British Journal of Educational Psychology (2011), 81, 531–557

The British Psychological Society

www.wileyonlinelibrary.com

Effects of a 'Learn to Think' intervention programme on primary school students

Weiping Hu[1,2]*, Philip Adey[3], Xiaojuan Jia[4], Jia Liu[4], Lei Zhang[4], Jing Li[4] and Xiaomei Dong[4]

[1]Shaanxi Normal University, Xi'an, PR China
[2]Beijing Normal University, Beijing, China
[3]Kings College London, UK
[4]Shanxi Normal University, Linfen, Shanxi, PR China

Background. Methods for teaching thinking may be described as out-of-context or infusion. Both approaches have potential to raise students' general cognitive processing ability and so raise academic achievement, but each has disadvantages.

Aims. To describe and evaluate a theory-based learn to think (LTT) curriculum for primary school students, which draws on the strengths of both out-of-context and infusion approaches.

Sample. One-hundred and sixty-six students in three classes of Grade 1 (6 + years old), Grade 2 (7 + years old), and Grade 3 (8 + years old) in a primary school in Shanxi province, China, randomly ascribed to experimental (90) and control (76) groups.

Methods. All students were pre-tested for non-verbal intelligence and academic achievement. Experimental students followed the LTT curriculum (one activity every 2 weeks) for 4 school years. All were post-tested on three occasions for thinking ability and four times for academic achievement.

Results. Grade 1 and Grade 2 students showed effects of LTT from 1 year after their start and increasing: on thinking ability $d = .78$–1.45; on Chinese $d = .68$–1.07; on maths .58–.87. Grade 3 students showed effects from 6 months after their start: on thinking ability .90–1.37; Chinese .77–1.32; maths .65–1.29. The effects were concentrated in students in the middle band of initial ability.

Conclusions. A curriculum for teaching thinking based on a structured theoretical model that combines elements of out-of-context and infusion methods has been shown to have long-term far transfer effects on students' thinking ability and academic achievement. More work is needed to meet the needs of a wider range of abilities.

The development of thinking ability is one of the most basic goals in children's education. For the individual, the justification is that developing the mind is the best possible

**Correspondence should be addressed to Professor Weiping Hu, Center for the Development of Teacher Professional Ability, Shaanxi Normal University, Xi'an 710062, PR China (e-mail: weipinghu@163.com).*

DOI:10.1348/2044-8279.002007

preparation for the often-unpredictable demands and opportunities that life holds (Kuhn, 2005; p. 4). Students have been traditionally assessed by how much information or facts they can reproduce in tests and examinations, but now there is a wide awareness that students should be assessed by how well they are able to think or process information.

Since the early 1950s, various programmes for teaching thinking have been developed. They have been comprehensively reviewed by Nickerson, Perkins, and Smith (1985), Coles and Robinson (1989), Nisbet and McGuinness (1990), and McGuinness (1999, 2000). They can be categorized into two types. One is an out-of-content (or 'bolt-on', Dewey & Bento, 2009) approach such as Lipman's Philosophy for Children (Lipman, Sharp, & Oscanyan, 1980; Topping & Trickey, 2007a, 2007b), De Bono's (1970) lateral thinking and cognitive research trust (De Bono, 1987), Feuerstein and colleagues' (Feuerstein, Rand, Hoffman, & Miller, 1980) Instrumental Enrichment, Greenberg's (1989) cognitive enrichment programme and Buzan's (1984) mind mapping. These programmes are usually based on analysis of component skills in thinking, which are taught and practised in special courses. The main criticisms of this approach (though this does not apply to all) are that it treats thinking as an 'add-on' element, that the skills approach is reductionist or fragmentary, and that transfer of these skills to new contexts is by no means guaranteed.

The other category is the infusion approach in which the thinking is integrated into one or more school subjects. For example, Guilford (1967) designed a model based on his structure of intelligence to teach thinking skills in children through problem solving. Williams (1972), based on his cognitive–affective interaction theory, designed a teaching model for promoting creative thinking that emphasized the application of creative thinking strategies in the classroom. Sternberg (1996a) developed a series of specific thinking strategies to increase children's thinking ability. Adey and Shayer (1994), based on Piagetian ideas of cognitive conflict and the schemata of formal operational thinking, and on Vygotskyan ideas of the social construction of understanding, developed an intervention named CASE (cognitive acceleration through science education). Chongde Lin (1999) developed a learning and development theory and applied it to classroom teaching in various subjects. Those who favour *infusion* argue that thinking cannot and should not be separated from its context, that this approach is more readily incorporated into current practice, and that transfer is more likely if thinking is embedded in all teaching and learning. Content and process are both important: thinking is always thinking about something. But often problem solving is seen only as a better method of teaching a subject, rather than as a way of teaching thinking – and then there is little transfer. Lipman (1985; p. 106) points out that the traditional curriculum is fragmented, taught in isolated subjects. Schools did not have formal instruction in thinking because educators wrongly believed that students would acquire thinking skills automatically through subjects (Ruggiero, 1995; pp. 5–6). Students have been taught what to think, but rarely how to think.

In response to criticisms of the two approaches to the training of thinking, a 'Learn to Think' (LTT) curriculum has been designed, which draws on the strengths of both types. The theoretical underpinnings of this study embrace notions from Chongde Lin's theory of intelligence, Piaget's cognitive development theory, and Vygotsky's social construction theory. The model, curriculum design, and principles of curriculum delivery, as well as the effects of intervention on primary school students will be introduced in this paper.

Thinking ability structure model

On the basis of Chongde Lin's theory of intelligence and viewpoints of good thinking, a three-dimensional thinking ability structure model (TASM) is proposed by Hu. Lin and Li (2003) proposed a thinking structure consisting of six components: self-regulation, purpose, materials, process, non-cognitive factors, and qualities and outcomes of thinking. Self-regulation of thinking is the supreme commander of the whole thinking structure, an invisible self-underlying the visible self. Purpose of thinking refers to the direction and the expected outcome of thinking activities, or the realization of such functions of thinking as adaptation. The materials of thinking can be divided into two categories: concrete and abstract materials. Concrete materials include senses, perceptions, images, etc. The abstract material mainly refers to concepts. Processes of thinking include searching, discriminating, representing, imaging, comparing, classifying, analyzing, and synthesizing of thinking materials. Non-cognitive factors of thinking include mental factors that are not directly involved in cognitive process yet have a direct effect on them, including motivation, interests, emotions, affections, will, temperament, personalities, etc. Qualities of thinking are the criteria of assessment for the outcome of thinking. There are five main qualities of thinking: profundity, flexibility, originality, criticism of thinking, and agility. Training the qualities of thinking is the breakthrough point in the cultivation of thinking ability.

But what counts as good thinking? There are three viewpoints. The first is understanding of knowledge. Since the 1980s, one strong contention has been that thinking ability is inevitably domain-specific. The philosophical argument for the function of knowledge in good thinking is expressed by McPeck (1994; p. 111), who emphasizes the 'so obvious and commonsensical' ideas that (1) there is no generalized thinking, only thinking about something; (2) a good thinker on one matter is not necessarily a good thinker on another matter; (3) the quality of thinking depends on the amount of knowledge about the topic and in the discipline to which it belongs; (4) teaching thinking must focus on teaching for understanding of the theoretical discipline. Some empirical studies suggest that the main factor in the good thinking of experts is knowledge, or rather their understanding of knowledge (Perkins & Salomon, 1989) and that it is knowledge that distinguishes the expert from the novice thinker.

The second viewpoint of good thinking incorporates the ideas of 'thinking skills'. This viewpoint suggests that instead of imparting bodies of knowledge, we must impart to our students abilities of good thinking. Thus, the 'educational market' would be filled with thinking skills of various qualities, for example, skills of critical thinking, of creative thinking or effective thinking (Marzano *et al.*, 1988). Sternberg (1996b; pp. 127–146) argued that successful intelligence included analytic intelligence, creative intelligence, and practical intelligence but that these three need to work together.

The third viewpoint on good thinking relates to dispositions. Scholars in the field of teaching thinking use various concepts to describe the dispositional dimension of thinking. Dewey (1998; pp. 29–33) wrote of three attitudes: open-mindedness, wholeheartedness, and responsibility. Sternberg (1996b; pp. 251–259) listed 20 personal qualities that qualify 'successful intelligence'. Perkins (1995; pp. 284–285) listed seven dispositions.

Some integration of these three viewpoints was provided as long ago as 1943 by Russell who proposed that critical thinking had four components: knowledge of the field or fields in which thinking is being done; attitude and habit of questioning and suspending judgment; and the application of logical and scientific method to problem situations, and

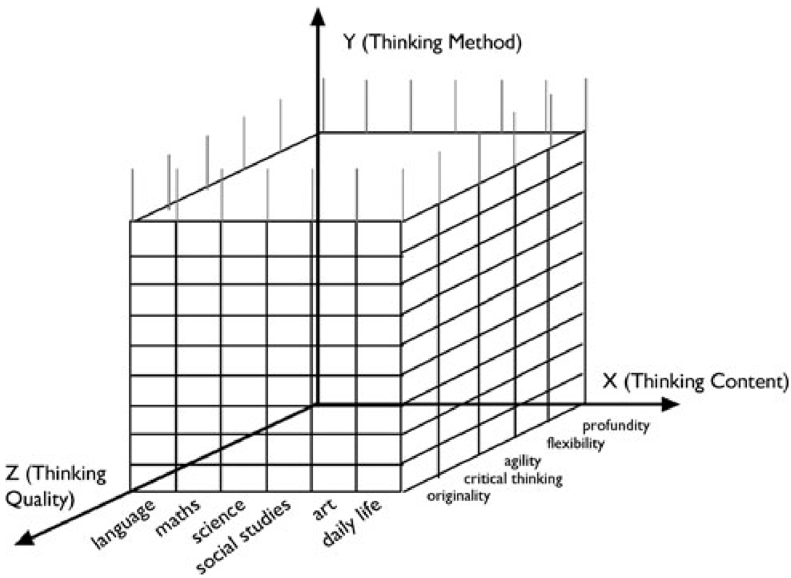


Figure 1. The thinking ability structure model (TASM).

taking action in light of this line of thinking (Russell, 1943; p. 746). Subsequently, he offered a concise definition: critical thinking seems to involve attitude plus knowledge of facts plus some thinking skills (Russell, 1960; p. 651). The TASM builds on this definition. It is an integrated model of thinking content (understanding of knowledge), thinking methods (thinking skills), and thinking quality (kinds of dispositions), which is also cognizant of the psychological foundation outlined at the start of this section (Lin) (Figure 1).

