Applied Research on Big-data-based Analysis of Chinese Basic Education Integrating Social Values
—Taking Chinese Education from 3rd to 6th Grade as Example

Shuqin ZHAO
College of Marxism, Fujian Normal University
Fuzhou, Fujian 350100, China
and
Huaqi CHEN
Faculty of Software, Fujian Normal University
Fuzhou, Fujian 350100, China
and
Jang-ruey TZENG
Faculty of Software, Fujian Normal University
Fuzhou, Fujian 350100, China

ABSTRACT

Primary education is the gold stage of personal growth which is the foundation of the comprehensive development of moral, intelligence, physical, art and labor. Therefore, grasping the daily education of primary students is the focus of primary education. The application of big-data analysis could focus on the micro and overall performance of students. Maybe, these data have no significance to the individuals, but the information of all the students could solve many problems in the teaching process. Accordingly the teachers could acquire the real learning level of most of the students in school and more accurately carry out the personalized education and facilitate the efficient study of students.Using the SAS statistical models, this paper mined and analyzed the 3-6 grade's big testing data on Chinese Language courses through the VOSMaP Database of the EduCube project. In this study, the results of two-way ANOVA analysis model can provide the finding and effectively assist educators to solve the child's learning problems. In China the Chinese Language courses of Basic Education focus on the cultivation of loving the motherland language, cultural and social values; it emphasizes on the development of intellectual, morality and sound personality as well as the balanced development of moral, intellectual, physical, aesthetic, labor and so on. Based on the big data analysis, this paper concludes that the Chinese Language education fused with social values can facilitate the effectiveness of this kind of integration education in China, as well as can provide the educators of this field a new thinking in the big data era.

Keywords: Basic Education, Social Values, Data Analysis, Study Experience, Two Factors Analysis of Variance

1. INTRODUCTION

Data has penetrated each industry and business function field today and become an important production factor. The mining and application of massive data by people is predicting the arrival of new wave of productivity growth and consumer surplus. The arrival of big data era and the vast data resource enable every field to start the quantization process, no matter it is academic circle, business circle or government, all fields will start this process, quantizing everything becomes the core of datamation. Education today should avoid single form of the resources, and try the best to eliminate the phenomenon of "Information Isolated Island", although primary school education is the education of basic knowledge, it requires educators to have intelligent and scientific ideas in a great extent. If it is lack of advanced educational idea and teaching skills and only relying on the teachers' explanation of teaching materials in classes alone, its effect will be very limited.

A series of important information in the daily homework and examination answers of students are often neglected. By mining and analysis of the big data, the authors find the potential values hidden in such information, and use them to provide assistance for improving the studying of students. Big data analysis injects new vitality into education reform, and also provides technical guarantee and data supporting for the applied research of integrating social values in basic education.

2. LITERATURE REVIEW

Chinese Education and Information-based Social Values

Chinese language is the most important interpersonal communication tool and an important part of human culture. Chinese language course is a comprehensive and practical course that studies the application of language and words. Chinese language course in basic education stage is a media for students to preliminarily use the language and words of motherland to communicate, absorb excellent cultures in ancient and modern times and in China and foreign countries, as well as enhance thinking and cultural cultivation and promote spiritual growth. It contains rich social values. Value, as phenomena of

consciousness, is a reflection of social existence. As Marx and Engels said: "People's idea, viewpoint and concept, in a word, people's consciousness, will change along with the change of people's living condition, people's social relationship and people's social existence" [1] (1995)With the rapid economic development, people's living standard had prominent enhancement, material demands obtained great satisfaction and spiritual demands are also increasing. Mayer Schoenberg said in the book *Era of Big Data* said that knowing "What" is enough, there is no need to know "Why". In the era of big data, we don't have to know the reasons behind phenomenon, but let data "speak" for itself. [2] (2013)

