kong secondary schools. *International Journal of Bilingual Education and Bilingualism*, *15*(1), 29–52. doi: 10.1080/13670050.2011.588307

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice*, *41*(4), 226–232. doi: 10.1207/s15430421tip4104_4

Ng, D. F. P. (2007). Medium and learning in Chinese and English in Hong Kong classrooms. *Language Policy*, *6*(1), 163–183. doi: 10.1007/s10993-006-9037-0

Okebukola, P. A., Owolabi, O., & Okebukola, F. O. (2013). Mother tongue as default language of instruction in lower primary science classes: Tension between policy prescription and practice in Nigeria. *Journal of Research in Science Teaching*, *50*(1), 62–81. doi: 10.1002/tea.21070

Poon, A. Y. K. (2013). Will the new fine-tuning medium-of-instruction policy alleviate the threats of dominance of English-medium instruction in Hong Kong? *Current Issues in Language Planning*, *14*(1), 34–51. doi: 10.1080/14664208.2013.791223

Poon, A. Y. K., Lau, C. M. Y., & Chu, D. H. W. (2013). Impact of the fine-tuning medium-of-instruction policy on learning: Some preliminary findings. *Literacy Information and Computer Education Journal*, *4*(1), 1037–1045. doi: 10.20533/licej.2040.2589.2013.0138

Rollnick, M. (2000). Current issues and perspectives on second language learning of science. *Studies in Science Education*, *35*(1), 93–121. doi: 10.1080/03057260008560156

Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE Life Sciences Education*, *11*(3), 294–306. doi: 10.1187/cbe.11-11-0100

Stoller, F. L. (2008). Content-based instruction. In N. van Deusen-Scholl & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 4, pp. 59–70). doi: 10.1007/978-0-387-30424-3_89

Swart, A. J. (2010). Evaluation of final examination papers in engineering: A case study using Bloom's taxonomy. *IEEE Transactions on Education*, *53*(2), 257–264. doi: 10.1109/TE.2009.2014221

Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of Further and Higher Education*, *33*(1), 57–69. doi: 10.1080/03098770802638671

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy and Practice*, *11*(1), 49–65. doi: 10.1080/0969594042000208994

Yip, D. Y., Coyle, D., & Tsang, W. K. (2007). Evaluation of the effects of the medium of instruction on science learning of Hong Kong secondary students: Instructional activities in science lessons. *Education Journal*, *35*(2), 77–107.

Yip, D. Y., & Tsang, W. K. (2007). Evaluation of the effects of the medium of instruction on science learning of Hong Kong secondary students: Students' self-concept in science. *International Journal of Science and Mathematics Education*, *5*(3), 393–413. doi: 10.1007/ s10763-006-9043-x

Yip, D. Y., Tsang, W. K., & Cheung, S. P. (2003). Evaluation of the effects of the medium of instruction on the science learning of Hong Kong secondary students: Performance on the science achievement test. *Bilingual Research Journal*, *27*(2), 295–331. doi: 10.1080/ 15235882.2003.10162808

Appendix 1: Taxonomy Table

The knowledge dimension	The cognitive process dimension					
	1. Remember	2. Understand	3. Apply	4. Analyze	5. Evaluate	6. Create
A. Factual knowledge						
B. Conceptual knowledge						
C. Procedural knowledge						
D. Metacognitive knowledge						

Sources: Adapted from Anderson and Krathwohl (2001).

Appendix 2: Exemplars of Test Paper Questions

Exemplar 1

Instruction for 3(a–c): You may choose to answer either level 1 or level 2. If you have answered both levels, *only level 1 will be marked*.