The X axis is thinking content, Y axis is thinking method, and Z axis is thinking quality. The model has three basic characteristics. First, it is a whole system consistent with the idea that thinking ability consists of content, method, and quality of thinking. They depend on each other, facilitate each other, develop together, and form an integrated system. Second, the model is not only static, but also has a developmental nature. Thinking ability should be a combination of both static and developmental structures. As knowledge is acquired, methods elaborated, and quality advanced, thinking ability develops and the model becomes more complex and fully integrated. Third, the model has the nature of self-regulation under the influences of purpose of thinking, non-cognitive factors, the environment, and social factors.

Each two-dimensional plane of the model has special meaning: the x–y plane (thinking content × thinking method) represents subject structure; the y–z plane (thinking method × thinking quality) represents general thinking ability, and the x–z plane (thinking content × thinking quality) represents understanding of the content.

The TASM is consistent with the theories and practices of teaching thinking summarized above. Relating it to Lin's six components of thinking, the thinking content dimension of the TASM represents the *materials* of thinking and *outcomes* of thinking. The thinking quality of the TASM is the thinking quality of Lin's model. The thinking method of TASM represents the *process* of Lin's model. Although the *purpose*, *self-regulation*, and *non-cognitive factors* of thinking are important parts of thinking structure, which influence the development of thinking ability, they do not belong to area of ability.

A 'LTT' curriculum

Based on TASM, Piaget's cognitive development theory, and Vygotsky's social construction theory, an 'LTT' curriculum has been designed consisting of a series of learning situations that are planned to be conducive to the development of students' thinking methods or strategies and improving the quality of thinking.

The structure model TASM is designed as a theoretical foundation for the cultivation of thinking ability. It tells us that the cultivation of thinking ability requires the teaching of thinking methods, the training of thinking quality and that these must be set within the context of a body of knowledge.

Piaget proposed that children proceed through four stages of cognitive development: sensorimotor, pre-operations, concrete operations, and formal operations. Each stage has major cognitive tasks that must be accomplished. In the sensorimotor stage, the mental structures are mainly concerned with the mastery of concrete objects. The mastery of symbols takes place in the pre-operational stage. In the concrete stage, children learn mastery of classes, relations, and numbers and how to reason. The last stage deals with the mastery of thought (Evans, 1973).

Vygotsky asserted that children's development can be fostered both by adults and more competent peers when working in the 'zone of proximal development' – the difference between 'the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers' (Vygotsky, 1978; p. 86). Based on the notion of a zone of proximal development, Vygotsky propounded a new formula, namely that the only 'good learning' is that which is in advance of development (p. 89). Vygotsky argues that social interaction is essential for children's development and that by verbalizing their reasoning they accept reasoning at a higher level than they start out with.

The theoretical foundations for the design of the difficulty of activities of the LTT curriculum embrace notions from Piaget's account of cognitive development that allows specification of the cognitive complexity of tasks, Vygotsky's zone of proximal development and the principle that good learning must be in advance of development. On these bases, the difficulty of activities is adjusted not just to match but to provide challenge and so promote the development of students' thinking ability.

Thinking contents—X axis

Because abilities as measured are largely domain-specific, activities of the thinking curriculum are contextualized in almost all areas studied by primary school students including mathematics, language and literature, science, society, art, other disciplines, and daily life experience. The X axis includes six factors.

Thinking methods—Y axis

The curriculum incorporates the thinking methods of concrete thinking (observation, association, imagination, space cognition), abstract thinking (comparison, combination, classification, inductive and deductive reasoning, analysis and synthesis, abstraction and generalization, composition, philosophic thinking), and creative thinking (reorganization, analogy, brainstorming, lateral thinking, divergent thinking, transfer, set break through, question asking, exploring activity, story inventing). So the *Y* axis contains 22 factors.

Thinking qualities—Z axis

Thinking qualities promoted by the curriculum include profundity, flexibility, critical thinking, agility, and originality. As a result, there are five factors in the *Z* axis.

In total, the three-dimensional TASM described above offers 660 cells, and the *x–y* plane representing subject structure provides 132 cells for each of which activities may in principle be designed. Drawing on their experience of primary and secondary education in China, we interviewed primary and secondary school teachers of different subjects, and asked them to select those which they thought would stand for the main thinking abilities in primary and secondary school students. The result suggested that some aspects were believed to be more important than others and on this basis, activities for all children in primary and secondary schools were designed. Every grade from 1st through 8th has its specific manual, each including 16 activities covering different thinking content and methods. The thinking contents are used as the carrier, the thinking methods are the main thread, and thinking qualities are trained in each activity. For example, the 16 activities of Grade 1 involve 4 thinking contents and 15 thinking methods; concretely speaking, there are 3 thinking methods (step-by-step observation, imagination, and story inventing) in the content of language and literature covering 2 activities, 1 thinking method (reasoning) in maths involving 2 activities, 4 (imagination, classification, reorganization, and divergent thinking) in art contained in 4 activities, and 10 (contrastive and sequential observation, space cognition, comparing, analogy, and so on) in daily life contained in 8 activities.

To achieve its aims of cultivating children's thinking ability, creativity, academic performance, motivation to learn, learning strategy acquisition, and sense of self-efficacy, we need not only well described curriculum content but also advanced teaching methods, expressed in the following five principles:

Stimulating interest and motivation

Cognitive research takes account of affective and social aspects, and, interest and motivation play an important part in thinking because thinking requires effort. Lin proposes non-cognitive factors, including motivation, interests, etc., have a direct effect on cognitive processes. Therefore, our thinking curriculum must aim not only just to teach skills and knowledge, but also to develop interest and motivation in their use. All aspects of activity selection, from choosing activity content, materials, and situations to producing cognitive conflict, teacher–children social construction, and reflection or transfer, are focused on stimulating children's learning interest and motivation, encouraging children to explore learning methods and strategies, and staying positive and active in the acquisition of scientific thinking.

Cognitive conflict

This term comes from CASE designed by Adey and Shayer (1994; p. 62). It is used to describe an event or observation that the students finds puzzling and discordant with previous experience or understanding. Cognitive conflict is an effective means to stimulate children to think actively and can lead to constructive mental work by students to accommodate their conceptual framework to the new type of thinking necessary. It is a feature both of Piaget's account of the impact of environmental stimulus and children's constructivist response on cognitive growth, and of teaching thinking programmes.

Social construction

This comes from Vygotsky's contention that social interaction is central to children's development. The teaching environment for social construction needs to be constructed, and teacher–child interactions and child–child interactions are emphasized in the delivery of the LTT curriculum. Teachers allow students to explain their reasoning to each other and to learn from each other's errors through co-operative learning. Discussion is a well-established method, but it must involve analysis of the processes of argument if it is to be effective in teaching thinking.

Self-regulation and meta-cognition

Underlying all these methods is the principle of meta-cognition, or self-regulation. Meta-cognition, a concept introduced by Flavell in 1976, is the awareness and control of your own thinking processes. Perkins and Salomon (1989) claim that meta-cognition is likely to be an essential element of any programme that is successful in improving general thinking skills. Adey and Shayer (1994; p. 68) also use this principle in CASE, and it is the overarching component of Lin's thinking structure. The aim here is to give pupils practice in monitoring their own thinking, with the teacher initially making their strategies explicit and the learners then internalizing them, making them part of their habitual mode of thinking. Thus, learning to learn means taking over from the teacher the control and management of your own learning and thinking. At the end of each activity, the students should reflect and summarize the thinking methods, thinking strategies, problem finding and solving methods, and what he/she has got out of the activities.

Application and transfer

Generally, an activity in LTT only belongs to a specific domain. The thinking methods and strategies studied in the activity need to be applied and transferred to daily life or other domains for training the thinking qualities and forming general habits of effective thinking. So there is broadening content (or 'bridging' in Shayer and Adey's terminology) in each activity.

The 'LTT' curriculum involves three types of training: basic thinking strategy training, problem-solving skills training, and creative thinking skill training. Basic thinking skills training is completed through establishing the learning situation, thinking method recognition, method deduction, method application, evaluation and reverse transfer consolidation, and transfer consolidation. Problem-solving skill training is completed through introducing a problem, problem analysis, brain storming, selection of the best method, evaluation and reflection, and consolidation *via* transfer practice. Creative thinking skill training is completed through task introduction, preparation activity, deductive reasoning, brain storming, choosing the best answer, producing a result, evaluation and reflection, and consolidation transfer.

