

Instructional Design and Practice of Data Analysis Course for the Cultivation of Computational Thinking Ability

Xiulin Ma, Jingjing Liu^{*}, Jing Liang

Faculty of Education, Beijing Normal University, Beijing, China

Email address:

maxl@bnu.edu.cn (Xiulin Ma), 1069321995@qq.com (Jingjing Liu)

^{*}Corresponding author

To cite this article:

Xiulin Ma, Jingjing Liu, Jing Liang. Instructional Design and Practice of Data Analysis Course for the Cultivation of Computational Thinking Ability. *Education Journal*. Vol. 8, No. 6, 2019, pp. 249-258. doi: 10.11648/j.edu.20190806.13

Received: August 23, 2019; **Accepted:** September 6, 2019; **Published:** September 20, 2019

Abstract: Due to the rapid development of computer, network and data processing, computational thinking (CT) ability has played an important element for talent cultivation in the era of big data since it came out. As computer foundation course, Data Analysis Courses has a certain degree of fit with computational thinking training. In order to cultivate computational thinking, our team tried to integrate computational thinking into computer basic courses. To begin with, this paper analyzes the convergence of Data Analysis Courses and computational thinking, and puts forward the instructional design of Statistical Product and Service Solutions (SPSS) data analysis course from the perspective of computational thinking. Then based on the problem-based learning, the instructional design is carried out from three levels: comprehension and understanding, simple application and comprehensive application. Finally, aiming at improving students' CT ability, the teaching activities are designed progressively in computational thinking knowledge, CT skills, CT consciousness and CT strategy. Based on the above work, we carried out the Data Analysis course for the development of computational thinking. After a semester of practice, the results demonstrates that the teaching model proposed in this article enhances the students' CT ability significantly, and Data Analysis Courses is of great significance to the cultivation of students' CT.

Keywords: Computational Thinking, Data Analysis Courses, Teaching Model, Problem-based Learning

1. Introduction

1.1. The Importance of CT

The rapid development of computer, network and data processing has promoted the emerging of CT. The important role it plays in talent training has been widely recognized worldwide.

In America, CT has gained support from many institutions, not only the panel discussion at Carnegie Mellon University, but also the American Computer Society (ACM), the American Institute of Mathematics (AIM), the National Association of Computer Science and Technology Teachers (CSTA) and many other groups. In 2011, the International Society for Technology in Education (ISTE) and CSTA jointly proposed the operational definition of CT [1]. On the other hand, Li Lian, Chen Guoliang, and Dong Rongsheng devoted themselves to the promotion of CT in China [2, 3]. Following the Joint Statement of the Nine School Alliance (C9)

Computer Basic Teaching Development Strategy in 2010, which pointed out the core position of CT training in computer basic teaching [4], the Department of Higher Education of the Ministry of Education launched the "University Computer Curriculum Reform Project" and started 22 projects in 2012. Besides, the first International Conference on Computational Thinking Education (CTE2017) was held at the Hong Kong University of Education, educators and researchers from the world shared their practical research experiences in different contexts for CT education [5].

1.2. The Importance of Data Analysis Ability in the Cultivation of CT

In this digital era, data is where value lies. How to extract the pattern from the data, provide references for decision making, or implement intelligent recommendations has become a popular research areas in social science. Since data analysis has penetrated into the natural sciences and humanities and social

sciences as a methodological attribute, there have been some interdisciplinary research methods, which leads to a huge change in research method of related disciplines, and even fundamentally alters its original paradigm.

The methods and strategies of data processing and analysis are considerable part of CT, which are of great significance to the improvement of learners' research capability. It is essential for cultivating talent to train CT. In the connotation of CT, "computation" based on computer and network technology is its core. The calculation method in which it plays a guiding role is the soul of CT, while the related technologies of data analysis and data mining play the role of skeleton and backbone.

The urgent need for data analysis in the computer science has also spawned software such as SPSS and Statistics Analysis System (SAS). Algorithms such as correlation analysis, differential significance test, attribution analysis, have become the foundation in humanities, the threshold for data analysis is gradually decreasing. Researchers in humanities such as education, economics, psychology, and sociology are expected to achieve professional-level quantitative analysis with the help of software.

1.3. Research Questions and Objectives

Based on the requirements for talent development in this era, our team has set up a data analysis course ("Social Science Statistical Analysis Software and Application") for undergraduate since 2013.

Firstly, from the perspective of CT training, the research analyzes the characteristics of data analysis courses, explores the convergence of course content and CT, and provides an effective teaching model of SPSS-oriented course. Secondly, designs the evaluation index system of data analysis course aiming at CT cultivation. Finally, combines with questionnaires and comprehensive practical assignments to measure the differences in students' CT before and after teaching, to demonstrate the effectiveness of this research in the cultivation of CT.

2. Theoretical Background

2.1. The Concept and Nature of CT

The concept of CT that is widely used internationally, was published and defined by Jeannette M. Wing, Director of the Department of Computer Science at CMU University, USA. Jeannette M. Wing proposed in 2006 that "Computational thinking involves solving problems, designing systems, and

understanding human behavior, by drawing on the concepts fundamental to computer science. Computational thinking includes a range of mental tools that reflect the breadth of the field of computer science" and pointed out three key parts of computational thinking, which are algorithms, abstractions and automation [6]. In 2010, Jeannette M. Wing gave a formal definition of computational thinking: Computational Thinking is the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent [7]. In the operational definition of CT proposed by ISTE and CSTA and other groups, the CT is a problem-solving process that includes (but is not limited to) the following characteristics [8]: (1) Formulating problems in a way that enables us to use a computer and other tools to help solve them. (2) Logically organizing and analyzing data. (3) Representing data through abstractions such as models and simulations. (4) Automating solutions through algorithmic thinking (a series of ordered steps). (5) Identifying, analyzing, and implementing possible solutions with the goal of achieving the most efficient and effective combination of steps and resources. (6) Generalizing and transferring this problem solving process to a wide variety of problems. These six elements were summarized as problem decomposition, algorithms, abstractions, and automation by Yadav, A., Hong, H., & Stephenson [9].