Basic Chinese Language Education and Application of Big Data Analysis

Big data is also referred to as massive information, it means the information assets [3] (2015) with massive quantity, high growth rate and diversity and need new processing model to have stronger decision-making power, insight and process optimization ability. In the book of *Era of Big Data* written by Viktor Mayer Schoenberg and Kenneth Cukier: big data is mostly not random samples but whole data [2] (2013) for the processing of massive data in the era of big data, we must have fundamental change in thoughts and ideas; preferring entirety than part, preferring efficiency than accuracy and preferring relevance than causality. The basic process and steps of big data processing include acquire, organize, analyze and decide, which is as follows in Figure 2-1.
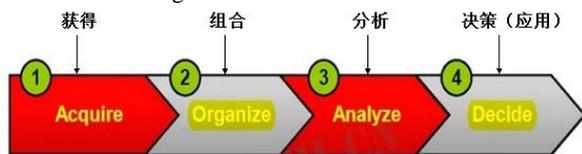
Figure 2-1 Cited from CSDN Blog Net – Flow Figure of Big Data Processing, modified by Chen Huaqi, 20150414

With the rapid development of big data field, promoting its application in education industry has become possible. It has gradually become a reality that schools have usable and high-quality massive data. However, the issue of what information to be mined and how to mine information proposes challenge to the work of education researchers. By mining the Traceable Learning Management based studying data of students, the authors can further improve the model and method of traditional teaching, so as to truly realize the "improvement of studying result and promotion of high-efficiency studying".

Characteristics of Modern Social Values Education

In the education of the social values, the traditional education mainly includes knowledge enhancement, model demonstration, character inspiration, environment influence, self-cultivation, self-guidance, etc., these methods are effective in the past and to great extent enhanced people's recognition and practice of the social values. [4] (2004) Values education must rely on knowledge, or else spiritual enhancement and outstanding

wisdom would have no place to start with. "Learn and Be Reasonable" means enhancement of knowledge quality is closely correlated with forming of advanced values. In the era of big data, information is in everywhere, how to efficiently obtain usable information, enhance studying efficiency, while training team quality, also promote people to have active changes in values, this must rely on big data analysis. Therefore, modern education of values in China must take full advantage of the modern information technology represented by internet.

Diversified Application of Big-Data Analysis System

The data of this research thesis is cited from the industry-study-research cooperation project between EduCube Global and Fujian Normal University: the VOSMaP database. It has over 100 thousand entries recording the Chinese subject testing question bank information and answers by students, the research value is very high. "Those have data will win it all". It can be predicted that under the wave of big data, all different industries are having in-depth reforms. In order to follow the step of era development, different fields on the related industrial chain all participated and hoped to occupy an unique position in the upcoming big data market, as shown in Figure 2-2
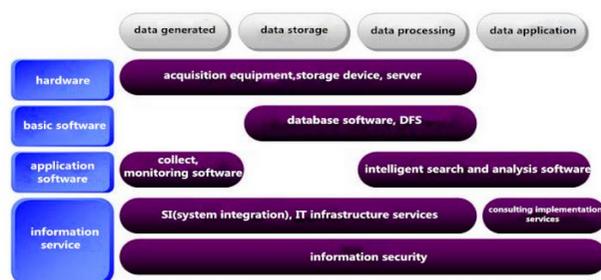

Figure 2-2 Cited from Everbright Securities – Schematic Diagram of Big Data Industrial Structure, 20140414

With the increasingly maturing big data technology, its market scale is also developing and expanding, which brings great promotional force to the related industries of big data and a series of investment opportunities. [5] (2014)

3. STRATEGIC THINKING OF CHINESE LANGUAGE EDUCATION INTEGRATING SOCIAL VALUES

Development Characteristics of Basic Chinese Language Education and Social Values

Chinese language course should be open and full of innovation and vitality. It should try its best to meet the demands of different regions, different schools and different students, and set the course objectives that are suitable for demands of the era. Students are the subjects of learning, Chinese language course must, based on the physical and mental development of students and the characteristics of Chinese language learning, fully inspire their consciousness of problems and entrepreneurial spirit, pay more attentions to individual difference and different learning demands. The construction of Chinese language course should inherit the excellent tradition of Chinese language

education in China; and pay close attentions to the demand of modern social development, and make the course evaluation method more diversified.

Feasible Model of Integrating Social Values into Basic Chinese Language Education Based on Big Data Analysis
In basic education, the rich humanistic connotation has wide and in-depth influence on the spiritual world of students, and the perception and understanding of Chinese materials by students are also diversified. Because Chinese course is a media of cultivation of thoughts and emotions for students, and the diversified model of Chinese course implementation also enriches the convenience and diversification of social values integrating into education, therefore, the integration of the two has certain feasibility.