In summary, LTT has five characteristics. (1) *Suitability*. The difficulty of each activity is in the 'zone of proximal development', and it is not only suitable for, but also can promote the development of thinking ability of students. (2) *Systematization*. The whole project covers almost all the thinking methods appropriate for students across the eight grades from the 1st grade of primary school to the 2nd year of secondary school. Each thinking method appears in different grades, and the thinking difficulty increases with grade. The whole curriculum trains the thinking ability of students step by step in a spiral manner. For example, categorization is arranged as follows: Grade 1 (age 6 years): categorize shapes and objects/things; Grade 2: categorize unfamiliar and living things and cross categorization; Grade 3: use categorization in problem-solving tasks; Grade 4: categorize based on purpose and the limitedness of categorization; Grade 5 (10 + years): categorize three-dimensional graphs and recognize the multi-dimensionality of standards. (3) *Diversity*. Drawing on the strong points of teaching thinking programmes, activities of LTT include thinking method training (such as *Elephant Vs Slide*: analogy; *Daedal Organism*: compare; *The Same World*: analysis and generalization), space cognition (*To help uncle goat find his seat*), problem finding (*Earth, I want to know*), problem solving (*Learn Ask*), scientific inquiry story inventing (*Courser and Brawler*), and so on. (4) *Elicitation*. Each activity is planned to induce some cognitive conflict in students, encouraging them to think actively (Adey & Shayer, 1994; pp. 62–65). (5) *Gradualness*. The arrangement of activities accords with the character of student thinking, from simple to complex, from shallow to deep, from easy to difficult, step by step, to enable students improve the ways of thinking gradually. The details of a lesson are given in Appendix A.

Methods

Participants

Participants for the study were 166 1st–3rd grade students (90 boys, 76 girls) in three classes from Huo Zhou Primary School in Shanxi Province in China. The age of Grade 1 students was 6 years and up (6 +), Grade 2 students' age was 7 and up (7 +), and Grade 3 students' age was 8 and up (8 +). By random sampling, 30 students in the class of Grade 1 (6 +) were selected as the experimental group, and the remainder in the same class (typically another 25–30 children) designated as the control group. By stratified random sampling, 30 students in each class of Grade 2 (7 +) and Grade 3 (8 +) were selected as the experimental group, and the others as the control group. While this approach does run a small risk that in their normal lessons, the control group will be 'contaminated' with better thinking by the activities of the experimental group in a shared classroom environment, as seemed to happen in Hautamäki, Kuusela, and Wikström's study (2002), the advantage is that experimental and control groups can be seen to have identical educational experiences in every respect except the LTT lessons. Even this risk, if realized, would lead to a type 2 error (missing an effect that was there) and so offers no threat in this case (see 'Results').

In order to explore if there were different effects on students of different academic achievement levels, before the experiment students of Grade 2 and Grade 3 were categorized into three groups (high-score group, mid-score group, and low-score group) according to their final examination results of the previous term, and students of Grade 1 were grouped according to their final examination results of first term. Of the sample, 90 participants (44 boys, 46 girls) followed the LTT curriculum, and 76 participants (46 boys, 30 girls) did not.

Materials

The 'Learn to Think' curriculum

This was described in the last section. Of course the key to its implementation is the effective delivery of the LTT curriculum by teachers. In this case, the LTT lessons were taught by members of the research team who had participated in a 3-day professional development course before the experiment, while the control groups were supervised by the students' head teachers as they learned by themselves or did their homework.

Raven standard progressive matrices test (RSPM)

Raven standard progressive matrices test (RSPM) was compiled by British psychologist J.C. Raven in 1938 and is still in wide use today. It is a non-verbal intelligence test that measures pattern recognition and inductive thinking. The reliability and validity of RSPM is well established.

Self-designed thinking-ability test

A thinking-ability test was developed, based on analysis of the structure of thinking-ability model. It has 54 multiple-choice (one from four) items, and each correct answer is scored one point. Thinking can be divided into abstract thinking and concrete thinking. Abstract thinking mainly includes analysis and synthesis, inductive and deductive reasoning, abstraction and generalization, comparison and classification abilities. Concrete thinking mainly includes imagination, association, space cognition, and so on. Items were developed in the light of the developmental level of primary students' thinking. The test comprises six sub-tests (with points distribution): comparing and classification (12), inductive reasoning (10), deductive inference (7), spatial cognition (9), analogical reasoning (9), and abstract-generalization (7). Cronbach's alpha reliability coefficient of the whole test was 0.89, the split-half reliability was 0.71, and the Pearson's correlation coefficient between the test and RSPM was 0.67, so it had a high reliability. The correlation coefficients among the sub-tests were lower, whereas the correlation coefficients between the whole test and each sub-test were higher, ranging from the lowest 0.64 to the highest 0.80, which suggests a high validity. Item analysis revealed that it had moderate difficulty and good discrimination. Confirmatory factor analysis showed that the fit index for six factors is very good (NFI, CFI, IFI, GFI, and NNFI $>$ 0.95, RMSEA $<$ 0.08[1]). This suggests that test has good construct validity. A sample of items from the test is given in Appendix B.

Academic achievement test

In China, primary and secondary school students usually take two formal exams (mid-term exam and final exam) in one term. We chose the students' final examination score as a measure of their academic achievement in this study. The final exam is a criterion-referenced test, which aims to test if knowledge they mastered in this term is up to curriculum standards and the degree they have reached.

---

[1] *The normed fit index, comparative fit index, incremental fit index, goodness of fit index, and non-normed fit index should be larger than 0.9. The root mean square error of approximation should be under 0.08.*

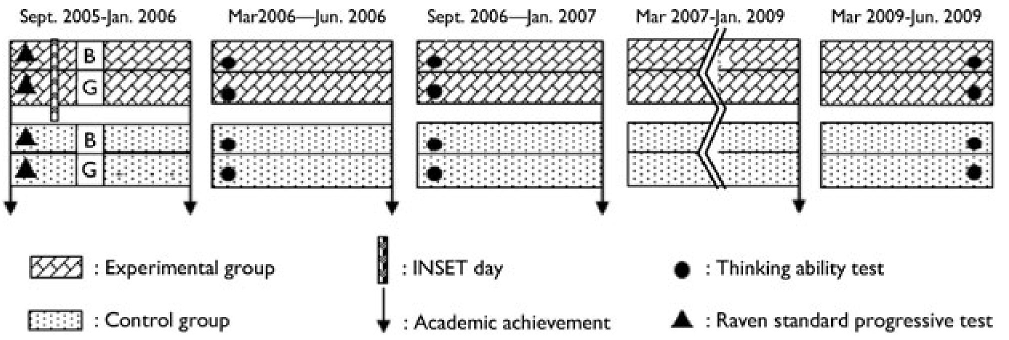


Figure 2. Experimental design of the LTT project (note that the horizontal time scale is not uniform).

Procedure

For each of the three classes that formed the sample, experimental and control groups had the same normal teacher, teaching conditions and so on, but the only difference was that the students of the experimental group (approximately half of the class) attended an LTT curriculum lesson once every 2 weeks, which was taught by a member of the research team, while students in the control group learned by themselves or did their homework, monitored by their head teachers. Note that the extra teaching received by the experimental group amounts to only an extra 16 h per year compared with a total of over 750 h of teaching in a school year (about 2%).

The school year starts in September and is divided into two semesters. Before the formal experiment, we collected 2nd and 3rd grade students' final examination results of the last term, and measured all of the participants with RSPM. Figure 2 displays the detailed experimental design. It shows that during the 4-year intervention, end of year examination results were collected each year, while participants were measured by the thinking-ability test three times. Previous research found that the effect of LTT mainly emerged after 1-year intervention and we wanted to confirm this (with a measures at 6 and 12 months) and then look for the long-term effect with the measure at the end of 4 years.

Data analysis

t-Tests (using SPSS 16.0) were used throughout the whole data analysis to examine differences between experimental and control groups on every test and the final examination results. All effect sizes were calculated using Cohen's *d* (Cohen, 1992; Thalheimer & Cook, 2002) and denoted as being small (.20), medium (.50), or large (.80) according to conventional operational definitions widely used.

Results

Pre-test scores

Comparison of RSPM scores between experimental group and control group

No significant differences were found between the groups' RSPM scores (Table 1) indicating that before the experiment the non-verbal intelligence of the experimental and control groups' students were not significantly different.

Table 1. Mean pre-intervention scores on RSPM

	Experimental ($M \pm SD$)	Control ($M \pm SD$)	E-C differences t	p
Grade 1 (6+)	16.45 ± 6.26	18.41 ± 5.85	1.139	.260
Grade 2 (7+)	32.35 ± 7.33	31.00 ± 4.48	.809	.422
Grade 3 (8+)	36.94 ± 4.23	35.08 ± 5.37	1.475	.146

Table 2. Mean pre-intervention scores on Chinese and maths tests

		Experimental ($M \pm SD$)	Control ($M \pm SD$)	E-C differences t	p
Grade 2 (7+)	Chinese	85.85 ± 8.94	85.70 ± 8.32	.066	.947
	Maths	86.35 ± 8.16	88.20 ± 9.37	.787	.435
Grade 3 (8+)	Chinese	86.93 ± 6.32	86.40 ± 4.69	.352	.726
	Maths	91.47 ± 7.83	89.52 ± 6.56	.944	.325

Comparison of the academic achievement of experimental and control groups

No significant differences were found in Chinese or maths scores between the experimental and control group students before the experiment indicating that they started from essentially the same level of academic achievement (Table 2).

Development of thinking ability over 4 years

All students did the thinking-ability test and its parallel questionnaires three times. Table 3 and Figure 3 show the scores of the experimental and control groups by grade and their differences on each testing occasion.