Jeannette M. Wing clearly pointed out that the most important and high-level thought process in computational thinking is the abstraction process. It is used to capture essential properties common to a set of objects while hiding irrelevant distinctions among them. It is the process of deciding which links need to be emphasized, which links can be omitted, and the basis of computational thinking [10].

2.2. Dimensions of CT

Jeannette M. Wing has already elaborated on the definition, connotation and extension of CT. Most of researchers agree with her point of view and directly quote this statement. Focusing on the construction of the CT framework, Professor Chen Guoliang and Professor Dong Rongsheng, well-known experts in computational thinking, jointly proposed "The System of Expressions in Computational Thinking". Based on the "Great Principles of Computing" [11], it constructs a hierarchical CT framework with computation as the core, abstraction, automation, design, communication, collaboration, memory, and evaluation as the basic concepts, as shown in Table 1.

Table 1. Framework of CT system.

Computational Thinking System Framework		
Items	Central concern	Key elements
Computation	What can be counted and what can't.	The complexity and efficiency of the synthesis problem, evolution, sorted by space and time; computational representation, transformation of expression, state and its transitions; Computability, computational complexity theory.
Abstraction	Focus on the essential characteristics of the object	Conceptual model and formal model, abstraction level; Reduction, embedding, transformation, decomposition, data structures (such as queues, stacks, tables and graphs, etc.), virtual machines.
Automation	Discover algorithms for information processing	From algorithms to physical computing systems, from human thinking to artificial intelligence algorithms; formalization (definition, theorem and proof), procedures, algorithms, iterations,

Computational Thinking System Framework		
Items	Central concern	Key elements
Programming	Build reliability and credibility	recursion, search, reasoning; strong artificial intelligence, weak artificial intelligence
Communication	Reliable information movement between different locations	Consistency and completeness, reuse, security, compromise and conclusion; modularity, information hiding, class, structure, aggregation
coordination	Effective use of multiple autonomous computers	Information and its representation, Shannon's theorem, information compression, information encryption, verification and error correction, coding and decoding
Recollection	Store and retrieve information	Synchronization, concurrency, dead-lock, arbitration; event processing, flow and sharing dependencies, collaborative strategies and mechanisms; network protocols, human-computer interaction, swarm intelligence
evaluation	Responsible for performance prediction of systems (including natural and artificial systems such as earthquakes, nuclear weapons)	Binding; storage architecture, dynamic binding (names, handles, addresses, locations), naming (hierarchy, tree), retrieval (name and content retrieval, inverted index); locality and caching, trashing jitter, data mining, Recommended system
		Visual model and simulation, data analysis, statistics, computational experiments; model methods, simulation methods, benchmark; prediction and evaluation, service network model; load, throughput, reaction time, bottleneck, capacity planning

2.3. Views on CT Training

2.3.1. Views on the Content of CT Education

There are three main viewpoints on the content of CT education. The first view considers that CT content refers to the elements, core concepts and abilities contained in the operational definition of computational thinking. The second view mainly regards abstraction as the content of CT education. The rest of the views constitute a third perspective, such as the classification of calculations by the "great principles of computing" and the CT education content (computational concepts, practices) used in the programming environment. Among them, the research carried out around the first viewpoint is the most popular [12].

2.3.2. Teaching for Computational Thinking

Computational thinking teaching activities are usually problem-oriented, that is, the PBL-based teaching model. Problem-based learning is a student-centered instructional approach. PBL students discuss professionally relevant problems in small groups. The problems are first discussed before any preparation or self-study has taken place to fully understand the problem, questions formulated for further individual self-study by the students in the group. After this individual self-study period, students gather again and discuss

what they have learned and come to an answer to the formulated learning issues [13]. Based on problem-solving teaching, the teaching activities are organized by means of asking questions, analyzing problems, and solving problems, and guiding the process of thinking development. Therefore, most studies have identified CT content as six elements (submit problems, organize and analyze problems, characterize data, automate solutions, analyze and implement solutions, transfer knowledge) or nine core concepts and capabilities (data collection, data analysis, data characterization, problem decomposition, abstraction, algorithms and procedures, automation, simulation, parallelization), and combined with subject knowledge [14].

3. Instructional Design from the Perspective of Computational Thinking

3.1. Curriculum Knowledge System

3.1.1. Teaching Content

According to the syllabus of the data analysis course of the Computer Basic Course of Beijing Normal University, its content is shown in Figure 1.

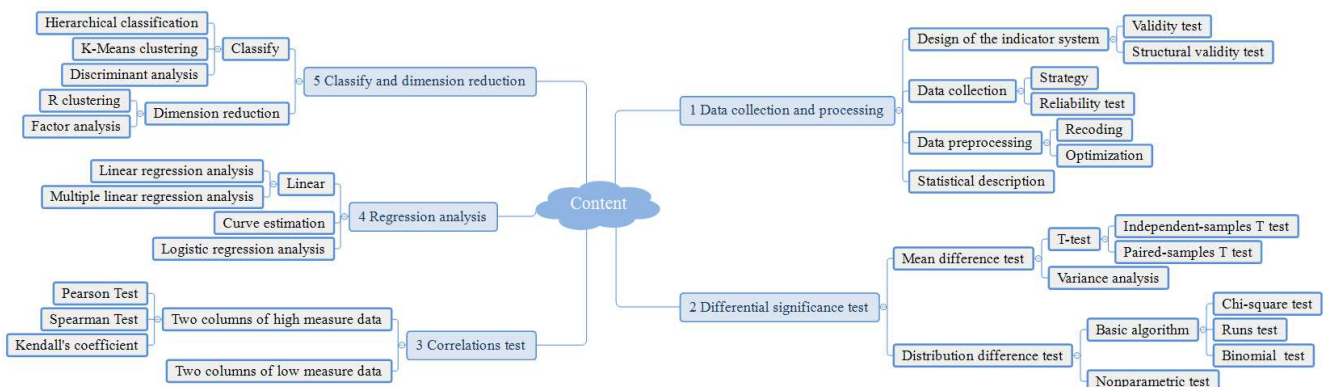


Figure 1. Data Statistics Analysis and Application Course Content Outline.