The social values contents and the diversified education model of Chinese are used to define the value orientation of Chinese course, inherit and promote the excellent Chinese cultural tradition, intangibly transform and reflect the Chinese internalization result of the socialism core values system. This scope of social values could reflect the full process of the common ideal of the socialism with Chinese characteristics, the national spirit with patriotism as the core and the era spirit with reform and innovation as the core, excellent ideological and moral trend integrating the knowledge system and subject development of Chinese education. It will integrate the Chinese traditional national culture and traditional virtues into Chinese education. Through Chinese education practices, based on the "knowledge" and "faith", the social values are practiced in the practice of Chinese language education, so as to form the model of "acknowledgement" --- "practice" --- "new acknowledgement" --- "new practice".

In the preliminary stage of the values education, it mainly involves the regulation on certain basic concepts and categories, which are mainly about the value judgment and value orientation of certain issue. On the practical level, with the appearance of multiple interest subjects in the current society, people's values orientation is bound to reflect the characteristics of multi-dimension and multi-layers. Therefore, big data could, from various aspects including the information collection, processing and analysis of knowledge ontology, provide support and guarantee for the basic Chinese language education integrated from diversified values and knowledge.

Social values education is integrated in the basic Chinese language course, big data could be used to analyze the individual traits and class studying record of the students, as well as adjust the integrating-type knowledge education behaviors based on individualized demands. Big data could be sourced from online tests of the students in and out of school, through this kind of process network implementation, the full recording of studying behavior and knowledge cognition with individualization and real-time feature, it could entirely and massively collect the testing results of the group students on the

knowledge cognition studying, understanding and application of social values, then certain value could be reached through professional big data analysis tools, technology and model application as well as statistics analysis. Big data enables basic Chinese education integrating social values education to be significant; the establishment of complete data structure enables the testing and analysis value of knowledge education to reflect scientific nature and application effectiveness.

4. DESIGN AND REALIZATION OF DATA ANALYSIS APPLICATION MODEL

Through the analytical research of the VOSMaP database, the authors hope to find the root reason that affect the correct and incorrect ratio of question answering of students. The thesis starts from specific questions, researches the corresponding articles of incorrect answers, difficulty coefficient of questions, incorrect ratio and other data information, find the true understanding ratio of knowledge ontology (including social values) of students in them, and provide orientation for the guidance work of teaching and life values of students.

The research objective of this part is to explore that the current Chinese education knowledge contents could analyze the value guidance of teaching and learning, and hope to, from the analysis result, find the measurability and creditability of social values after integrating social values into Chinese language education.

Data Contents and Analytical Application
Integrating the entity relationship diagram of VOSMaP and the detailed design document of database, analysis is implemented to conclude the database diagram shown in Figure 4-1, including the six items, including the test answer sheet: t_testmain; question table: t_question; answer details table: t_test_detail; gradient table: t_gradient; course table: t_course; course type table: t_courseType.
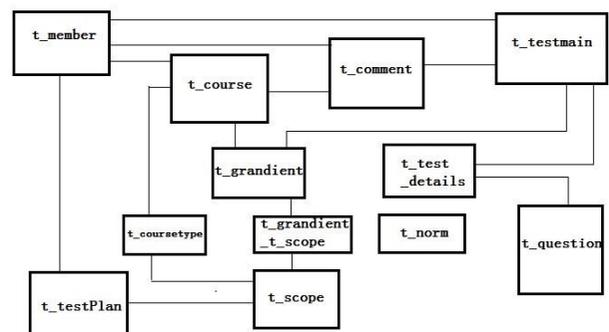

Figure 4-1 is cited from the core ER Figure of VOSMaP database source EduCube Project..

The "data set" collection needed for data analysis include:
Related Record Inquiry of the Testing Questions of Corresponding Grades: The grade researched in this thesis is

the Chinese subject testing data of the 3rd Grade to 6th Grade of the Attached Primary School of Fujian Normal University, as shown in Figure 4-2, it shows the inquiring process of the data base data through SQL inquire statement, and reaches the objective of data processing purification.