After nearly 1 year of the LTT intervention, the overall level of the thinking ability of experimental group students in Grade 1 and Grade 2 had significant differences from the control group ($p < .01$); in Grade 3 a significant difference emerged after just 6 months ($p < .01$); after 4 years, the thinking abilities of all the experimental students were

Table 3. Mean scores on the thinking-ability tests for experimental and control groups (intervention started September 2005)

Grade		Experimental ($M \pm SD$)	Control ($M \pm SD$)	t	p	d
1 (6+)	March 2006	23.17 ± 6.15	22.59 ± 5.74	.343	.733	0.10
	September 2006	33.17 ± 3.27	30.59 ± 3.50	2.738	.009**	0.78
	June 2009	41.35 ± 3.27	38.05 ± 4.87	2.767	.008**	0.83
2 (7+)	March 2006	39.13 ± 4.86	38.96 ± 3.54	.149	.882	0.04
	September 2006	40.17 ± 3.57	36.56 ± 4.71	3.228	.002**	0.89
	June 2009	44.17 ± 2.28	39.65 ± 3.98	4.149	.000***	1.45
3 (8+)	March 2006	42.75 ± 2.90	38.74 ± 6.17	3.226	.002**	0.90
	September 2006	43.71 ± 3.00	39.00 ± 4.06	4.987	.000***	1.37
	June 2009	45.56 ± 3.48	40.20 ± 5.49	3.222	.004**	1.21

Note. $^{**}p < .01$; $^{***}p < .001$.

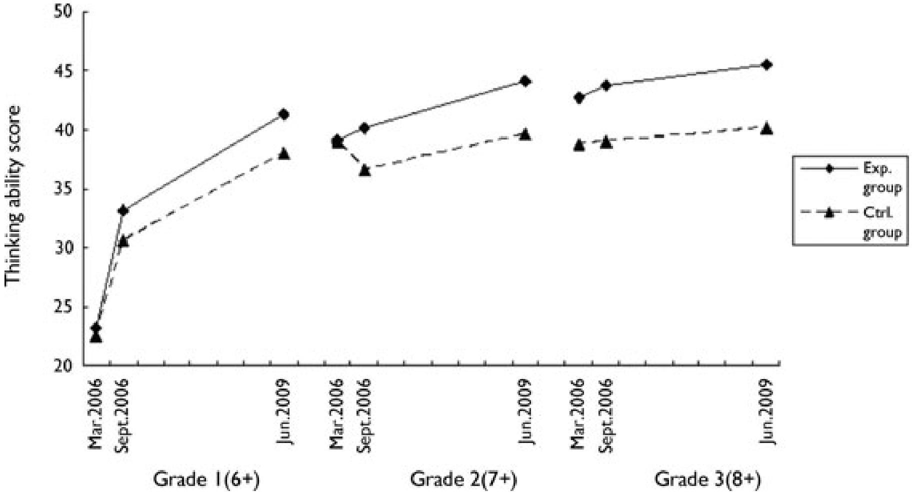


Figure 3. Comparison of thinking ability.

significantly higher than the control group ($p < .01$). Note that the thinking ability of Grade 2 students in the control group presented an obvious drop in September 2006, whereas the experimental group students' rose.

Effects of different initial ability groupings

To further explore the effects of the LTT intervention on the thinking ability of students from different initial academic achievement levels, we looked separately at the high-score, mid-score and low-score groups of students in each grade (Table 4).

It will be seen that the low-score group did not show any significant effect of the intervention at any grade. The effects of the LTT intervention on thinking ability were mainly concentrated in mid-score group students.

Table 4. All of the effects of LTT on thinking ability that reach significance for students in different starting ability bands. *t* Values of experimental-control differences

	Score-group	Low	Medium	High
Grade 1 (6 +)	March 2006			
	September 2006		2.52*	
	June 2009		3.20**	
Grade 2 (7 +)	March 2006			
	September 2006		3.22**	
	June 2009		3.27**	3.24*
Grade 3 (8 +)	March 2006		2.11*	2.65*
	September 2006		3.73***	3.25*
	June 2009		2.28*	3.26*

Note. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 5. Mean scores on the thinking-ability sub-tests for experimental and control groups (intervention started September 2005); Grade 1 (6+)

		Experimental (*M* ± *SD*)	Control (*M* ± *SD*)	*t*	*P*	*d*
Comparing and classification	March 2006	6.20 ± 1.73	6.09 ± 1.85	.218	.828	0.06
	September 2006	8.93 ± 1.20	7.91 ± 1.60	2.638	.011*	0.75
	June 2009	8.46 ± 1.30	8.41 ± 1.46	.117	.908	0.04
Inductive reasoning	March 2006	4.00 ± 1.93	4.23 ± 2.02	−.411	.683	−0.12
	September 2006	5.43 ± 1.50	5.23 ± 1.27	.521	.605	0.14
	June 2009	7.96 ± 1.25	7.00 ± 1.67	2.256	.029*	0.68
Deductive inference	March 2006	2.33 ± 1.21	1.68 ± 1.25	1.890	.065	0.54
	September 2006	3.43 ± 0.90	3.14 ± 1.25	1.000	.322	0.28
	June 2009	5.19 ± 1.55	4.33 ± 1.11	2.210	.032*	0.64
Spatial cognition	March 2006	3.47 ± 1.43	3.32 ± 1.36	.377	.707	0.11
	September 2006	3.30 ± 0.84	3.23 ± 0.87	.305	.762	0.08
	June 2009	5.85 ± 1.32	5.00 ± 1.22	2.258	.029*	0.68
Analogical reasoning	March 2006	3.00 ± 1.72	3.14 ± 1.55	−.294	.770	−0.09
	September 2006	6.17 ± 1.39	5.77 ± 1.27	1.046	.301	0.30
	June 2009	7.50 ± 0.86	6.81 ± 1.36	2.115	.040*	0.64
Abstract-generalization	March 2006	4.17 ± 1.49	4.14 ± 1.32	.076	.940	0.02
	September 2006	5.90 ± 1.27	5.32 ± 1.36	1.585	.119	0.45
	June 2009	6.38 ± 0.80	5.86 ± 0.91	2.108	.041*	0.62

Note. *$p < .05$

Table 6. Mean scores on the thinking-ability sub-tests for experimental and control groups (intervention started September 2005); Grade 2 (7+)

		Experimental (*M* ± *SD*)	Control (*M* ± *SD*)	*t*	*p*	*d*
Comparing and classification	March 2006	9.83 ± 1.74	10.00 ± 1.29	−.396	.694	−0.11
	September 2006	10.07 ± 1.01	9.16 ± 1.34	2.777	.008**	0.79
	June 2009	9.83 ± 1.50	9.35 ± 1.32	1.002	.324	0.35
Inductive reasoning	March 2006	6.73 ± 1.41	6.88 ± 1.11	−.441	.661	−0.12
	September 2006	7.60 ± 1.13	6.32 ± 1.60	3.465	.001**	0.96
	June 2009	8.78 ± 0.94	7.47 ± 1.37	3.297	.002**	1.15
Deductive inference	March 2006	3.77 ± 1.28	3.50 ± 1.03	.851	.399	0.23
	September 2006	4.03 ± 1.24	3.96 ± 1.30	.218	.829	0.06
	June 2009	5.78 ± 0.73	5.00 ± 1.06	2.537	.016*	0.89
Spatial cognition	March 2006	5.90 ± 1.24	5.81 ± 1.06	.297	.768	0.08
	September 2006	5.50 ± 1.59	5.00 ± 1.26	1.273	.209	0.35
	June 2009	6.06 ± 1.39	5.06 ± 0.90	2.499	.018*	0.87
Analogical reasoning	March 2006	6.37 ± 1.13	6.31 ± 1.41	.174	.863	0.05
	September 2006	6.53 ± 0.97	5.80 ± 1.41	2.269	.027*	0.63
	June 2009	7.67 ± 0.69	7.12 ± 0.86	2.098	.044*	0.73
Abstract-generalization	March 2006	6.53 ± 0.90	6.38 ± 0.94	.604	.548	0.17
	September 2006	6.57 ± 0.68	6.20 ± 0.91	1.706	.094	0.48
	June 2009	6.06 ± 0.42	5.65 ± 1.11	1.420	.171	0.51

Note. *$p < .05$; **$p < .01$.

Table 7. Mean scores on the thinking-ability sub-tests for experimental and control groups (intervention started September 2005); Grade 3 (8+)

		Experimental ($M \pm SD$)	Control ($M \pm SD$)	t	p	d
Comparing and classification	March 2006	9.41 ± 1.24	8.52 ± 1.16	2.676	.010*	0.75
	September 2006	10.68 ± 0.98	10.44 ± 0.96	.909	.367	0.25
	June 2009	9.31 ± 1.58	9.20 ± 1.97	.176	.862	0.06
Inductive reasoning	March 2006	9.00 ± 0.76	8.13 ± 1.66	2.612	.012*	0.73
	September 2006	8.00 ± 1.24	6.84 ± 1.57	3.089	.003**	0.85
	June 2009	8.94 ± 1.24	7.53 ± 1.55	2.795	.009**	1.04
Deductive inference	March 2006	4.66 ± 1.26	4.13 ± 1.18	1.567	.123	0.44
	September 2006	5.23 ± 1.09	4.20 ± 1.00	3.638	.001**	1.03
	June 2009	6.25 ± 1.00	4.80 ± 1.15	3.759	.001**	1.39
Spatial cognition	March 2006	5.44 ± 1.11	5.00 ± 1.24	1.374	.175	0.38
	September 2006	6.10 ± 1.27	5.12 ± 1.27	2.857	.006**	0.79
	June 2009	6.25 ± 0.77	5.13 ± 1.51	2.571	018*	0.98
Analogical reasoning	March 2006	7.81 ± 0.86	6.91 ± 1.70	2.329	.027*	0.72
	September 2006	7.10 ± 0.87	6.16 ± 1.31	3.066	.004**	0.88
	June 2009	8.50 ± 0.52	7.40 ± 1.55	2.617	.018*	1.00
Abstract-generalization	March 2006	6.44 ± 0.67	6.04 ± 1.22	1.533	.131	0.43
	September 2006	6.61 ± 0.56	6.24 ± 0.83	2.003	.050*	0.54
	June 2009	6.31 ± 0.70	6.13 ± 0.74	.689	.496	0.26

Note. $^{*}p < .05$; $^{**}p < .01$.