3.1.2. Integration of Data Analysis Courses and CT

"Data analysis" is mainly the application of mathematical statistics theory and methods to solve practical problems. It introduces the basic theories and methods of data analysis, and explains the analytical methods in detail such as linear regression, ANNOVA, attribution analysis, discriminant analysis, and cluster analysis in data analysis. Apart from this, it pays more attention to the whole process of problem solving, and refines valuable information and knowledge from it [15].

As an applied and practical course, it explores the inherent regularity of data by defining requirements, collecting, processing and analyzing data, and then interpreting results to achieve scientific understanding of phenomena. In university, the data analysis courses equips students' data analysis skills and provide technical support for empirical research. Applying empirical research often involves analysis problems, problem decomposition, data collection, data analysis, data representation, and even abstraction, algorithms and procedures, which is the main component of CT education content.

3.1.3. Match Point of Data Analysis Course Content and CT Content

In the data analysis course, students need to solve real problems with computer technology and computing strategies from the following four aspects: (1) Computational awareness; (2) Computational algorithms; (3) Abstraction process from concrete problem to computational problem; (4) Interpretation of data analysis results (reflect the conclusion from data).

Therefore, based on the "Great Principles of Computing", the data analysis support students' CT ability in five dimensions including computation, abstraction, automation, design and evaluation. Among them, the calculation algorithm is embodied in logical thinking ability and computing ability. The transformation from specific problem to computational problem is embodied in abstraction and design thinking. As a process from abstract to concrete, from data to practical problems, the interpretation of results is a data-based evaluation.

3.2. Teaching Activities Based on PBL Concept and Its Process Design

CT teaching activities are usually problem-oriented, which means the problem-based teaching model are suitable for this.

3.2.1. Four Dimensions of CT Training

This study divides the content of CT into four dimensions, which are constitute of CT knowledge, CT skills, CT awareness, and problem solving skills. CT knowledge mainly refers to the basic concepts of computer science, including theoretical knowledge of computing disciplines, computers and related devices. The phenomenon of "high score but low ability" indicates that knowledge cannot represent thinking, and the purpose of training on CT skills is to enable learners solve problems, design system and understand human behavior more effectively with basic concepts of computer

science. In the cultivation of computational thinking ability, it is required to cultivate disposition to master skills. It's expected to establish the basic consciousness of "computation". Also it is believed that computing technology can enhance people's "ability". The usage of mechanized methods to solve problems, such as abstract description, discrete, mechanical executable, has its unique advantages. As a comprehensive skill, the formation and mastery of CT strategy is demanding for learners' knowledge reserve and knowledge application.

However, these four dimensions interact with each other. The acquisition of knowledge and skills is the prerequisite for problem-solving strategy. The formation of CT consciousness is accompanied by the theory and practice of knowledge obtain, skill acquisition and problem solving. As shown in figure 2.

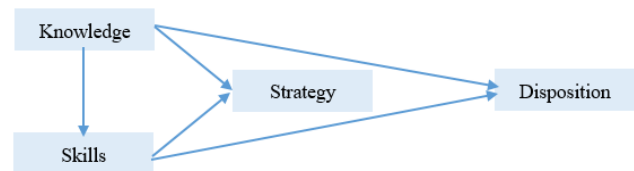


Figure 2. Interaction of four dimensions on CT.

3.2.2. "Three-step" Strategy for CT Training

In addition to theories, it is more important to grasp the analysis methods to solve specific problems. According to our experience, we set up two lessons for theoretical study and two lessons for practical operation. During the theoretical class, students are arranged to learn from teacher and discuss with group members. The group discussion is facilitated by a teacher (i.e., so-called tutor) and is aimed to acquire knowledge, to better understand the problem, and to acquire skills to solve the problem [16]. Then they are requested to complete the practical application and operation of the knowledge and principles learned in practical phase, which lasts for two lessons about 90 minutes. At the end of each unit, there will be an assignment expected to solve by group. Therefore, to develop CT in the data analysis course is basically implemented in accordance with the "three-step" strategy, corresponding to classroom teaching, practice with SPSS and group collaboration activities, respectively for comprehension and understanding, simple application and comprehensive application. In terms of the content, there are four dimensions, namely CT knowledge, CT skills, CT consciousness and problem solving ability.

More specifically, the teaching process of the data analysis course can be divided into three stages. In the first step, the teacher mainly demonstrates and imparts relevant knowledge and skills about data analysis. By means of classroom discussion, they can consolidate the knowledge they have learned more accurately. In the second session, students are provided with problem situations that promote the transfer of knowledge and skills. Comparing actual operations with skill points, under the guidance of "cognitive coaches" (teachers), the operation methods skills are strengthened and transformed into logical

skills. It is worth mentioning that the requirements should be designed deliberately to so that make the problems-solved explicitly. In the third step, the comprehensive problem solving should reflect the thinking mode and process of the subject and the decomposition and synthesis related to CT, abstraction and decomposition of the problem, problem conversion, recursive thinking back, heuristic reasoning and other ideas. In this way, the students will be proficient in adopting different methods for different types of problems, and form a comprehensive problem-solved strategy in the end.