The diagram below shows the water cycle. (20 marks)
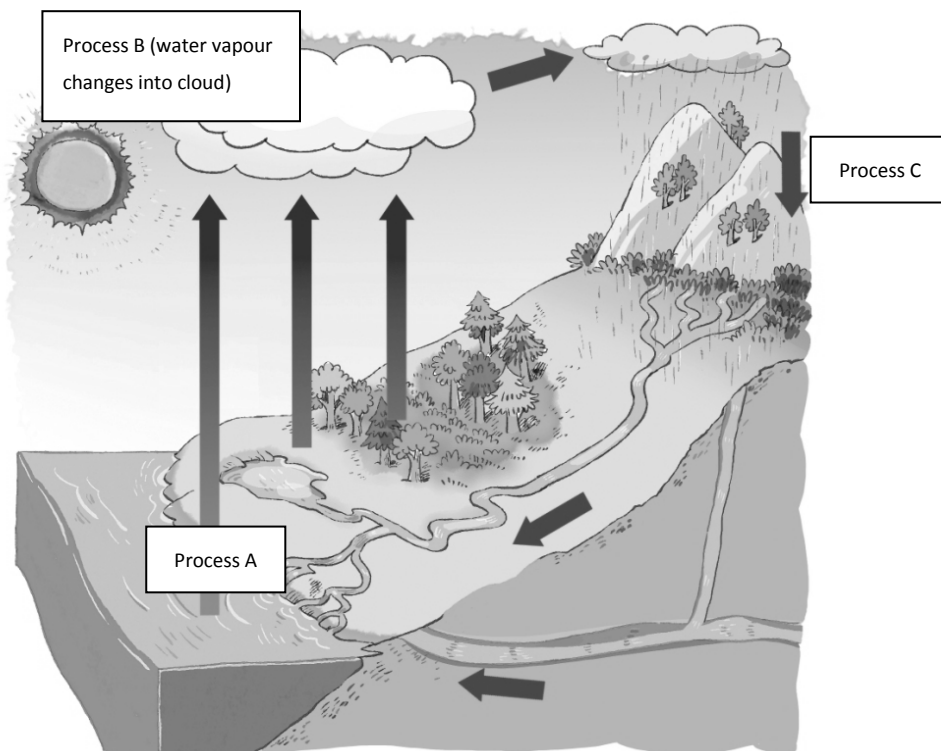
(a) **Level 1**

What is process A? (2 marks)

Process A is _____.

Level 2

What is process A? Describe how process A takes place. (6 marks)

Process A is ____(i)____. When the ____(ii)____ heats up the water in the ____(iii)____, river and on land, the water absorbs ____(iv)____ energy and ____(v)____ to become water vapour.

(b) **Level 1**

What is process B? (2 marks)

Process B is _____.

Level 2

What is process B? Describe how process B takes place. (6 marks)

Process B is ____(i)____. Water vapour cools down and ____(ii)____ which join together to form ____(iii)____.

(c) **Level 1**

What is process C? (2 marks)

Process C is _____.

Level 2

What is process C? Describe how process C takes place. (6 marks)

Exemplar 2

The picture below shows a wind turbine.

[A picture of a wind turbine]

(a) Explain why wind power is called a "green power"?

(b) Why wind power is not commonly used in Hong Kong? Give two reasons.

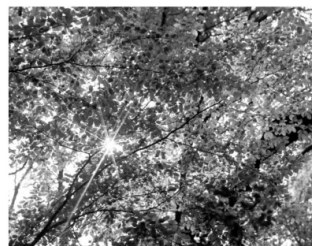## *Exemplar 3*

Write down the energy conversion involved in each of the following situations.