Figure 4-2 SQL inquire statement of Chinese subject of the 3rd Grade, formed by Chen Huaqi

Using Data to Gradually Establish the Inquire Result Data Set: The data base of the testing questions, knowledge points, question difficulty and question incorrect ratio records of corresponding grades and subjects is established by using SAS data set function, such as establishing the logic library connected with database; data is stored in the content corresponding to formed logic library; read the statements in the database, establish database and other statistical analysis to provide data basis.

PROC GLM（Procedure General Linear Model）Statistic Process and Two-factor Analysis of Variance(ANOVA)
Two factors analysis of variance is to make variance analysis on the observed sample, compare the systematic difference of the sample observation value under the influence of two factors, then analyze if each factor has significant influence on the result. This thesis uses the two factors A, B of the knowledge point and difficulty coefficient corresponding to the question as the independent variables to research if the influence of the two on the question incorrectness ratio or the collective influence of the two factors on the question incorrectness ratio is significant, so as to find the causes that affect the answering incorrectness ratio of students.

Two Factors Analysis of Variance in PROC GLM Statistic Process: Multivariate analysis of variance is made with the "Statistics - ANOVA - Factorial ANOVA" function module of SAS software. In order to facilitate the subsequent drawing of

the connection diagram of mean difference of different levels of each factor, this research adopts SAS statistic module "analyzer" for operation analysis, the data result operation procedure is as shown in Figure 4-3:
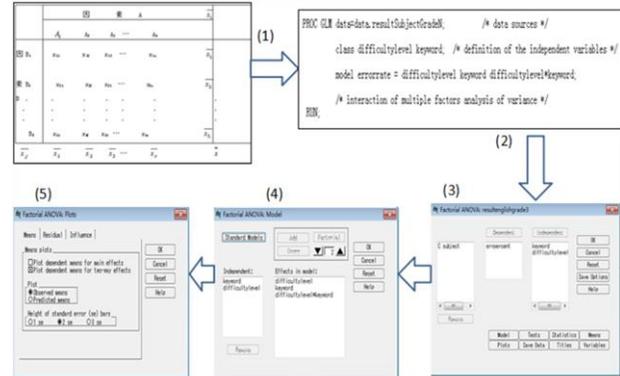
Fig. 4-3 ANOVA Data Result Process

1) Establishment of mathematical model of two factors analysis of variance;
2) Data analysis by using the SAS software's PROC GLM statistic process;
3) Setting of independent variables and dependent variables;
4) Setting of model effect column;
5) Setting of connection diagram of mean difference;

Analysis of Research Result and Application Value
The SAS output result includes the variance analysis report and mean difference connection diagram. In the variance analysis report, the data with statistical significance is [Pr > F] value, which means the value of P in statistic sense, P=0.05 is boundary condition of the acceptable relevance. The [Pr > F] value ＜0.05, it means the influence of such indicator on the dependent variable is significant; if [Pr > F] value ＜0.01, it means the influence of such indicator on the dependent variable is highly significant; if [Pr > F] ＜0.001, it means the whole model is highly significant, besides, the influence of difficulty coefficient and knowledge point on average question incorrectness ratio is significant, but the collective influence of the two is not significant. The horizontal axis corresponding to the three vertical lines of the dependent variable mean difference connection diagram means different difficulty coefficients, the value 1,3,5 from left to right respectively mean simple, middle and difficult. Connection lines with different colors correspond to different knowledge points, the vertical axis of the crossing point means the average incorrectness ratio of question under the corresponding knowledge point and difficulty, the Figure is also attached with the vertical bar with mean value as the center and two times of standard deviation as the length. Two times of standard deviation: in mathematics, it means the arithmetic square root of variance ×2, it could reflect the dispersion degree of data set, the longer the vertical bar, it means the bigger dispersion degree of the incorrectness ratio corresponding to the difficulty coefficient of knowledge point and question. Summarizing the variance analysis result of the Chinese language education and learning for the 3rd-6th Grade, the following conclusion and suggestion are made as shown in Table 4-1.