3.2.3. The Cultivation Model of the CT

According to the main content and methods of the CT cultivation, and the four dimensions of CT, the cultivation

model of the CT ability based on problem solving is designed. Firstly, focusing on the four dimensions of CT, with the guidance of the cognitive tutor (teacher), the data analysis class shapes the students' disposition of CT by means of questioning and organizing group discussions. Secondly, through the penetration of teaching method, such as problem-based learning, inquiry-based learning, students are demanded to complete the teacher's arrangement of the homework, and achieve a simple application of the method, so as to develop CT skills. Thirdly, through organizing group collaborative learning, teachers use a variety of data strategies to achieve a comprehensive application of data analysis methods. The specific process is shown in Figure 3.

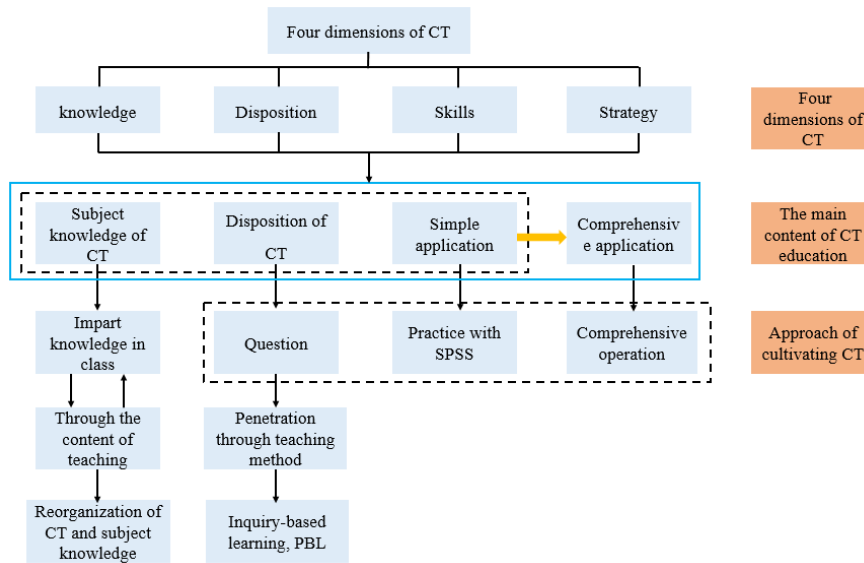


Figure 3. Design of teaching activities based on PBL.

Based on the above process, the corresponding teaching activities are shown in Figure 4.

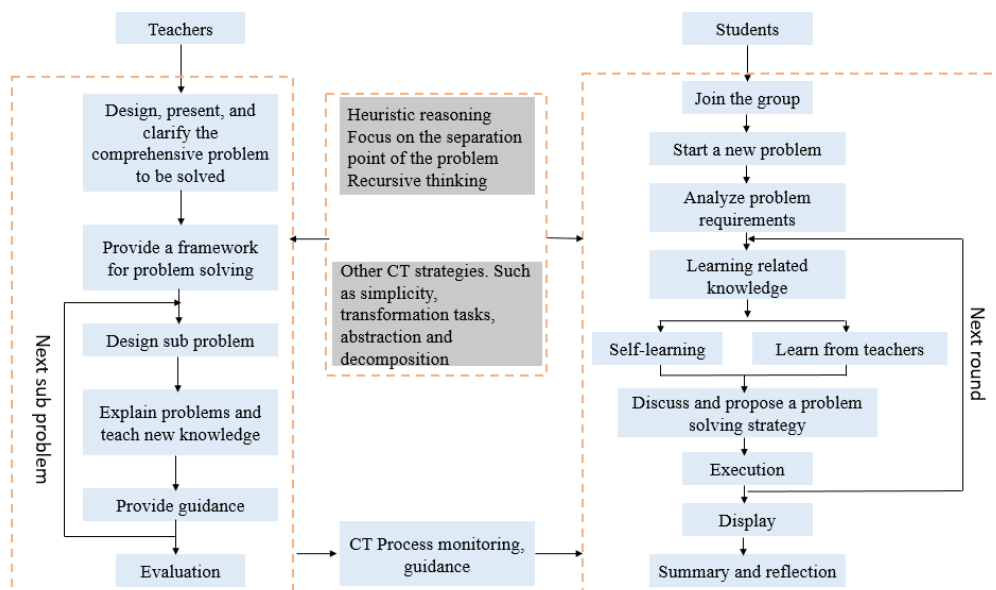


Figure 4. Activities of teacher and students in the model.

3.3. Evaluation

3.3.1. Questionnaire

In this part, a questionnaire was designed to test students' CT ability from five dimensions. To make sure it's reliability and validity, 28 preliminary samples were tested.

i. Structure of the Questionnaire

In order to evaluate the improvement of students' CT ability, we designed a questionnaire including a total of 26 questions from five aspects including CT disposition (C1), CT knowledge (C2), data abstraction and modeling strategy (C3), data analysis knowledge and skills (C4), and the ability of problem-solved using data analysis (C5). In the form of a five-level scale, the degree of affirmation from left to right is sequentially increased according to the degree of recognition of the subject's semantics.

In a self-reported manner, students are required to truthfully fill in the questions that reflect their cognitive and strategic understanding of CT. And then we analyze the differences in students' CT ability between pre- and post-teaching.

ii. The Reliability and Validity of the Questionnaire

In order to ensure the scientific and effective measurement, the research first analyzed the reliability and structural validity of the questionnaire questions. 28 preliminary samples were obtained after the initial survey to test the reliability and validity of the questionnaire.

Firstly, the result of Factor Analysis for the preliminary survey data demonstrates that $KMO=0.881$, which means the variables are suitable for Factor Analysis. Then run factor analysis and check the structural validity of the output.

