


Spring 2005

Evaluation of 2+2 Alternative Teacher Performance Appraisal Program in Shanxi, People's Republic of China

Weiping Wang
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/urbanstudies_etds

 Part of the [Educational Administration and Supervision Commons](#), and the [Teacher Education and Professional Development Commons](#)

Recommended Citation

Wang, Weiping. "Evaluation of 2+2 Alternative Teacher Performance Appraisal Program in Shanxi, People's Republic of China" (2005). Doctor of Philosophy (PhD), dissertation, , Old Dominion University, DOI: 10.25777/b2xn-cc14
https://digitalcommons.odu.edu/urbanstudies_etds/1

This Dissertation is brought to you for free and open access by the School of Public Service at ODU Digital Commons. It has been accepted for inclusion in Theses and Dissertations in Urban Studies by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

**EVALUATION OF 2+2 ALTERNATIVE TEACHER PERFORMANCE
APPRAISAL PROGRAM IN SHANXI, PEOPLE'S REPUBLIC OF CHINA**

By

Weiping Wang

B.A. July 1985, Beijing Foreign Studies University

M.A. July 1990, Beijing Foreign Studies University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

URBAN STUDIES-EDUCATION CONCENTRATION

OLD DOMINION UNIVERSITY

May 2005

Approved by: /

Dwight W. Allen (Chair)

Maurice Berube (Member) /

Linda Bol (Member)

ABSTRACT

EVALUATION OF 2+2 ALTERNATIVE TEACHER PERFORMANCE APPRAISAL PROGRAM IN SHANXI, PEOPLE'S REPUBLIC OF CHINA

Weiping Wang
Old Dominion University, 2005
Chair: Dr. Dwight Allen

This dissertation examined the effectiveness of the 2+2 Alternative Teacher Performance Appraisal System that has been implemented as one of the educational reform efforts in Shanxi province in China. All 78 teachers in Grade One in the six schools participated in the study. The outcomes investigated in the dissertation included teachers' pedagogical beliefs and attitudes, the teachers' professional performance, teachers' collaboration, and the 2+2 feedback. A mixed research design incorporating both quantitative and qualitative approaches was used to evaluate the program. Data triangulation was achieved through varied data sources including focus group meetings, 2 + 2 feedback forms, observations, and questionnaires. Six high schools (13 teachers in each school) in Shanxi province were selected for the research. Three of the schools participated in the 2+2 program while another three served as the comparison. The researcher conducted focus groups with the teachers in the 2+2 program to explore their perceptions of the program.

The results showed that 2+2 program significantly improved teachers' professional performance, enhanced teachers' collaboration, and increased the feedback between the peers, though no similar impact was found on teacher's beliefs and attitudes. After exposure to the program, the teachers in the 2+2 group performed better in all of the nine in-class instructional performance functions that were measured by the

evaluators. Teachers felt overwhelmingly that 2+2 helped them with their professional growth. The quality of feedback was perceived as high quality by the teachers themselves and the researcher. The teachers in the 2+2 group had far more classroom observations than the comparison group. At pre-test, the teachers in the 2+2 group conducted observations on an average of 1.5 times per month, while at post-testing, they reported observing others teaching about nine times a month. The comparison group teachers observed one another only about twice a month. It is interesting to note that teachers produced more compliments than suggestions, which may indicate that unlike professional evaluators, teachers tend to be more willing to offer compliments rather than suggestions in the 2+2 setting. Over 60% of the respondents characterized 2+2 feedback as encouraging. The productive and constructive feedback made the collaboration in the 2+2 program more frequent and successful.

Copyright, 2005, by Weiping Wang, All Rights Reserved.

This dissertation is dedicated to my mother.

ACKNOWLEDGEMENTS

First of all, I would like to thank all the teachers of the six high schools in Shanxi Province who participated in the 2+2 program. They made tremendous contributions of their time, experience, and ideas to the program and this study. The entire time with them has been an exceptional journey. My appreciation goes to the administration of the schools as well, for recognizing the value of research and encouraging teacher participation and fully supporting the study.