Effects on sub-sets of thinking

Tables 5–7 show the effects of LTT on each of the sub-scales of the thinking-ability test.

After 6 months of intervention, no significant differences were found in any sub-test of thinking ability between experimental group and control group students in Grade 1. One year later, a significant effect was found in comparing and classification, with no significant differences in the other sub-tests. After 4 years intervention, significant differences were found between experimental and control groups in all sub-scales of the thinking test except comparing and classification.

In Grade 2, after 6 months of intervention, no significant effects were found in any sub-test of the thinking ability. But after 1- year intervention, significant differences were found in comparing and classification, inductive reasoning and analogical reasoning. After 4 years intervention, significant differences between experimental and control group students were found in all sub-tests except for comparing and classification and abstract-generalization.

In Grade 3, after 6 months of LTT intervention, significant differences were found in comparing and classification, inductive reasoning, and analogical reasoning. After 1 year of intervention, significant differences were found between the experimental and control groups in all sub-tests except comparing and classification. After 4 years of intervention, significant differences were found in inductive reasoning, deductive inference, spatial cognition, and analogical reasoning with no significant differences in comparing and classification and abstract-generalization.

Table 8. Mean scores of academic achievement between experimental and control groups; Grade 1 (6+)

		Exp. ($M \pm SD$)	Cont. ($M \pm SD$)	t	p	d
Chinese	January 2006	78.60 ± 15.08	77.74 ± 12.89	.214	.832	0.06
	June 2006	88.31 ± 8.58	85.11 ± 7.97	1.369	.177	0.39
	January 2007	85.38 ± 11.07	82.15 ± 9.04	1.118	.269	0.32
	January 2009	87.44 ± 5.14	82.26 ± 10.08	2.282	.027*	0.68
Maths	January 2006	74.20 ± 14.70	71.09 ± 17.63	.692	.492	0.20
	June 2006	92.20 ± 5.60	89.59 ± 6.63	1.534	.131	0.44
	January 2007	97.06 ± 4.51	94.04 ± 6.29	2.018	.049*	0.58
	January 2009	92.17 ± 6.73	86.86 ± 10.74	2.073	.044*	0.62

Note. $^{*}p < .05$.

Table 9. Mean scores of academic achievement between experimental and control groups; Grade 2 (7+)

		Exp. ($M \pm SD$)	Cont. ($M \pm SD$)	t	p	d
Chinese	January 2006	88.96 ± 6.29	86.16 ± 8.63	1.403	.166	0.38
	June 2006	81.75 ± 6.98	76.30 ± 8.41	2.574	.013*	0.73
	January 2007	90.70 ± 5.79	84.82 ± 7.98	3.195	.002**	0.87
	January 2009	86.75 ± 4.98	81.12 ± 5.84	3.077	.004**	1.07
Maths	January 2006	89.72 ± 7.71	87.73 ± 5.29	1.009	.277	0.30
	June 2006	92.41 ± 6.20	90.65 ± 5.06	1.109	.273	0.31
	January 2007	88.32 ± 10.69	82.89 ± 8.25	2.084	.042*	0.57
	January 2009	89.81 ± 4.29	82.26 ± 12.12	2.426	.025*	0.87

Note. $^{*}p < .05$; $^{**}p < .01$.

Table 10. Mean scores of academic achievement between experimental and control groups; Grade 3 (8+)

		Exp. ($M \pm SD$)	Cont. ($M \pm SD$)	t	p	d
Chinese	January 2006	88.09 ± 6.05	86.10 ± 5.33	1.311	.195	0.35
	June 2006	88.38 ± 4.48	85.42 ± 3.11	2.807	.007**	0.77
	January 2007	79.35 ± 7.19	69.55 ± 8.03	4.898	.000***	1.32
	January 2009	85.97 ± 5.57	77.50 ± 7.59	3.597	.001**	1.31
Maths	January 2006	95.00 ± 4.64	90.93 ± 5.39	3.082	.003**	0.83
	June 2006	93.74 ± 4.56	90.62 ± 5.24	2.381	.023**	0.65
	January 2007	93.57 ± 3.88	87.85 ± 5.18	4.808	.000***	1.29
	January 2009	93.28 ± 3.83	86.22 ± 8.09	3.155	.005**	1.15

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Effects of LTT on academic achievement

Students' Chinese and maths scores were collected four times, in January 2006, June 2006, January 2007, and January 2009. Academic growth and differences between the experimental and control groups are presented in Tables 8–10 and Figure 4–6.

As can be seen, there was a very small difference in academic achievement between the experimental group and control group of Grade 1 in January 2006. Differences between the two groups increased gradually from June 2006 through January 2007, and January 2009.

From Table 9 and Figure 5 we see that in January 2006, no significant differences were found in Chinese or maths scores between the experimental group students and control group students of Grade 2, but in June 2006 the Chinese score of the experimental group students was significantly higher than the control group students, while the experimental groups' achievements on both Chinese and maths were significantly higher than control group students in January 2007 and January 2009. The effects on Chinese scores were large from January 2007, and on maths scores were large in January 2009.

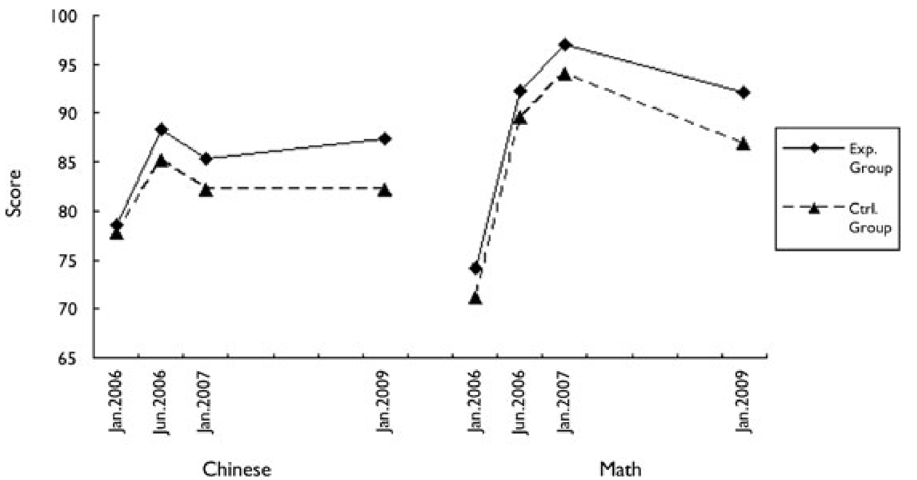


Figure 4. Academic achievement of Grade 1 (6 +).

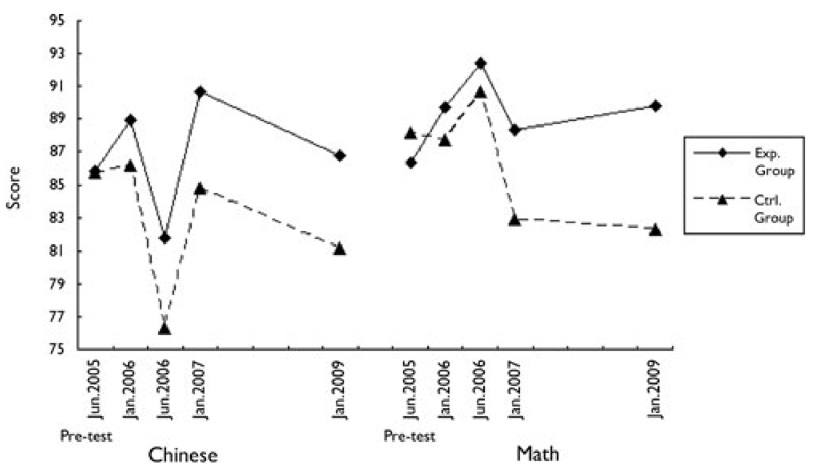


Figure 5. Academic achievement of Grade 2 (7 +).

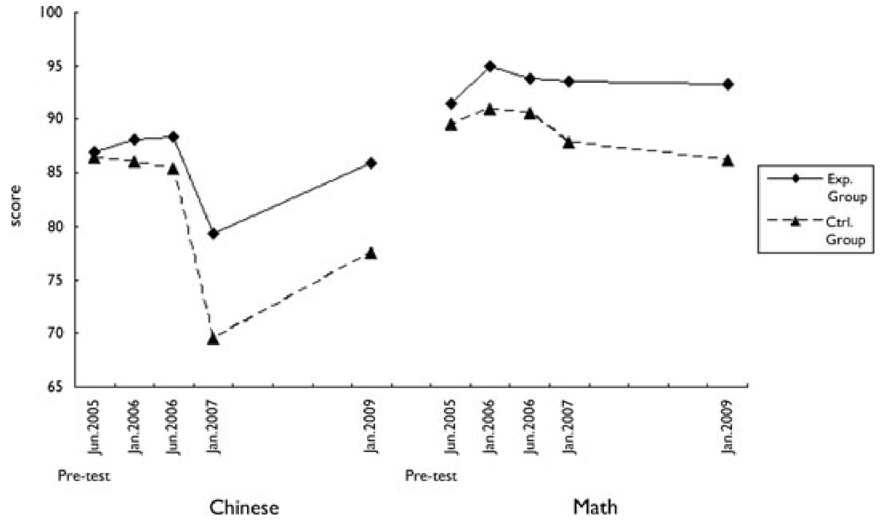


Figure 6. Academic achievement of Grade 3 (8+).