After running the Principal Component Analysis, the

variables a7, a8, a9, a10, a11 belong to the first principal component, and the variables a2, a3, a4, a5, a6 belong to the second one, a17, a18, a19, a20, a21 belong to the third component., a12, a13, a14 belong to the fourth, a23, a25, a26 belong to the fifth. According to the description of each variable in the original data, it is found that the results are basically consistent with the preset survey index system, which means it has good structural validity.

Based on the preliminary survey data, the consistency analysis of the variables in the survey indicators is carried out, and the results of the final reliability analysis are shown Table 2.

Table 2. Questionnaire reliability statistics.

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.901	.902	23

There are 23 items in Table 2, which means 23 topics representing parameter reliability analysis. The Cronbach's $\alpha=0.902$ (larger than 0.8), indicating that the reliability of the data is good. Therefore, the scales covered in this questionnaire are effective and credible in the measurement of CT and data analysis capabilities.

3.3.2. Design of Comprehensive Evaluation

In addition to evaluating the program and the statistical analysis process, the evaluation mainly focus on the usage of CT strategies (decomposition and synthesis, etc.) and data analysis thinking in the process of comprehensive problem solving. As shown in Table 3.

Table 3. Evaluation form of Comprehensive tasks.

Type of thinking	Procedure	G1	G2	G3	G4	G5	G6
Process of CT	Whether it is possible to solve research problems in the practice of following the "principle-method-construction", decomposition and synthesis thinking processes.						
Data collection	Collecting data according to research purposes and designing survey plan						
	Process	Qualitative-quantitative-qualitative					
Data analysis	Analytical thinking	Hypothesis-verification					
	Logical thinking	Judging the distribution pattern-calculating the test probability-inferring					
Evaluation	Reliability (confidence interval, significance)						
	Validity analysis						

4. Teaching Practice

This study lasted for one semester and the teaching content includes five modules as shown in Figure 1. Limited to the length of the paper, this section takes the "differential significance test" module as an example to organize through the "three-step" strategy to enhance the four dimensions of CT.

4.1. Preparation for Teaching

4.1.1. Objectives

This teaching practice activity is carried out around the "Differential Significance Test" module in the course "SPSS

Data Analysis Theory and Practice". Based on the "Great Principle of computing" proposed by Denning, this chapter attempts to support students' CT ability in five dimensions: computation, abstraction, automation, design, and evaluation. Meanwhile, developing computational thinking skills as one of the core objectives of this course requires adhering to the "three-step" strategy mentioned above.

4.1.2. Participants and Course Organization

LiYun class (2016), admitted to Beijing Normal University in 2016, was selected as an experimental class to carry out teaching practice. Constitutes 48 students, including 9 boys and 39 girls, it was divided into 6 groups according to the principle of

heterogeneous grouping. At the same time, consisting of 46 students, LiYun class (2015) was chosen to be control class.

The experimental class adopts the “2+2” mode, which arranges 2 lessons for classroom lectures and discussions, and 2 lessons for practicing skills. The assignments mainly include SPSS operations and large social reports that require an analytical report. Students are required to choose their own themes, design a questionnaire, and complete data analysis. In the teaching process, the teaching activities are organized according to the norms of thinking training and the instructional design oriented to the development of CT, and the students' computational thinking ability is taken as the

main training goal.

4.2. Implementation of Teaching Activities

4.2.1. The First Step: Knowledge Transfer and Understanding

At this stage, the relevant knowledge and skills of differential significant analysis are mainly imparted by teacher. Through classroom discussion, students will master how to test the significance of differences. Take "Independent Samples T Test" as an example, the teaching process is shown in Table 4.

Table 4. Teaching process of “Independent Samples T Test”.