Secondly, my gratitude to my dissertation committee members: Dr. Dwight Allen, Dr. Maurice Berube, and Dr. Linda Bol for continual support and patience with my research. To Dwight Allen, my advisor and chair of my committee, who created the 2+2 program and brought me to Old Dominion University. He has encouraged me continually to expand my dreams both in the States and in China. My sincere thanks go to Dr. Linda Bol, who devoted a great deal of her precious time and expertise to improve this dissertation. She was full of concrete ideas and always ready to help. Her enthusiastic and continuous support helped me evolve my own ideas and complete the final product. To Maurice Berube, whose wisdom and humor made the difficult dissertation process a not-so-scary journey.

I would like to extend my appreciation to my colleagues who helped greatly in the dissertation research. To Dr. Alyce LeBlanc, who shared her valuable experience on 2+2 program with me. To Dr. Zhaoxin Ma, my long-time friend, who helped me to administer surveys and organize interview meetings with the teachers. To Ms. Xueqin Chang, who played a great role as the program manager and made the program memorable.

Last but not least, I would also like to acknowledge the faculty, staff and students of Darden College of Education for creating a supportive environment. Particularly to Barbara Webb for always being positive, encouraging and making a doctoral program manageable.

TABLE OF CONTENTS

	Page
LIST OF TABLES	IX
CHAPTER I	1
INTRODUCTION.....	1
China's Educational Reform Efforts in Teacher Education	1
Context of the 2+2 Program in Shanxi Province	5
The Content of the 2+2 Program in Shanxi Province.....	7
Statement of the Problem.....	12
Purpose of the Study	15
Research Questions.....	16
Research Design	16
Significance of the Study.....	17
CHAPTER II.....	20
LITERATURE REVIEW	20
Peer Coaching.....	20
Teacher Evaluation	24
Teacher Collaboration	29
Teacher Isolation	34
Feedback	37
Teachers' Professional Development	42
Previous Evaluation of the 2+2 Program.....	46
CHAPTER III	50
METHODOLOGY	50
Research Plan.....	50
Research Questions.....	51
Research Design	52
Data Collection	56
<i>Setting</i>	56
<i>Sample</i>	56
<i>Instrumentation</i>	58
<i>Interview</i>	62
Procedure	63
<i>Implementation of the 2+2 Program</i>	63
<i>Administration of Measures</i>	64
Data Analysis.....	65
<i>Quantitative Data</i>	65
<i>Qualitative Data</i>	66
Role of the Researcher.....	69
CHAPTER IV.....	70
RESULTS	70

Program Impact on Professional Performance	70
<i>Overall program Impact of 2+2</i>	71
<i>Impact of 2+2 program component on the teachers in the 2+2 Program</i>	72
Program Impact on Teachers' Pedagogical Beliefs and Attitudes	79
Program Impact on the Frequency of Classroom Observation.....	80
Program Impact on the Frequency of Feedback	82
Program Impact on Teacher Collaboration.....	84
Program Impact on the Kind of Feedback Received	86
Teachers' Perceptions of the Advantages and Disadvantages of the 2+2 system	91
Suggestions Teachers Provided for Improving the 2+2 Program.....	102
Summary of the Findings.....	105
CHAPTER V	107
IMPLICATIONS AND CONCLUSIONS	107
Professional Development	107
Observation	110
Feedback	112
Teachers' Collaboration.....	113
Teachers' Pedagogical Beliefs and Attitudes	115
Recommendations for Future 2+2 Practice	116
Limitations and Recommendations to Further Research.....	117
<i>Limitations</i>	117
<i>Recommendations to Future Research</i>	120
Conclusions.....	122
BIBLIOGRAPHY	123
APPENDICES	140
A: Teacher Professional Performance Scale.....	140
B: Teachers' Pedagogical Belief and Attitude Scale	144
C: 2+2 Program Response Survey	146
D: 2+2 Feedback Form	150
E: Focus Group Question Protocol	151
VITA.....	153