Table 10 and Figure 6 show that the maths scores of the experimental group students in Grade 3 were significantly higher than the control group students every time, while the Chinese scores were significantly higher every time except in June 2006. The effects on Chinese scores were large from January 2007, and on maths scores were large at every time except in June 2006.

Effects on students of different initial ability

The effects of the LTT intervention on Chinese and maths achievement of students from different initial academic achievement levels are summarized in Table 11, which shows wherever a significant experimental-control difference occurred in this 3 grade × 3 ability level × 3 occasions × 2 subjects matrix.

Table 11. All of the effects of LTT on subject scores that reach significance for students in different starting ability bands. *t* Values of experimental-control differences

		Chinese			Maths		
		Low	Medium	High	Low	Medium	High
Grade 1	June 2006						
(6+)	January 2007		2.07*				
	January 2009		2.11*			2.7*	
Grade 2	June 2006		2.05*				
(7+)	January 2007		2.78**			2.34*	
	January 2009		3.30**			2.20*	
Grade 3	June 2006		2.47*	3.83**			
(8+)	January 2007		3.36**	2.89*		4.04***	3.11*
	January 2009		2.78*	3.22*		2.19*	3.17*

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

The great majority of the effects were again concentrated in the middle-ability group in all grades, but there were also some effects in the high-ability group in Grade 3. There were no effects at all on the low-ability group.

Discussion

Effects of LTT on primary students' thinking ability

The LTT curriculum claims to advance methods of teaching thinking, and training thinking quality. In line with the concept of the spiral curriculum (Bruner, 1960) students were taught the methods of thinking scientifically in a hierarchical and circular way to promote the development of their thinking ability most effectively. The overall level of the thinking ability of the experimental group students was significantly higher than the control group students after the LTT intervention, but the effects on different grades appeared at different times. After nearly 1 year of LTT curriculum training, the experimental group students in the lower grades (1 and 2, now aged 7+ and 8+ years) showed significant differences from the control group students in their thinking ability. Grade 3 students showed significant effects after just 6 months of the LTT intervention (aged 8.5 years). A possible reason for this difference between Grade 1/2 and Grade 3 is differences in adaptation to school life. 1st–2nd grade students who have just entered the primary school are facing to the challenge of adapting to formal schooling, and have not formed favourable learning habits, nor even learnt how to learn, or how to answer questions. The faster effect in Grade 3 relates to the critical age at which the students' thinking is transforming from concrete imagery thinking to abstract logical thinking. It is generally acknowledged that this critical age emerges in Grade 4, about 10 years old, and if the education is appropriate, it can be advanced to Grade 3 (Lin, Wang, & Ye, 1993).

The drop in thinking-ability test scores after 1 year shown in Grade 2 of the control group suggested a general decrease in thinking ability in students at this age (8+), which is consistent with Epstein's research that shows a brain growth plateau between 8 and 9 years (Epstein, 1986). Meanwhile, the rise shown in Grade 2 of the experimental group and the significant difference between the two groups revealed that LTT clearly promoted primary school students' thinking ability. It suggests that without appropriate intervention the 'natural' development of children's thinking ability is slow and unstable, and that appropriate training is necessary to maximize children's development.

Analyzing the effects of the intervention on the different sub-scales of thinking ability is an important way to understand the development of students' thinking, giving a better chance of improving the development of thinking ability. From 1st to 3rd grades, significant differences between experimental and control groups were found on nearly all of the six sub-tests. In detail, the difference in comparing and classification was significant in the lower grades, but the significance disappeared in the middle Grade 3; significant differences on inductive reasoning and analogical reasoning appeared for the first time in Grade 2 in which the students had experienced LTT for 1 year, and the significance continued until Grade 6 (aged 11+) for students who had stated the intervention in Grade 3; significant differences on deductive reasoning, spatial cognition and abstract-generalization emerged in Grade 3 in which the students had experienced LTT for 1 year, and continued until Grade 6, except for abstract-generalization on which the significance disappeared in Grade 5 for the students who started the 4-year intervention in Grade 2. These results may be discussed as follows: firstly, the inconsistency shown in the six sub-scales is closely related to the primary students' thinking characteristics. Primary students are at the concrete operation stage, in Piaget's terms, a characteristic of which is that thinking is based on observations and images that are thought to be correct, without the formation of hypotheses. Primary students' thinking can show logicality and organization, but it is difficult for them to think about something abstract that is not immediately within perception. Comparing and classification ability is the first thinking ability developed; therefore, in the lower grade of primary school, the students who attended LTT did not show significant differences compared with control students, when they were asked to do deductive inference, spatial recognition, and abstract-generalization sub-tests that need higher abstract thinking ability, and significant difference on comparing and classification existed. Secondly, the reason why the significant differences in the five sub-tests existed in Grade 3 (aged 8 +) and continued until Grade 6 (11 +) can be discussed as follows: on the one hand, it is related with the critical age. According to Sloutsky's (2001) tag attribute model, 7–8 years old is the cut-off point of the children's special inductive reasoning ability. Third grade students who are 8–9 years old are probably in the period of accelerated development on thinking of all aspects, so the experimental 3rd grade students who attended LTT showed significant differences on all of the five sub-tests compared with the controls. On the other hand, the LTT curriculum is arranged with each activity aimed to be in the zone of proximal development, and the whole curriculum structured step by step in a spiral manner, so the significant differences continued from 3rd to 6th grade. Thirdly, the reason why the significant difference on the abstract-generalization sub-test disappeared in Grade 5 and Grade 6 may be that the abstract-generation sub-test itself is relatively easy. Because the abstract-generalization ability in our test is assessed mainly by examining the relationship between concepts and generic relations, and the test materials are from common objects that can be seen in our daily life.

Effects of LTT project on primary students' academic achievement

The LTT curriculum appears to improve primary students' academic achievement as shown by tests of Chinese and mathematics. The findings from this study fitted well with Adey and Shayer's (1994) findings suggesting that an educational intervention rooted in well-established theories of cognitive development can have long-term and replicable effects on young adolescents' academic achievement. An explanation for the findings could be as follows: firstly, thinking is one of the most important influence factors on academic achievement, and some previous researches showed that reasoning ability (Kuhn & Holling, 2009; Marjoribanks, 1976) and spatial ability (Casey, Pezaris, & Nuttal, 1992) were important for predicting academic achievement. LTT influences students' thinking ability directly through teaching students thinking methods, while the development of the thinking ability lays a solid foundation for the improvement of the students' academic achievement. Secondly, LTT is by no means a simple form of training. Besides thinking method and thinking quality, the three-dimensional TASM also includes a very important element, thinking content. Every activity of LTT is closely related to the students' study life, involving the majority of subjects' knowledge, such as Chinese, mathematics, science, society, art, and so on. The students have learnt thinking methods within a rich teaching situation, rather than simply learning the theory, process, and principle of thinking. Thirdly, LTT pays attention to stimulating students' interest and motivation from beginning to end. Students are set in a learning or problem situation *via* cognitive conflict, which sets the students' minds a puzzle that is interesting and attackable, to arouse their maximum interests. The conflict is at least partially resolved as students' minds go beyond their previous thinking capability, and it is an effective way to make students think actively. Fourthly, LTT sets up a kind of open, democratic and positive activity atmosphere, and encourages students to spend more time discussing problems with partners, thinking independently, speaking out their own ideas bravely, and judging others' views, rather than blurting out answers without thinking just to get the teacher's attention. In such a context, students are not afraid to fail to answer questions any more, and they will be much more positive and active in obtaining knowledge, applying knowledge and solving problems. Some research has shown that students who used the approach of classroom discussion (Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009) and participated co-operatively in classrooms (Ladd & Dinella, 2009) displayed higher academic achievements. This feature is closely related to what Adey and Shayer (1994) describe as social construction. Fifthly, LTT trains students to realize and control their own thinking processes, that is, meta-cognition which is in the sense of conscious reflection on the problem-solving process. Students are encouraged to reflect and summarize how the successful methods or strategies they used helped them solve the problems, and what they have got out of the activity. Sixthly, LTT teaches students conscious transfer of a thinking method from a context in which it is first encountered to a new context, and then the transfer is most likely to be effective if the thinking method has previously been made explicit, conscious and verbalized (Chen & Klahr, 1999; Klahr & Nigam, 2004; Toth, Klahr, & Chen, 2000; Zohar & Peled, 2008). Therefore, this curriculum helps the students to apply and transfer the thinking methods that have been learnt in the LTT curriculum to other subject areas or to daily life, and to form the habits of effective thinking, thereby improving academic achievement. In summary, our teaching does not tell students the thinking processes directly, but requires students to exert his/her wisdom to realize the requirement of the thinking method, in order eventually to be able to achieve the level of explaining the thinking processes to others. This kind of teaching really mobilizes students' enthusiasm to think, makes them actively solve problems that they encounter, and trains them to transfer the thinking methods they have learned consciously and effectively.

It took longer for LTT to improve Grade 1 and Grade 2 students' academic achievement than it did for Grade 3 students (see Tables 8–10, Figure 4–6). The probable reason is that, on the one hand, the 3rd or 4th grade is the critical period of the development of students' thinking ability, and the development of thinking ability can improve the academic achievement effectively, but there is a delayed effect; on the other hand, the LTT curriculum of Grade 1 and Grade 2 is more related to students' daily life, whereas the Grade 3 curriculum is more related to subject knowledge. This result also indicates that the training of thinking method must be combined with subject knowledge in order to improve academic achievement more effectively.