Table 4.1 Variance analysis of Chinese language education and learning of 3rd-6th Grade | General review of different grades

3rd Grade

The GLM Procedure
Dependent Variable: errorrate errorrate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	63	18.17778608	0.28853629	2.28	<.0001
Error	333	42.11293172	0.12646526		
Corrected Total	396	60.29071779			

R-Square	Coeff Var	Root MSE	errorrate Mean
0.301502	95.09562	0.355620	0.373960

Source	DF	Type III SS	Mean Square	F Value	Pr > F
difficultylevel	2	4.01506780	2.00753390	15.87	<.0001
keyword	22	5.60648816	0.25484037	2.02	0.0049
keyword*difficultyle	39	6.93301476	0.17774961	1.41	0.0609

3rd Grade
1. Knowledge point and difficulty coefficient both have significant influence on incorrectness ratio;
2. The incorrectness of simple questions is relatively low, but the incorrectness ratio of middle and difficult questions is relatively high.

4th Grade

The GLM Procedure
Dependent Variable: errorrate errorrate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	68	6.97155135	0.10252281	1.13	0.2440
Error	277	25.08670937	0.09056574		
Corrected Total	345	32.05826072			

R-Square	Coeff Var	Root MSE	errorrate Mean
0.217465	151.3262	0.300941	0.198869

Source	DF	Type III SS	Mean Square	F Value	Pr > F
difficultylevel	2	0.21667947	0.10833973	1.20	0.3039
keyword	23	1.56002909	0.06782735	0.75	0.7924
keyword*difficultyle	43	4.60460415	0.10708382	1.18	0.2140

4th Grade
1. The model reflects insignificant characteristic to the result;
2. Most texts are well understood, few texts are poorly understood, and pointed teaching could be implemented.

5th Grade

The GLM Procedure
Dependent Variable: errorrate errorrate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	77	18.93362704	0.24589126	2.67	<.0001
Error	296	27.24176989	0.09203301		
Corrected Total	373	46.17539693			

R-Square	Coeff Var	Root MSE	errorrate Mean
0.410037	169.4708	0.303369	0.179010

Source	DF	Type III SS	Mean Square	F Value	Pr > F
difficultylevel	2	0.04665276	0.02327638	0.25	0.7767
keyword	25	8.76491138	0.35059646	3.81	<.0001
keyword*difficultyle	50	7.02465366	0.14049307	1.53	0.0178

5th Grade
1. Difficulty coefficient didn't have significant influence on the incorrectness ratio of students;
2. The average incorrectness ratio is low, the incorrectness ratio of middle and difficult questions is low, which means students have relatively transparent understanding on the texts.

6th Grade

The GLM Procedure
Dependent Variable: errorrate errorrate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	58	10.32897997	0.17808586	1.89	0.0003
Error	305	28.75789000	0.09428816		
Corrected Total	363	39.08686997			

R-Square	Coeff Var	Root MSE	errorrate Mean
0.264257	213.5769	0.307064	0.143772

Source	DF	Type III SS	Mean Square	F Value	Pr > F
difficultylevel	2	0.39705322	0.19852661	2.11	0.1235
keyword	19	5.21689157	0.27457324	2.91	<.0001
keyword*difficultyle	37	2.86961449	0.07755715	0.82	0.7603

6th Grade
1. The difficulty didn't have significant influence on the incorrectness rate of students;
2. Most texts were received well, few need to be strengthened.

The incorrectness rate of question answering by the 3rd grade students was obviously influenced by the knowledge points and difficulty, the understanding of different texts had relatively big variety, the simple questions had relatively low incorrectness rate, but the middle and difficult questions had relatively high incorrectness rate, which proved that the understanding of text by the students was superficial and the understanding degree was not deep enough and they were also very easy to be affected by the question disturbance factors. The influence of difficulty on incorrectness rate of the 4th to 6th grade was not high, which proved that the learning ability of students was relatively strong, and they were able to have relatively deep studying and research on texts. It's just that they had different understanding degree of the text, therefore, the teaching degree pointed to different texts should be different. The knowledge of Chinese language education is also the case, the knowledge ontology corresponding to social values and combined learning result could also be used in this analytical method.