LIST OF TABLES

Table	Page
1: Research Questions and Research Design Table.....	54
2: Data Analysis Matrix.....	68
3: Comparison of the Teachers' Professional Performance	74
4: Comparison of Preparedness for Instruction, Management of Instruction Time, and Management of Student Behavior	76
5: Comparison of Instructional Presentation and Monitoring of Student Performance ...	77
6: Comparison of Providing Reinforcement and Feedback, Facilitating Instruction, Communicating with Students and Chalk Board Skill	78
7: Comparison of Observation between 2+2 and Comparison Group.....	81
8: Comparison of Feedback Between 2+2 and Comparison Group	83
9: Comparison of Collaboration in One Month	85
10: Responses in Compliments and Suggestions Categories.....	88
11: Responses of Non-Productive Compliments by Category	90
12: Responses of Non-Productive Suggestions by Category.....	90
13: Teachers' Description of the Implementation of 2+2 Program	93
14: Teachers' Comparison of "2+2" with the Traditional Appraisal System.....	94
15: Teachers' Perceptions of the Strengths of 2+2 System	96
16: Teachers' Perception of the Weaknesses of 2+2 System	97
17: Teachers' Perceived Progress with 2+2 Program	98
18: Teachers' Description of the Benefits from 2+2 Program.....	99
19: Teachers' Reaction to Being Observed.	100
20: Teachers' Reaction to Being an Observer.	101
21: Suggestions Provided for Improving 2+2 Program	102
22: Plans for the Next Steps for "2+2"	104

CHAPTER I

INTRODUCTION

三人行，必有我师焉。择其善者而从之，其不善者而改之。

---孔子《论语》

Among any three people walking, I will find something to learn for sure. Their good qualities are to be followed, and their shortcomings are to be avoided.

---Confucius 'Analects'

China's Educational Reform Efforts in Teacher Education

The Chinese education system is going through fundamental changes following the economic and social reforms. Since the 1980s, among other things, education has been a focal point of the nation's social reform plan. The late Chinese leader Deng Xiaoping asserted soon after he reclaimed his leadership in the Chinese Communist Party that the nation's modernization hinges upon education. The key for future success of China's modernization is additional talented personnel. To improve the education of its people, China has to create quantities of teachers who are well-motivated individuals with skills and adaptability to match the development of science and technology (Zhang, 1994).

Education reform has been moving at a fast pace during the last two decades, and its central focus has been improvement of the overall quality of teachers. In 1983, Deng Xiaoping declared that "teachers hold the key to a school's success in training qualified individuals of socialist construction and training workers who develop morally, intellectually, and physically, and with both socialist consciousness and culture" (CPC Central Committee, 1985). Ever since then, there has been a rapid development of teacher

education in China. A nationwide network of teacher preparation and professional development has been established.

Altogether six types of regular public institutions presently exist in the country where people can get preservice teacher preparation: normal universities, vocational universities, normal colleges, normal schools, special education normal schools, and preschool normal schools. These institutions vary in length of program from two to four years, in degrees from B.A. to B.S. to certificate, and in the positions that graduates will be assigned. Normal universities provide four-year degree programs to train teachers for senior high schools; vocational universities usually have departments of education that are specialized in training teachers for vocational schools; normal colleges are involved in training junior high school teachers who get a diploma certificate for teaching; normal schools admit both junior and senior high school graduates to train them to become elementary school teachers. The country also has three kinds of inservice teacher training institutions: provincial teacher training colleges, regional and municipal teacher training colleges, and county-level teacher training schools (State Education Commission, 1996).

It is a tremendous success for the country to establish a large number of new normal institutions in the past two decades. Up till 2001, 96.81% of the 5,797, 700 elementary school teachers are qualified in terms of the diploma they have obtained, while in 1977 only 47.1% were regarded as qualified. The percentage of qualified junior high school teachers has raised from 9.8% in 1977 to 88.81%, and that of the senior high school teachers has raised from 45.9% to 70.71% (China Education and Research Network, 2001). However, many have only met the minimum requirements for normal institutions. Many of the new teachers they have turned out are unqualified or of low

quality. A great number of teachers find themselves unable to carry out their instructional duties successfully (Yang & Wu, 1999).