Effects of LTT on students of different initial academic achievement levels

When the differential effect of LTT on students of different initial ability was tested, it was found that a significant effect on academic achievement (as measured by scores on Chinese and mathematics) only occurred with the medium-level students of Grade 1 and Grade 2, and only with the middle- and high-ability level groups in Grade 3. A very similar pattern of effects is shown with measures of thinking ability. The consistency between gains made by experimental groups over controls in academic achievements and thinking ability reinforces the idea that better thinking leads to higher achievement.

It is clear that the LTT curriculum, in this implementation, has failed to have a significant impact on the initially lower ability students. Notwithstanding our intention to place the demand levels of the activities with the zones of proximal development of all of the students, results suggest that with classes of 30 the teacher/researchers, in their first experience of using the LTT, found it difficult to adapt the demand levels within each lesson to meet the needs of a wide range of students. Unsurprisingly, they appear to have been 'teaching to the middle'. Lower ability students probably have reduced meta-cognitive abilities (Slife, Weiss, & Bell, 1985) and so would need more concrete scaffolding to get on to the 'thinking ladder' and so be more likely to use rehearsal strategy, that is, learn by rote, which is influenced by thinking weakly.

These results are in accordance with the previous finding (Onatsu-Arvilommi, Nurmi, & Aunola, 2002) suggesting that the use of maladaptive achievement strategies hampered children's subsequent improvement in reading and mathematical skills. Also, low-score students have a lower ability to transfer and their motivations are usually external (Butkowsky & Willows, 1980; Carr, Borkowski, & Maxwell, 1991), so it is difficult for them to realize how they are thinking consciously and to transfer the correct thinking ability to other context. Thirdly, teachers' beliefs that higher order thinking was inappropriate for low-score students (Zohar, Degani, & Vaaknin, 2001) may influence students' achievements. Furthermore, in the 4-year intervention, we also explored the effects of LTT on primary school students' learning strategies and learning motivation, which were closely related with achievements, and the results showed that it had an unapparent effect on low-score students too. It is just a shortcoming that our curriculum has a low effect on the low-score students, so the curriculum and or its delivery by teachers needs some attention.

Conclusion

Notwithstanding the failure to have impact on lower achieving students, we have shown overall that an intervention grounded in psychological theory and represented by a multi-faceted theoretical model can have a consistent, long-term, and growing effect on primary students' general thinking ability and on their academic achievement as assessed by standard school tests. Results presented here add to a growing body of evidence that specific attention to the development of children's thinking, even of a very modest intensity, can have far-reaching and cost-effective positive effects on their learning. After further attention to the lower ability students and to a professional development programme to make the materials and methods more generally accessible, we suggest that the LTT curriculum will prove to be valuable across China and further afield.

Acknowledgement

This research is supported by the MOE Project of Key Research Institute of Humanities and Social Sciences of University and the Program for New Century Excellent Talents in University.

References

Adey, P., & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement*. London: Routledge.

Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Butkowsky, I. S., & Willows, D. M. (1980). Cognitive–motivational characteristics of children varying in reading ability: Evidence for learned helplessness in poor readers. *Journal of Educational Psychology*, *72*, 408–422.

Buzan, T. (1984). *Use your perfect memory*. N.Y.: E. P. Dutton.

Carr, M., Borkowski, J. G., & Maxwell, S. E. (1991). Motivational components of underachievement. *Developmental Psychology*, *27*, 108–118.

Casey, M. B., Pezaris, E., & Nuttal, R. L. (1992). Spatial ability as a predictor of math achievement: The importance of sex and handedness patterns. *Neuropsychologia*, *30*(1), 35–45.

Chen, Z., & Klahr, D. (1999). All other thing being equal: Children's acquisition of the control of variables strategy. *Child Development*, *70*, 1098–1120.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Coles, M. J., & Robinson, W. D. (Eds.) (1989). *Teaching thinking. A survey of programmes in education*. Bristol: Bristol Press.

De Bono, E. (1970). *Lateral thinking—A textbook of creativity*. London: Ward Lock Educational Limited.

De Bono, E. (1987). *CoRT thinking program: Workcards and teachers, notes*. Chicago: Science Research Associates.

Dewey, J. (1998). *How we think*. New York: Houghton Mifflin Co.

Dewey, J., & Bento, J. (2009). Activating children's thinking skills (ACTS): The effects of an infusion approach to teaching thinking in primary schools. *British Journal of Educational Psychology*, *79*, 329–351.

Epstein, H. T. (1986). Stages in human brain development. *Developmental Brain Research*, *30*, 114–119.

Evans, R. (1973). *Jean Piaget: The man and his ideas*. N.Y.: E. P. Dutton.

Feuerstein, R., Rand, Y., Hoffman, M., & Miller, M. (1980). *Instrumental enrichment: An intervention programme for cognitive modifiability*. Baltimore, MD: University Park Press.

Hautamäki, J., Kuusela, J., & Wikström, J. (2002). CASE and CAME in Finland: "The second wave". Paper presented at 10th International Conference on thinking. Harrogate.

Greenberg, K. (1989). The Cognitive Enrichment (COGNET) funded in part by the U.S. Department of Education Follow Through Program. The University of Tennessee Follow Through Sponsor Project, Katherine Greenberg, Ph.D., Director.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. *Psychological Science*, *15*, 661–667.

Kuhn, D. (2005). *Education for thinking*. Cambridge: Harvard University Press.

Kuhn, J.-T., & Holling, H. (2009). Gender, reasoning ability, and scholastic achievement: A multilevel mediation analysis. *Learning and Individual Differences*, *19*, 229–233.

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade? *Journal of Educational Psychology*. *101*(1), 190–206.

Lin, C. (1999). *Learning and Development*. Beijing: Beijing Normal University Press.

Lin, C., & Li, T. (2003). Multiple intelligence and the structure of thinking. *Theory and Psychology*, *13*(6), 829–845.

Lin, C., Wang, Y., & Ye, Z. (1993). *Psychology for primary school students*. Hangzhou: Zhejiang Education Press.

Lipman, M. (1985). Thinking skills fostered by philosophy for children. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills: Volume 1: Related instruction to research*. London: Lawrence Erlbaum Associates.

Lipman, M., Sharp, M., & Oscanyan, F. (1980). *Philosophy in the classroom*. (2nd ed.). Philadelphia: Temple University Press.

Marjoribanks, K. (1976). School attitudes, cognitive ability, and academic achievement. *Journal of Educational Psychology*, *68*(6), 653–660.

Marzano, R. J., Brandt, R. S., Hughes, C. S., Jones, B. F., Preseissen, B. Z., Rankin, S. C., *et al.* (1988). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

McGuinness, C. (1999). *From thinking skills to thinking classrooms: A review and evaluation of approaches for developing pupils' thinking*. Norwich: HMSO.

McGuinness, C. (2000). ACTS: A methodology for enhancing thinking skills across-the-curriculum. *Teaching Thinking*, *1*(1).

McPeck, J. (1994). *Critical thinking and the 'Trivial Pursuit' theory of knowledge*. In Kerry S. Walters (Ed.), *Rethinking reason: New perspectives in critical thinking*. Albany, NY: SUNY Press, 101–117.

Murphy, K. P., Wilkinson, Ian, A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology*, *101*(3), 740–764.

Nickerson, R. S., Perkins, D. N., & Smith, E. E. (1985). *The teaching of thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nisbet, J., & McGuiness, C. (1990). Teaching thinking: The European scene. *Teaching Thinking and Problem Solving*, *12*(3), 12–14.

Onatsu-Arvilommi, T., Nurmi, J.-E., & Aunola, K. (2002). The development of achievement strategies and academic skills during the first year of primary school. *Learning and Instruction*, *12*, 509–527.

Perkins, D. N. (1995). *Outsmarting IQ: The emerging science of learnable intelligence*. New York: Free Press.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, *18*(1), 16–25.

Ruggiero, V. R. (1995). *The art of thinking: A guide to critical and creative thought*. 4th ed. New York: Harper Collins College Publishers.

Russell, D. (1943). Critical thinking in childhood and youth. *The School*, *31*(9), 744–750.

Russell, D. (1960). *Critical thinking. Encyclopedia of educational research*. New York: The Macmillan Co.

Slife, B. D., Weiss, J., & Bell, T. (1985). Separability of metacognition and cognition: Problem solving in learning disabled and regular students. *Journal of Educational Psychology*, *77*(4), 437–445.

Sloutsky, V. M., *et al*. (2001). How much does a shared name make things similar? Linguistic labels, similarity and the development of inductive inference. *Child Development*, *72*, 1695–1710.

Sternberg, R. J. (1996a). *Teaching for thinking*. Washington: American Psychological Association.

Sternberg, R. J. (1996b). *Successful intelligence*. New York: Simon & Schuster.

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research articles: A simplified methodology. Retrieved from http://www.work-learning.com/effect_sizes.htm.

Topping, K. J., & Trickey, S. (2007a). Collaborative philosophical enquiry for school children: Cognitive effects at 10–12 years. *British Journal of Educational Psychology*, 77(2), 271–278.

Topping, K. J., & Trickey, S. (2007b). Collaborative philosophical enquiry for school children: Cognitive gains at 2-year follow-up. *British Journal of Educational Psychology*, 77(4), 787–796

Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction*, *18*(4), 423–459.

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

Williams, F. E. (1972). *A total creativity program for individualizing and humanizing the learning process*. Englewood Cliffs, NJ: Educational Technology Publications.

Zohar, A., Degani, A., & Vaaknin, E. (2001). Teachers' beliefs about low-achieving students and higher order thinking. *Teaching and Teacher Education*, *17*, 469–485.

Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students. *Learning and Instruction*, *18*, 337–353.