Teaching process	Contents						
Explanation	<ol style="list-style-type: none"> i. Both variables are independent sample. ii. Both satisfy the normal distribution. 1-tailed or 2-tailed test? <ol style="list-style-type: none"> a. Check if the distribution is the same, choose 2-tailed test. b. Check if the difference, choose 1-tailed test. iii. Identification of test probability <ol style="list-style-type: none"> a. 95% or 99% confidence interval of the difference. b. Result: <ol style="list-style-type: none"> 1. Sig<0.05 (0.01), indicates significant differences. 2. Sig>0.05 (0.01), indicates that there is no significant difference. 						
Present the problem to be solved	<p>Case: The Chinese scores of a class are shown in Table 1. Please check whether there are significant differences in Chinese scores of different genders.</p> <p>Differences test about Chinese scores should be performed between boys and girls. For the t-test, the following problems should be mainly solved:</p>						
Analysis	<ol style="list-style-type: none"> 1. ‘Chinese’ are in a normal distribution. 2. Grades belonging to different genders are independent samples. 3. Choose the ‘independent samples t-test’ <p>Note: Is the variance homogeneity affecting the analytical methods and results?</p>						
Solution	<table border="0" style="width: 100%;"> <tr> <td style="width: 33%;">Based on SPSS</td> <td style="width: 33%;">Based on functions in Excel</td> <td style="width: 33%;">Based on the analysis tool library in Excel</td> </tr> <tr> <td> (1) Test whether the ‘Chinese1’ and ‘Chinese2’ meet the normal distribution (One-Sample K-S Test) (2) Run ‘Independent-Samples T Test’: a. Analyze-Compare Means-Independent-Samples T Test b. Test Variable (s): Chinese1 c. Grouping Variable: sex d. Set ‘Confidence Interval Percentage’: 95% 1. Click ‘OK’ (3) Interpret the Output 2. Check the ‘sig.’ in Levene’s Test for Equality of Variances e. If the ‘sig.’ > 0.05, then choose ‘sig. (2-tailed)’ in the first row of the t-test for Equality of Means. f. If the ‘sig.’ < 0.05, then choose ‘sig. (2-tailed)’ in the second row. </td> <td> (1) Execute function: 1. In one free cell, type “FTEST (array of ‘Chinese1’, array of ‘Chinese2’)” 2. If the FTEST value>0.05 a. Type “TTEST (array of ‘Chinese1’, array of ‘Chinese2’, 2, 2)” 3. else b. Type “TTEST (array of ‘Chinese1’, array of ‘Chinese2’, 2, 3)” (2) Interpret the Output 4. Verify the value of ‘p’ 5. Compare with means </td> <td> (1) Execute ‘Independent-Samples T Test’: 1. In one free cell, type “FTEST (array of ‘Chinese1’, array of ‘Chinese2’)”: 2. If the FTEST value>0.05 a. Data-data analysis-T Test: Two-sample equal variance hypothesis b. Else 3. Data-data analysis-T Test: Hypothesis of non-homogeneous variance: c. Array 1: array of ‘Chinese1’ d. Array 2: array of ‘Chinese2’ e. Click ‘OK’ (2) Interpret the output 4. Verify the value of ‘p’ </td> </tr> </table>	Based on SPSS	Based on functions in Excel	Based on the analysis tool library in Excel	(1) Test whether the ‘Chinese1’ and ‘Chinese2’ meet the normal distribution (One-Sample K-S Test) (2) Run ‘Independent-Samples T Test’: a. Analyze-Compare Means-Independent-Samples T Test b. Test Variable (s): Chinese1 c. Grouping Variable: sex d. Set ‘Confidence Interval Percentage’: 95% 1. Click ‘OK’ (3) Interpret the Output 2. Check the ‘sig.’ in Levene’s Test for Equality of Variances e. If the ‘sig.’ > 0.05, then choose ‘sig. (2-tailed)’ in the first row of the t-test for Equality of Means. f. If the ‘sig.’ < 0.05, then choose ‘sig. (2-tailed)’ in the second row.	(1) Execute function: 1. In one free cell, type “FTEST (array of ‘Chinese1’, array of ‘Chinese2’)” 2. If the FTEST value>0.05 a. Type “TTEST (array of ‘Chinese1’, array of ‘Chinese2’, 2, 2)” 3. else b. Type “TTEST (array of ‘Chinese1’, array of ‘Chinese2’, 2, 3)” (2) Interpret the Output 4. Verify the value of ‘p’ 5. Compare with means	(1) Execute ‘Independent-Samples T Test’: 1. In one free cell, type “FTEST (array of ‘Chinese1’, array of ‘Chinese2’)”: 2. If the FTEST value>0.05 a. Data-data analysis-T Test: Two-sample equal variance hypothesis b. Else 3. Data-data analysis-T Test: Hypothesis of non-homogeneous variance: c. Array 1: array of ‘Chinese1’ d. Array 2: array of ‘Chinese2’ e. Click ‘OK’ (2) Interpret the output 4. Verify the value of ‘p’
Based on SPSS	Based on functions in Excel	Based on the analysis tool library in Excel					
(1) Test whether the ‘Chinese1’ and ‘Chinese2’ meet the normal distribution (One-Sample K-S Test) (2) Run ‘Independent-Samples T Test’: a. Analyze-Compare Means-Independent-Samples T Test b. Test Variable (s): Chinese1 c. Grouping Variable: sex d. Set ‘Confidence Interval Percentage’: 95% 1. Click ‘OK’ (3) Interpret the Output 2. Check the ‘sig.’ in Levene’s Test for Equality of Variances e. If the ‘sig.’ > 0.05, then choose ‘sig. (2-tailed)’ in the first row of the t-test for Equality of Means. f. If the ‘sig.’ < 0.05, then choose ‘sig. (2-tailed)’ in the second row.	(1) Execute function: 1. In one free cell, type “FTEST (array of ‘Chinese1’, array of ‘Chinese2’)” 2. If the FTEST value>0.05 a. Type “TTEST (array of ‘Chinese1’, array of ‘Chinese2’, 2, 2)” 3. else b. Type “TTEST (array of ‘Chinese1’, array of ‘Chinese2’, 2, 3)” (2) Interpret the Output 4. Verify the value of ‘p’ 5. Compare with means	(1) Execute ‘Independent-Samples T Test’: 1. In one free cell, type “FTEST (array of ‘Chinese1’, array of ‘Chinese2’)”: 2. If the FTEST value>0.05 a. Data-data analysis-T Test: Two-sample equal variance hypothesis b. Else 3. Data-data analysis-T Test: Hypothesis of non-homogeneous variance: c. Array 1: array of ‘Chinese1’ d. Array 2: array of ‘Chinese2’ e. Click ‘OK’ (2) Interpret the output 4. Verify the value of ‘p’					
Display operation	Demonstration by teacher						
Discussion	Share the case of the wrong operation and discuss the similarities and differences between different solutions						

4.2.2. The Second Step: Simple Problem Solving

Under the guidance of teachers, students are equipped with strategy about problem solving and the method of “Differential Significance Test”, which promotes the conversion of operational skills into logical skills. During this

phase, students need to write out the problem-solving steps and the reasons for the selected actions in the document to make the thinking process explicit. An example of the problem need to solve is shown in Figure 5.

Experiment 2 — Differential significance test

Please download the pending data files mydataA.sav, mydataB.sav and mydataC.sav, then complete the following operations and briefly explain the operation methods, procedures and results you have used in the ans.docx documentation. For statistical inference questions, please explain your conclusions.