5. BIG DATA ANALYSIS APPLICATION OF THE CASE

Interconnection and Source of Basic Chinese Language and

Social Values Education

The *Standard of Chinese Language Course of Compulsory Education* formed by the Ministry of Education of the People's Republic of China indicated that the learning process of Chinese language shall cultivate the thoughts and morality of patriotism, collectivism and socialism and develop the individuality, cultivate the spirit of innovation and cooperation, form the active life attitude and correct values, as well as develop thinking ability, learn scientific thinking method, cultivate the scientific attitude of pragmatism and upholding of genuine knowledge, etc., and also stress the ability of collecting and processing information, use new technology and multi-media in learning and other course objectives. It clearly required using basic Chinese language education to integrate values education and establish scientific values as well as cultivate good teaching habits.

Zhang Yong, Chinese education measurement expert, "Using big data technology to implement scientific evaluation on students in the six functions of testing, evaluation, diagnosis, screening, selection and identification could analyze the knowledge application, skill application and ability inclination of each student, and could timely adjust educational behaviors, so as to truly realize individualized education." [6] (2015)

Feasibility of Chinese Education Data Analysis Model and Value Diversion Model

Under the teaching model of basic education, the assessment and evaluation of student learning and growth by teachers are mostly in accordance with average score of class, which has certain subjectivity and limitation. Relying on the big data analysis of many students of same grade, the cognition and internalization of students on the basic Chinese knowledge can be massively and commonly found. This thesis uses data analysis and is based on the big data norm, explores the incorrect answers of students in the knowledge points and different difficulty coefficients in the texts of different grades, teachers could more truly understand the understanding of different text knowledge points of students, so as to realize the individualized teaching and promote individualized education.

Implementing value-based knowledge combination utilization method on the foresaid analytical models, when the knowledge measurement questions are designed, as long as the authors change different traditional social values into knowledge presentation, and integrate them into observation of Chinese knowledge points, and design different difficulty coefficients for them, the cognition of students on different social values could be quantitatively understand or their internalization level could be measured. Through measurement of the same model, the authors could use the Traceable Learning Management to trace the learning result of students and the method to find the hidden problems, and propose the improvement methods. It could be still established into the big data analytical model that measures social values.

The Chinese course in the nine-year compulsory education stage is facing the entire students, integrating values contents in Chinese course design could not only enable students to receive the basic Chinese accomplishment, but also cultivate the social values of students. Integrating excellent cultural knowledge and the natural succession and influence of values through basic Chinese course will be able to promote the balanced development of students, and by using big data analysis, we could synchronize the Chinese education and social value education, enable students to receive attention and guidance in growth, and enable students to gradually form rich knowledge combination and complete character.

6. CONCLUSION AND SUGGESTION

By using big data analytical tools, the authors could undertake mathematical model prediction and structuring on massive data. When the mathematical model on a question need to be analyzed and established, the essence through the phenomenon of things, and transform actual question into the related application model shall be seen. After data analysis and analytical analysis of the results, the authors could decompose the essence of problem and find the relatively effective solution.

In the field of education, the prospect of big data application is very wide, but the path is still unclear. The current educators are mostly lack of the awareness of retaining data and collecting data. Let's imagine, if the tests of examining students' learning can be changed into data resource and implement Traceable Learning Management, the authors can compose individual knowledge into data set and then change into knowledge database, then implement knowledge map-style teaching and application, it will bring immeasurable value for the teaching in the future.

This kind of model that promote integrating social values education into basic education and generally inspecting teaching effect is worthy for the government agencies to think about and consider.

7. REFERENCES

[1] Marx Engels Selected Works [C]. Beijing: **People's Press**，1995 (1);291

[2] Viktor Mayer-Schonberger,Kenneth • Kuke Ye;Sheng Yang Yan,Zhou Tao（Translate）. Big Data Era. [M] Hangzhou：**Zhejiang People's Press**，2013.1；P67, P27

[3] Baidu Encyclopedia《Big Data》[EB]. http://baike.baidu.com/link?url=4L7lBax7OcAh,2015-04-15

[4] Tangzhi Long. Time Characteristics of Value Education [J]. **Learning Forum**, 2004, 20(9): 62-65

[5] She Huijie. Thinking big data era - Opportunities and challenges [EB]. http://www.docin.com 2014-01-21

[6] People. Experts：Big data will guide the individualized nature of education return [EB/OL]. http://liuxue.people.com.cn. 2015-03-25