In order to help those inexperienced and under-qualified teachers to grow, many schools set up projects to have experienced and qualified teachers to work with their colleagues and peers who are regarded as professionally under-qualified. Instructional experts from outside of schools are also included to assist in their professional development activities. In Shanxi province, the provincial government has been funding various school-based professional development projects for years in order to improve teaching and learning.

Shanxi province, located in the northwestern part of China, has a population of about 30 million people, of which over five million are receiving primary and secondary education (Shanxi Education Commission, 2000). To provide adequate education and training for such a huge population is an extremely hard task for the province government, but it is an ultimate aim which is being pursued consistently. Serious consideration is given to issues concerning educational reform and school improvement. The most important of these is to empower teachers to provide better service to their students (Shanxi Education Commission, 2000).

In China, the national standards for teacher qualification require that primary teachers must hold a certificate from a normal school; middle school teachers must have a college diploma and high school teachers should have obtained a bachelor's degree or a qualification equal to a degree. It now is acknowledged widely that a great number of those teachers who have obtained the required diploma and certificate and thus regarded as qualified may lack the adequate knowledge and skills to perform their professional

responsibilities, while in Shanxi about 15% of the teaching force do not have the required diploma and degree that they are classified as unqualified according to the national standards (Shanxi Education Commission, 2000). These teachers are required to continue their education in higher teachers training institutions and to participate in in-service professional development programs before they are regarded as qualified (Shanxi Education Commission, 2000).

Teachers face more changes and challenges in their profession. Teaching has become a continuous learning process for teachers as well as for students. Over the past five years, attempts have been made to link teacher development more closely with the implementation of educational policies. In 1995, a decision was made by the Province's Educational Commission that teachers' promotion and salary increase of pay are determined not only by years of service, but also the amount and quality of in-service training received (Shanxi Education Commission, 2000).

One major initiative of the province's educational reform package is the 2+2 Alternative Teacher Performance Appraisal Program (2+2). The 2+2 program is an experimental alternative to the province's teacher performance appraisal system. A study has been planned in which three high schools in the province that are now using 2+2 as an alternative in assessing and evaluating teachers' professional performance were compared to three other matched high schools. Based on frequent peer and administrator observation and feedback, the 2+2 program was developed to provide more frequent, less formal feedback to teachers. Specifically, 2+2 performance appraisal protocol is two compliments plus two suggestions. The protocol was designed to help to reduce teacher isolation and increase feedback. The program was also designed to foster a collaborative

culture that will lead to an exchange and implementation of successful instructional strategies. In Shanxi province, the 2+2 protocol existed first (Sep, 1998-Sept, 2001) as one of the components of a major teacher development project—the Lead Teachers for the 21st Century Shanxi Province Training Program (LTTP). Then beginning from September 2001, the 2+2 program has been implemented as an extension program of the LTTP to serve individual schools that can help in-service teachers grow professionally (Shanxi Research Center for Secondary Education, 2001b).

Context of the 2+2 Program in Shanxi Province

The following description of the Lead Teachers for the 21st Century Shanxi Province Training Program (LTTP) provides a context and background for the 2+2 Alternative Teacher Appraisal Program. Taiyuan Normal University (the former Shanxi Institute of Education) and the Division of Teachers' Professional Development under Shanxi Provincial Educational Commission have been engaged in the implementation of the program. Sources for the program include the LTTP Project Mission Statement and Guidelines, developed by the project advisory group led by Weiping Wang, the current researcher (Shanxi Research Center for Secondary Education, 2001b).

The LTTP was part of a reform effort proposed at the Province's Secondary Educational Reform Conference hosted by the provincial government in January 1998. In March of 1998, a funding proposal for the LTTP program was submitted to Shanxi Provincial Educational Commission. The proposal was approved and the program was launched in September of 1998. As one important piece of its secondary education reform package, the government declared that a total of 500,000 *yuan* (about \$55,000) would be provided for the training of 300 high school lead teachers in three years. Those lead

teachers would be expected to initiate educational reforms at school levels, and to play a role in helping their colleagues to grow professionally (Shanxi Research Center for Secondary Education, 2001b).