1. recode the string variables in mydataA and mydataB into numeric variables, and name them in English. Then add the description information to the new variables and their values in variable view.
2. Test the distribution of scores in each subject in mydataA, and the distribution of IQ and enrollment scores in mydataB.
3. In mydataA, test the difference between variable 'Chinese1' and 'Chinese2', 'Chinese2' and 'history', 'mathematic' and 'physic', and 'physic' and 'chemistry'.
4. In mydataA, check whether the students' major has an impact on the variable 'Chinese1' and 'history'.
5. In mydataA, check whether the students' major and sex has an impact on the variable 'physic' and 'chemistry'.
6. In mydataA, check whether there is a significant difference between 'Chinese 1', 'Chinese 2', and 'history', and check whether there are significant differences in 'mathematic', 'physic', and 'chemistry'.

Figure 5. Experiment of 'Differential Significance Test'.

4.2.3. The Third Step: Comprehensive Application

The solution of the comprehensive problem is infiltrated throughout the course. At the beginning of, the teacher presents the topic and requirements of the problem, and points out the key points of problem solving, so that the students can clear their

goals. On the one hand, it can be used as a guide for student problem selection and problem solving, and on the other hand, it can ensure that students have the right ideas about the problem. The following is a comprehensive operation of the social survey, the specific content is shown in Figure 6.

Group Activity 1 — Campus survey and data analysis

Select research topics from the management, study, life, and entertainment facilities of the school and design a questionnaire. At least 40 participants should be selected for the investigation.

Requirements:

1. Able to obtain scientific conclusions on the research questions through questionnaires and data analysis
2. The questionnaire is designed to be reasonable and contains check items, which can represent survey questions.
3. The respondents were no less than 40 people. The questionnaire covered different levels of research questions and was well represented.
4. In data analysis, there should be at least a difference test (parametric or non-parametric), attribution (correlate and regression), and cluster analysis.
5. Complete a research report that meets the specifications of academic papers.

Note:

This task is expected to be reported by the group in **mid-December**. Please arrange the time reasonably.

Figure 6. Example of comprehensive operation.

5. Data Analysis and Discussion

5.1. Analysis and Conclusion Based on Questionnaire

5.1.1. Comparison Between the Experimental Class and the Control Class

In order to demonstrate the teaching effect, the author

conducted a post-test on the LiYun class (2016), and conducted a computational thinking test on the LiYun (2015) class with the questionnaire designed in 3.3.1. An 'Independent Samples T Test' was performed on the post-test data collected for the experimental and control classes. The analysis results are shown in Table 5.

Table 5. Result of 'Independent Samples T Test'.

Items	Experimental class (Mean)	Control class (Mean)	Significance	
			t	Sig. (2-tailed)
C1	4.38	3.81	3.56	0.000
C2	4.17	3.41	4.32	0.000
C3	4.32	3.83	3.57	0.000
C4	4.22	4.13	1.97	0.051
C5	4.51	4.12	2.01	0.049
C	4.31	4.01	2.32	0.041

As can be seen from Table 5, there are significant differences between the experimental and control classes in four dimensions (C1, C2, C3, and C5), although there is no statistically significant difference in data analysis knowledge and skills (C4). But its value of test probability is also close to 0.05 (significant standard).

Table 6. Result of 'Paired Samples T Test' on the experimental class.

	Paired differences				T	df	Sig.(2-tailed)		
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
				Lower				Upper	
Pair 1	QC1 - HC1	-.489	.608	.113	-0.721	-.258	-4.334	28	.000
Pair 2	QC2 - HC2	-1.027	1.079	.200	-1.438	-.617	-5.127	28	.000
Pair 3	QC3 - HC3	-1.137	.893	.166	-1.478	-.798	-6.863	28	.000
Pair 4	QC4 - HC4	-.4761	.710	.1318	-.7459	-.2058	-3.610	28	.001
Pair 5	QC5 - HC5	-.725	.868	.162	-1.055	-.395	-4.499	28	.000
Pair 6	QC - HC	-.770	.566	.105	-.985	-.555	-7.331	28	.000

Note: Q: pre-test; H: post-test; C: a comprehensive evaluation of the evaluation dimension. For example, QC1: pre-test data for question 1; HC1: post-test data for question 1.

It can be seen from Table 6 that the mean of the paired samples is negative, the significance of two-tailed is less than 0.05, and $t < 0$. Therefore, there is a significant difference between the pre-test and post-test values, and the post-test is significantly higher than the pre-measurement in all dimensions. To sum up, after teaching practice, not only the knowledge, skills and ability of data analysis to solve problems are significantly improved, but also the computational thinking thoughts such as decomposition and synthesis, abstraction and modeling have been significantly improved.

5.1.3. Differential Analysis by Gender

'Independent Samples T Test' is performed on the pre- and

As for the cultivation of CT and abstraction ability, it is of great value to instill CT strategy and method in teaching practice, organize class following the paradigm of thinking training, and emphasize the importance of CT in big data era.

5.1.2. Data Analysis of Experimental Class

For the experimental class, the author conducted pre-test and post-test respectively using the questionnaire, and obtained 48 sets of valid data. The results of 'One-Sample Kolmogorov-Smirnov Test' shows that the pre- and post-test data of the samples were normally distributed. Therefore, choose the 'Paired Samples T Test' to test the corresponding data of the experimental class, and the analysis results are shown in Table 6.

post-test data by sex, and found that they did not reach the significant level (Sig values were 0.138 and 0.179, respectively). Therefore, no significant difference can be seen between boys and girls in the CT ability.

5.2. Analysis and Conclusion Based on Comprehensive Evaluation

As all groups meet the requirements in terms of completion, in the aspect of comprehensive task evaluation, it mainly analyzes the process of students' thinking when they complete. In the table of the comprehensive operation, "1" represents that there happens the above thinking process, and "0" represents doesn't. The analysis results of the six groups are shown in Table 7.

Table 7. Result of 'Comprehensive tasks evaluation'.