Received 16 December 2009; revised version received 28 May 2010

Appendix A

Details of a lesson are given as follows.
Content of activity: shopping at the supermarket (one of activities in Grade 1) Objectives of activity: (1) Methods objective: train the classification ability of students—classify objects/things at the supermarket according to different criteria. (2) Ability objective: cultivate students' ability to observe things carefully—classify things from different angles basing on comparison and observation. (3) Emotion objectives: culture students to love life, care about life and be willing to collaborate with peers.
Important points of activity: learn to classify objects/things by different criteria.er
Difficult points of activity: use as many classification methods as you can, meanwhile call in question others' views, and thus develop students' critical thinking ability.
Preparation of activity: a certain number of daily supplies, multimedia courseware

Procedures of activity:

Step 1: Set up a learning situation via cognitive conflict to arouse children's maximum interests.

Kiddy, I have a small question; are you willing to help me? (Ss: Yes.) Now, we want to divide all of us into two groups; how will you classify it? Think quickly. (T: male and female by gender, adult and children by age, teacher and student by identity, and so on. *Guide the students to speak freely*.)

Design intent: Establish a learning situation to arouse children's interest in thinking, then some daily life questions that children pay little attention to can excite their cognitive interest, so as to stir up their enthusiasm of keeping on participating the activity.

Step 2: Activity process, involving facilitating children to observe, think, discuss, and conduct experiment.

The first activity:

Wonderful. Now, let's get down to today's activity. Meimei is a lovely girl; today, her mother buys a lot of things at the supermarket, then mother let Meimei act as a cashier to classify these things. If you are Meimei, how will you do? Look at the large screen. Are you familiar to these goods? (Ss: Yes.)

Well, let's have a competition in groups. Before the classification, there is something required: first, choose one recorder from your group quickly whose duty is to write down the classification; second, collaborate with each other; third, keep secret before allowing opening your group's result; fourth, lowering your voice when discussion. *(Group discussion begins, and the teacher guide between the groups.)*

Do you have finished? Let's together share your achievements. Each group sends a representative to speak first, and then others add, including how many categories your group classifies, according to what criterion, and what objects belong to the same category. Who is the first?

Great, each group has done very well. Think further about how you classify, or why you sort in this way.

Design intent: In the form of cooperation within group and competition between groups, students are expected to know there are standards for classification by exchanging, and deeply feel the great power of cooperation.

The second activity:

After the interaction, I feel that you all have yourselves own views on classification. Next we will do an interesting activity. Let's classify these dolls together. Look at the large screen. This time classify it by yourself, and can't discuss with your classmates. Let's see who can complete it first, now begin.

When students utter their results, teacher should remind them to present their views in complete statement and make them know the real connotation of classification.

Design intent: consolidate the 3 main points of classification by the second activity; expect students can summarize the main point after completing the activity.

Step 3: Evaluation and reflection.

Well, let's have a break for two minutes. Close your eyes. Have a rest while think, from the beginning of the class to the present, what is your profound experience? Or what you have learned? Or what you would like to share with others?

(Ss: I find that when classify the goods, we should classify all the goods, and can't lose one of them; I have learnt that classifying should be according to a certain standard; what I want to share with you is that when we classify things, there may be not only one method, but also others. . . .)

What everyone has said is very good, but each said just one point. Who can summarize them, and tell us when classify, what should we pay attention to?

Design intent: ask students to recall and reflect what they have learnt; summarize the emphasis and main points, in order to cultivate students' metacognitive ability.

Step 4: Consolidation transfer (activity broadening).

Well, let's sort out your own bags. Who does quickly and well?

Design intent: Expect students can successfully apply the classification methods learnt in class to daily life by sorting out their own bags.

Appendix B

Part of the thinking-ability test

Comparing and classification

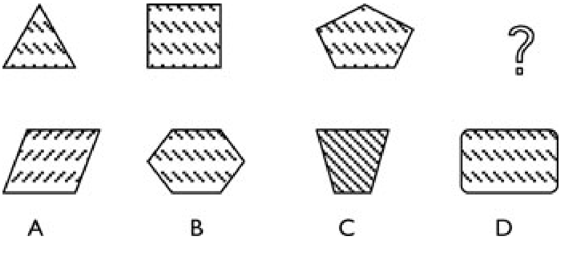
(1) Which one of the following four choices do you think is different from others? ( )

(2) Which one of the following four choices do you think is different from others? ( )
(A) bird (B) sparrow (c) kite (D) swallow

Inductive reasoning

(3) Observe the changes in the following figures, then choose the best answer: ( )
2, 9, 16, 23, ___
(A) 24 (B) 25 (C) 30 (D) 31

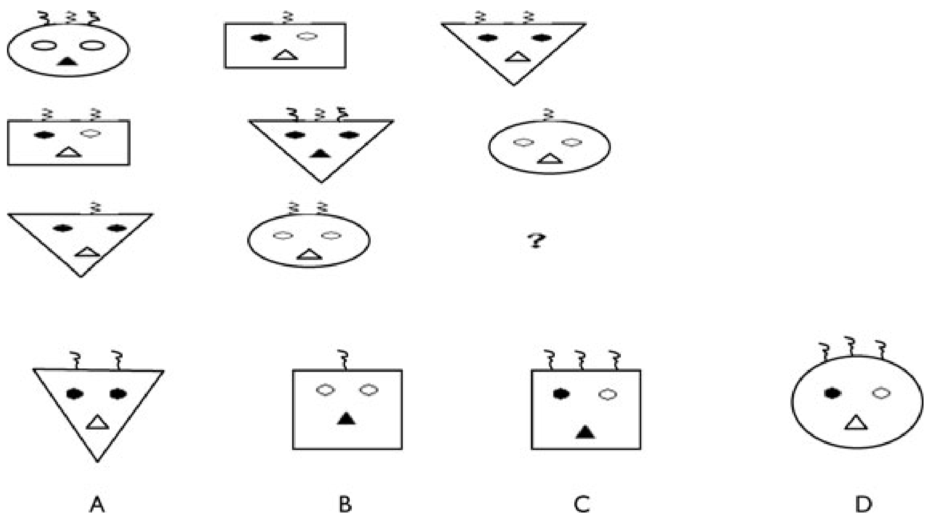
(4) Observe the changes in the following figures, then choose the best answer: ( )



Deductive inference

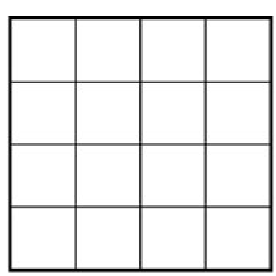
(5) Lucy is older than Lily and Mike is older than Lucy, so we can conclude that ( )
(A) Lucy is older than Mike. (B) Mike is older than Lily.
(C) Lily is older than Mike. (D) Lily is as old as Mike.

(6) Observed the changes in the following figures, then choose the best answer: ( )



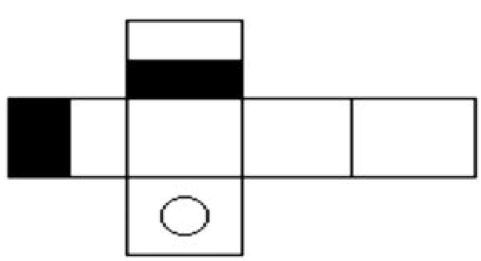
Spatial cognition

(7) How many squares are there in the figure? ( )

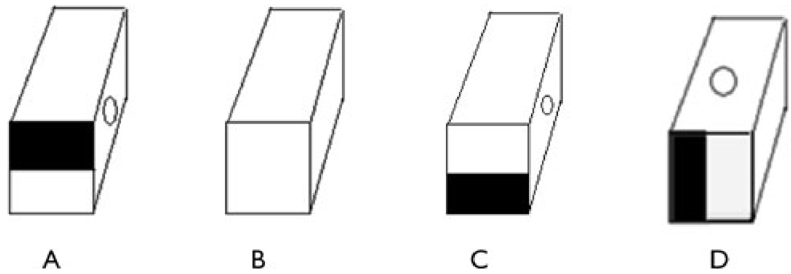


A 30 B 17 C 26 D 34

(8)



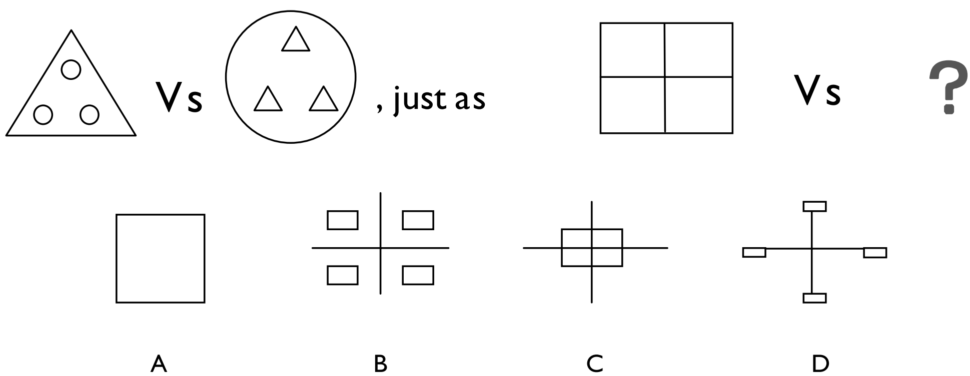
If you fold the figure above into a cube, it probably is ( )



Analogical reasoning

(9) Car vs. land, just as airplane vs. ( )
(A) wing (B) sky (C) land (D) astronaut

(10) Observed the following figures, then choose the best answer: ( )



Abstract-generalization

(11) Which one of the following four choices do you think is different from others? ( )
(A) running (B) do sports (C) playing basketball (D) playing football

(12) Meat, mushroom, bread, and apple belong to ( )
(A) vegetables (B) fruits (C) food (D) carbohydrate