The principal of each of the 43 provincial-nominated key high schools was invited to recommend six participant candidates, which made up 258 of the total candidates, and principals of the other 242 ordinary schools could each recommend one candidate. A pedagogic knowledge test was conducted on those 500 candidates to examine their knowledge level in educational and pedagogical content. Finally 300 teachers were selected, based on their test scores, from 500 candidates to participate in the training program. They were required to participate in workshops organized by the center during summer and winter vacations (Shanxi Research Center for Secondary Education, 2001b).

The educational leaders of Shanxi Province understand that 21st century teachers' professional development programs must train teachers to improve their performance in the everyday classroom settings, where they must be empowered with new ideas along with new instructional strategies, methods, and techniques. The traditional mode of teaching which incorporates elements such as teacher-centered instruction, single-sense stimulation, single media, single-path profession, isolated work, information delivery, and passive learning must be transformed into an alternative that includes student-centered instruction, multi-sensory stimulation, multimedia, multi-path profession, collaborative work, information exchange, and active learning (Shanxi Education Commission, 2000).

According to the merit pay policies of the province's Educational Commission, all the 300 lead teachers that have participated in training are now paid more for more responsibilities than other ordinary teachers. One of their responsibilities is to help other

teachers in their professional growth and to initiate educational reform efforts in their own schools. They are instructed to introduce whatever they have learned, including the 2+2 system, to their colleagues and demonstrate for them how to apply those new theories and skills in their daily work settings. By September 2001, of the teachers who have experienced the LTTP project, about one third has been using the 2+2 in their own school on a voluntary basis. Three schools in the province have made 2+2 part of their school assessment and evaluation policy.

The 2+2 represents a new system that requires some systematic changes in the school system. The administration and teachers need to have a different mentality about observation and evaluation for them to understand that 2+2 represents a new concept of classroom observation which focuses on the power of feedback, encouragement, discussion and discourse, and the importance of perspective and collaboration in the improvement of instruction (Allen, Nichols, & Leblanc, 1997). The implementation of 2+2 in schools requires some organized efforts from almost all parties in education. Thus the provincial Educational Commission decided that 100,000 *yuan* (about \$13,000) would be provided to fund an experiment on 2+2 as a continuation and follow-up of the LTTP project. The official decision made by the Educational Commission stipulated that the program would be conducted in a small number of schools before it can be made into a policy for all schools.

The Content of the 2+2 Program in Shanxi Province

The 2+2 protocol was first developed by Allen in Namibia in 1994 while he was working with completely untrained teachers who had little access to trained supervisors. He then transported the protocol to China in 1995, while serving as the Chief Technical

Adviser of the educational programs funded by the United Nation's Children's Fund (LeBlanc, 1997a).

The purpose of the 2+2 protocol is straightforward. It is designed to maximize professional interactions, decrease teacher isolation, and increase meaningful feedback that will lead to improved instruction (Shanxi Research Center for Secondary Education, 2001b). The essence of the 2+2 Alternative Teacher Performance Appraisal Protocol is a series of regular classroom observations by teachers and administrators. The observer visits a classroom and makes two compliments and two suggestions for improvement or change. Observers just stay as long as needed to write down two compliments and two suggestions. It usually takes 10 to 15 minutes. The 2+2 feedback forms are completed and left in a bag or box designated for the purpose or handed to the teacher who was observed some other time during the day, allowing for almost immediate feedback to the teacher. The process is the same whether the observer is an administrator or teacher.

The premise of the 2+2 protocol is simple. It is a shared belief among those 2+2 users that there is no such thing as perfect teaching that nothing can be changed or improved; and there is no such thing as teaching so bad that nothing about it can be complimented. Everyone understands from the outset that each classroom visit will result in two compliments and two suggestions. Compliments are expected to be authentic and sincere. Both negative and positive suggestions are anticipated. In 2+2, no summative rating or comparison is required or desired. Theoretically, improvement of teacher performance comes from a mutual recognition of strengths and the offer of suggestions that can be implemented successfully in similar settings or identified as agendas for future improvement and development (LeBlanc, 1997a). It is very important

that compliments are put before suggestions. Those teachers interviewed by LeBlanc (1997a) responded that encouragement is far more powerful than criticism in learning new skills and forming new behaviors. A teacher is more ready to accept suggestions when compliments are offered first.