Type of thinking	Procedure	G1	G2	G3	G4	G5	G6
Process of CT	Whether it is possible to solve research problems in the practice of following the "principle-method-construction", decomposition and synthesis thinking processes.	0	1	1	1	1	1
Data collection	Collecting data according to research purposes and designing survey plan	0	1	1	1	1	1
Data analysis	Process Qualitative-quantitative-qualitative	1	1	1	1	1	1
	Analytical thinking Hypothesis-verification	1	1	1	0	1	1
Evaluation	Logical thinking Judging the distribution pattern-calculating the test probability-infering	1	0	1	1	1	1
	Reliability (confidence interval. significance)	0	1	1	1	0	1
	Validity analysis	0	1	1	1	0	1

The result shows that in the comprehensive problem-solving research report, most students have mastered the ability to solve problems by using statistical methods of statistical analysis and complex problem solving, decomposition and synthesis. It indicates that the goal of

improving the ability of CT are met.

5.3. Conclusion of Data Analysis

The above data show that in the teaching of data analysis courses, consciously infiltrating students' knowledge, theory and

strategy of CT is conducive to the formation of their CT consciousness and ability. The survey data of the experimental class and the control class shows that in the data analysis course, whether the teacher organizes the teaching according to the process of thinking teaching has a significant influence on the data analysis course. In addition, consciously teaching or presenting CT knowledge and ideas in class also exerts the same effect.

In fact, for data analysis courses, follow the rules of computational thinking teaching, consciously strengthen students' thinking patterns, and encourage students to think about problems from the perspective of abstraction and modeling, which will not only affect the quality of teaching, but will improve the quality of learning.

6. Conclusion

The development of science and technology puts forward new requirements for the cultivation of CT of college students. However, relevant teaching practices have confirmed that the CT courses separated from specific courses and situations are not conducive to the development of students' CT ability and the formation of CT consciousness.

Therefore, it is possible to set up a CT training course among college students only when adopt an effective teaching model. From the current Chinese university computer course system, it is feasible and effective to integrate the computational thinking ability into the university computer public class by means of the university computer-based class.

This study evaluates the teaching effect through the instructional design of the SPSS data analysis course, supplemented by quantitative data, which demonstrates that this mode has a significant effect on the cultivation of CT ability. It provides a new way to develop CT skills in computer basic courses in the end. Cultivating CT in computer-based classes is not a one-time process. Taking the data analysis courses, the program design course, and even the multi-media course as an effective carrier for CT, will certainly be positive for the benign development of college students' CT.

Acknowledgements

We would like to thank all the teachers and students who have supported our teaching practice. It is your enthusiasm and encouragement that has always supported us. Thanks to Education Department of Beijing Normal University for providing good support and guarantee for our research. Thank all of you!

References

- [1] ISTE&CSTA (2011a). Operational Definition of Computational Thinking for K-12 Education [EB/OL]. [2017-07-01]. <http://www.iste.org/docs/ct-documents/computational-thinking-operational-definition-flyer.pdf?sfvrsn=2>.
- [2] Wu L. C., Li L. (2017), Computer Liberal Curriculum Construction under the Background of New Engineering. *China University Teaching*: 12, 62-69.
- [3] Chen G. L., Dong R. S. (2011), Computational Thinking and Computer Basic Education in University. *China University Teaching*: 01, 7-11+32.
- [4] Dong R. S. (2010), Joint Statement of the Nine School Alliance (C9) Computer Basic Teaching Development Strategy. *China University Teaching*: 10, 14-15.
- [5] Chen P., Huang R. H., Liang Y., and Zhang J. B. How to develop computational thinking-based on 2006-2016 research literature and the latest international conference papers. *Modern Distance Education Research*: 01, 98-112.
- [6] Wing, J. M. (2006). Computational thinking. *Communications of the ACM*: 49 (3). 33-35.
- [7] Cuny, J., Snyder, L., & Wing, J. M. (2010). Demystifying computational thinking for non-computer scientists. Unpublished manuscript in progress, referenced in <http://www.cs.cmu.edu/~CompThink/resources/TheLinkWing.pdf>.
- [8] ISTE&CSTA (2011a). Operational Definition of Computational Thinking for K-12 Education [EB/OL]. [2017-07-01]. <http://www.iste.org/docs/ct-documents/computational-thinking-operational-definition-flyer.pdf?sfvrsn=2>.
- [9] Yadav, A., Hong, H., and Stephenson, C. (2016). Computational thinking for all: pedagogical approaches to embedding 21st century problem solving in K-12 classroom. *Tech Trends*: 60 (6), 565-568.
- [10] Wing, J. M. (2010). Computational thinking: What and why?. Unpublished manuscript. Pittsburgh, PA: Computer Science Department, Carnegie Mellon University. Retrieved from <https://www.cs.cmu.edu/~CompThink/resources/TheLinkWing.pdf>.
- [11] Denning, P. J., and Martell, C. H. (2015). *Great Principles of Computing*. ACM.
- [12] Liu M. N., Zhang Q. W. (2018). Research Progress in Foreign Computing Thinking Education. *Open Education Research*: 24 (01), 41-53.
- [13] Loyens, Sofie M. M., Marcq, Helene, Gijbels, David, Dolmans, Diana H J M. (2016). Deep and surface learning in problem-based learning: A review of the literature. *Advances in Health Sciences Education*: 21 (5), 1087-1112.
- [14] Liu M. N., Zhang Q. W. (2018). Research Progress in Foreign Computing Thinking Education. *Open Education Research*: 24 (01), 41-53.
- [15] Zhang L., Wei L. F. (2015). Exploration of the "Data Analysis" Course in the Age of Big Data. *Education Teaching Forum*: 25, 154-155.
- [16] Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. In L. Wilkerson & W. H. Gijsselaers (Eds.), *New directions for teaching and learning* (Vol. 68, pp. 3-12). San Francisco: Jossey-Bass.