(a) a piece of charcoal
     (炭) is burnt in a
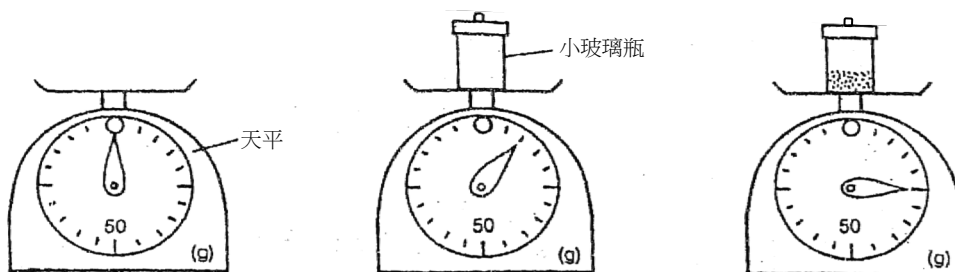   barbecue fire
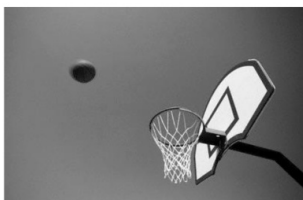
(b) a falling basketball

(c) sunlight falls on the
     leaves of a tree

Exemplar 4

下圖所示為量度糖重量所需的各步驟。

(The following pictures show the steps of measuring the weight of the sugar.)

糖的重量是多少？請列出算式。

(What is the weight of the sugar? Show your steps of calculation.)

Exemplar 5

下圖所示為兩個位於河邊的 A 城和 B 城。該處的居民都是從河中獲取日常用水的。A 城排放了許多污水到河中，這對 B 城會造成甚麼影響？A 城在排放污水前應該做些甚麼？

(The locations of City A and City B are shown in the following map. Citizens in both cities attain water from the river for their daily-life use. As City A has discharged plenty of polluted water into the river, what would be the impact on City B? What should people in City A do before discharging the polluted water?)

[A simple map indicating the location of
City A (labelled as "A 城") and City B (labelled as "B 城")
and the water flow of the river]

Exemplar 6

Consider the following path of a file.

C:\school\software.avi

(a)   Write down the drive name, directory name and file name.
(b)   Which part of a file name shows its file type?
(c)   What is the type of the above file?

Exemplar 7

細閱下圖，然後回答各部分問題。

(Read the following pictures carefully and answer the questions)

C 觀點 － 施永青

教育能提升社會流動嗎？

近年，年輕人對社會的怨氣很大。他們覺得社會的向上流動能力差了，即使讀完大學，也不易找到有發展前景的工作，很多人對自己的將來都沒有掌握。政府在教育上的投資，似未能收到預期的效果。

雖然，知識一直被視作改變命運的工具，但這只對個人有效，一上升至社會層面，情況就變得複雜。影響一個社會命運的，可不只教育一項，還有政治因素、經濟因素、自身因素、外圍因素、歷史因素、當今因素……。菲律賓的大學就培養了很多大學生，但他們畢業後，卻選擇去做不用多少知識的家庭傭工。

我認識一個朋友就經常感嘆，他兒子大學畢業後，在社會上找到的機會還不如他以前小學畢業！

資料來源：節錄自〈教育能提升社會流動嗎？〉（http://hk.centanet.com/home/ColumnArticle.aspx?id=25179）

列舉三項這篇文章採用了的各式功能。

(Suggest three formatting functions that have been used in this paragraph.)

Acknowledgments

香港微調教學語言政策：科學學科校本試卷轉變的個案研究

陳以謙

摘　要

　　本研究以一所中學為例，探討香港微調教學語言政策於實施前後，初中校本科學學科（即綜合科學科和電腦與資訊科技科）試卷的轉變。由於該校初中班別部分學術科目從母語教學轉為以英語授課，相關校本試卷亦轉以英語提問和作答。試卷的分析主要分為兩個階段：首先，研究集中描述試卷在設計和格式上的變化；其次，試卷中所有題目以修訂版布盧姆教育目標分類學來分析其認知和知識兩個範疇。研究結果顯示，新政策下試卷的問題類型、考核內容及得分比例均作重新分布，其中認知和知識範疇較高層次的題目比例相對減少。本文最後總結該初中校本試卷變化可能帶來的影響，並指出新教學語言政策的影響仍有需要繼續研究。

關鍵詞：評估；學科內容語言教學；教學語言；微調政策；布盧姆教育目標分類學的　　　　　修訂版；香港