The 2+2 appraisal system was designed to provide more opportunities for teachers to give and receive feedback. Lack of feedback contributes to the failure of school reform efforts that intend to change and improve classroom practice. "People have always operated on feedback. Sometimes they do it deliberately. Then they reflect upon their basic purposes and the values they hold dear; they ponder over how a situation had worked out to this point and the problems they have to solve to achieve their goal; finally, either slowly or in a flash, they make up their minds as to what is the best thing to do next" (Wilhelms, 1967, p. 292). Teachers need frequent feedback to grow professionally. 2+2 is designed to provide more feedback to teachers. "Teachers receive multiple feedback on their work and are able to gain an appreciation for innovative and diverse approaches used by other teachers" (Beerens, 2000, p. 44). The reality, however, in most Chinese schools is that an average teacher gets feedback only once or twice a year from the administration. With 2+2, marginal teachers, new teachers and lead teachers are expected to experience more observations (Shanxi Research Center for Secondary Education, 2001b).

The 2+2 system intended to serve as a channel for teachers to value one another and contribute to each other's job performance to help to increase professional interaction and collaboration among teachers. The assumption is that as teachers engage themselves in others' instructional activities they have opportunities to value others' strength as well

as weakness, which helps to establish and build a climate of mutual understanding, trust, and commitment to one another and the school (LeBlanc, 1997a).

The 2+2 program was planned as an alternative to conventional teacher appraisal systems that would better serve assessment and professional growth objectives. It is expected to become an alternative to current teacher evaluation that exists in most Chinese schools. Although traditional teacher performance appraisal often has negative connotations, rarely does either high or low quality teacher performance have any consequences. Despite this, administrators utilize valuable time in the standard appraisal process, and teachers feel threatened by it (LeBlanc, 1997a). If the program is successful with those three urban schools by September of 2002, all the other key high schools would include the 2+2 protocol as one part of their assessment and evaluation policy. Eventually, all the schools in the province are expected to use 2+2 for teachers' professional growth and educational improvement.

Being the most experienced teacher appraisal organization in Shanxi, the Research Center for Secondary Education in Taiyuan Normal University was entrusted with supervising the program and serving as the central office that delivers training materials and advises concerning the implementation of the program. In June 2001, *the 2+2 Program Implementation Handbook* was issued by the program central office, in which a detailed description of the system was provided, and guidelines were laid down for the implementation of the program. The guideline pointed out that the one lead teacher in each school was included in the program not as a participant, but trainer and consultant. The trained lead teachers in the treatment group schools were instructed first to of all introduce the 2+2 protocol to the first grade (equal the 10th grade in the United

States) teachers and help them use the system for the purpose of professional growth and instructional improvement.

The 2+2 alternative teacher performance appraisal system is now part of the assessment and evaluation procedure of the 3 treatment group schools. As is described in the *School Evaluation and Assessment Policy* issued by the Project Central Office on behalf of the participating schools, “Educational reform starts with classroom instructional improvement, which is for the most part in the hands of teachers. Teachers should be empowered to change and improve classroom teaching continuously. It also helps to alleviate isolation, promote professional growth, encourage mutual feedback, and reestablish vital connections and collaboration among teachers” (Shanxi Research Center for Secondary Education, 2001b, p. 2). *The 2+2 Program Implementation Handbook* (Shanxi Research Center for Secondary Education, 2001a) also includes a model for routine feedback by administrators and resource personnel.

In October 2001, the current researcher visited all of the three participating schools which had made 2+2 part of the assessment and evaluation policy. An evaluability assessment (EA) was conducted by the researcher. An evaluability assessment is a process concerned with identifying the “program objectives, expectations, and causal assumptions of policy-makers and managers in charge of the program” (Wholey, 1987, p. 77). The evaluation needs of managers and measurement criteria have been established, while probable uses for information about the program’s performance has been identified (Wholey, 1987).

The first step in the EA included a review of all program documentation related to the 2+2 teacher performance appraisal system. The purpose was to deliver an actual

description of the program's operation, the participants, and to determine the components of the 2+2 program which were evaluable. A Documental Model was developed to deliver a description of the program. A flow chart graphically depicted the process and components of the program and presented a visual representation of the program's formally stated components and goals and their linkages.

The second phase of the EA was to discuss the Document Model with program managers. The program manager Ms. Xueqin Chang and three school principals were interviewed. The interviews were conducted to: (a) define program goals and objectives; (b) identify program components which were not included in the model; (c) confirm the operation of program components; (d) clarify the linkages specified between program components and goals. The Program manager's model was developed as a result of those interviews.

The EA was employed by the researcher to discuss with the stakeholders about how the program would be evaluated, specifically, what was expected from each participant in carrying out program evaluation tasks. The whole process was completed in September, 2001. Based on the evaluability assessment, the 2+2 was determined as evaluable.

Statement of the Problem

Reform in the Chinese educational system has been occurring over the past two decades. Largely, the central government and its ministry of education have been trying to change organizational structure and curriculum by legislative order and standardized tests based on the assumption that improvement in schooling would inevitably follow. Beginning from the mid 90's, local governments have exercised their authority and the

influence to change teaching practice by educational policies. Much of their focus has been on teachers' professional development coupled with rewarding and promoting policies. However, all these restructuring efforts and systemic reform of schools have created limited success in initiating many positive changes at the school level. Teachers' attitudes, practices and competencies have not been changed much as has been expected by those educational reformers on all levels (Shanxi Research Center for Secondary Education, 2001b).

School reform can not succeed without changing the school culture. Researchers (Eisner, 1992; M. Fullan, 1994, 1996; Sarason, 1995, 1996) have identified the need for change in school culture to occur before lasting instructional change can take effect. Changing an individual teacher's attitudes and performance, which is grounded in inquiry, reflection, and experimentation, is the root of changes in school culture.

The school culture of teacher isolation is another major inhibitor of school improvement. It is clear that the daily routines of schools provide little time and few opportunities for teachers to interact and share ideas with each other, and teachers are not empowered to exert influence on each other's improvement process of teaching practice (Little, 1990; Lytle & Fecho, 1991). No system exists for peer support in pursuing professional growth and instructional improvements. The 2+2 program is designed to help to change the current school culture reflected in teacher isolation, and build a positive and productive relationship among teachers (LeBlanc, 1997a). 2+2 serves as a channel for teachers to value one another and contribute to each other's job performance. The premise is that the extent to which teachers engage themselves in others' instructional activities offers opportunities to value others' strength as well as weakness,

determines in large measure the capacity that can be established and built upon a climate of mutual understanding, trust, and commitment to one another and the organization.

Teacher evaluation as currently practiced in most Chinese schools is flawed. Administrators usually give teachers periodic evaluations or appraisals on their classroom performance. But activities of this nature do not happen often. When an evaluation does take place, the evaluation report consists of so many things that a teacher can hardly determine where to begin with improvements. Educational Evaluation in China indicated that teachers tend to be confused when too many things come up for them to consider, and it is still harder to change too much at one time (Shera, 1992). In the current process of evaluation, teachers play a very passive role. So most teachers tend to resist evaluations and appraisals for the simple reason that they are often troublesome and not very helpful (Shera, 1992). The evaluators know very well about this. For all the practical purposes, the ratings must be completely positive and nondiscriminating that makes it nonsignificant in helping teachers improve their job performance (Shera, 1992).

The 2+2 initiative also is created to change the existing flawed mechanisms of teacher evaluation. It is to turn the evaluation efforts that are essentially punitive, creating anxieties and confusion for the sake of the evaluator's sense of authority into a process that can engage both the evaluator and the evaluated. Teachers interviewed by Leblanc responded that they are less likely to be threatened when being observed and evaluated and that they have more opportunities to observe each other's teaching and learn from one another (LeBlanc, 1997a).

It was expected that the participating teachers' instructional performance would be improved, and their knowledge level about teaching and education in general would be

enhanced as a result of the 2+2 program. It was predicted that the 2+2 program would facilitate the development of a collaborative school culture supportive of systemic educational reform. To address the need for program evaluation, a utilization-focused evaluation was designed to provide a means for program assessment.

Purpose of the Study

The purpose of this study is to examine the effectiveness of the 2+2 Alternative Teacher Performance Appraisal System that has been implemented as one of the educational reform efforts in Shanxi province in China. Information was gathered and analyzed to reveal how the program has been implemented. Results were presented to the stakeholders, including Shanxi Educational Commission, the local school districts, and the participating school, to help them to determine whether or not the program should be further implemented.

The study focused on outcomes associated with the 2+2 program. The outcomes included an analysis of the pre and post-test scores, an analysis of the feedback forms and a comparison of the two groups of teachers in terms of the pedagogical beliefs and attitudes, and the teachers' performance in classroom. A survey was conducted to study the participating teachers' attitudes toward performance appraisal systems and toward the 2+2 program, and whether isolation was reduced and whether collaboration was increased as a result of the program. The effectiveness of the 2+2 system would be demonstrated if those teachers who had been applying it perform better than those who had not been using it. Implication recommendations were made to help enlighten thinking about the issues of teacher evaluation and professional development more generally.

Research Questions

1. How effective is the program in helping teachers to improve their professional performance?
2. How has the program influenced teachers' pedagogical belief and attitude?
3. Does the frequency of feedback provided differ between teachers in the treatment group and those in the comparison group?
4. Do teachers in the treatment group collaborate more in their teaching practice than those in the comparison group?
5. What kind of feedback is provided to teachers who participate in the 2+2 program?
6. What are teachers' perceptions of the advantages and disadvantages of the 2+2 system?
7. What suggestions do teachers provide for improving the 2+2 program?

Research Design

The evaluation was a quasi-experimental design in which six key urban high schools were selected and randomly assigned to either the treatment group or the comparison group. The research questions that resulted from the initial investigation were addressed by employing mixed methods within the quasi-experimental design.

A quasi-experimental design was employed to examine the effectiveness of the program of the 2+2 Alternative Teacher Performance Appraisal System that had been implemented as one of the educational reform efforts in Shanxi province in China from its inception through to its completion. Comparison was established to evidence the effect of the 2+2 program on teachers' professional performance in terms of their demonstration of skills or competency in class. Six urban high schools were selected by the Central

Office for the Program's Implementation from among the 43 provincial key high schools to participate in the program: The Provincial Experimental High School in the city of Taiyuan, the No. 1 High School in the city of Xinzhou, The No.1 High School in the city of Jincheng, the No.1 High School in the city of Yuci, the No. 1 High School in the city of Datong, and Quwo High School in the city of Quwo. These schools were selected because they share some common characteristics in terms of their size, students' achievement level, and teachers' educational background. Most important they were all key schools in urban areas in Shanxi.

Significance of the Study

Traditional performance appraisal systems have played and still play a very important role in teachers' professional development enhancement. However, the long process in addition to the images of evaluation, supervision, assessment, threat and comparison makes it difficult for teacher performance appraisals to occur as frequently as it is needed(LeBlanc, 1997a). The 2+2 system is not to replace the current appraisal systems but to adjust the formality and time-consuming nature of it.

One of the reasons this evaluation is important is that it adds to the body of knowledge regarding teacher performance appraisal systems. 2+2 is very recent. Very little empirical research exists regarding this alternative teacher appraisal system. This study will provide data that can be added to the pool of research concerning teacher performance appraisal systems. The 2+2 protocol may not replace those traditional performance appraisal systems, but it can stand out as a way of conducting day-to-day appraisal practices in schools.

Second, this evaluation is timely for key urban high schools in Shanxi province. Systemic reform projects such as LTTP need to be examined carefully in terms of their assumptions and objectives and be prepared to adjust program elements and implementation as a function of the dynamic and complex nature of educational reform efforts.

A third reason that this evaluation is significant is that it provides formative information that may be utilized for program improvement as well as summative information about program effectiveness. Schools in Shanxi province can readily include the 2+2 program in their educational reform undertaking once the program is proved effective. It is expected to provide teachers with more opportunities to receive feedback and encouragement for their peers and administrators.

A fourth reason that the evaluation is useful is to help teachers to be aware of their own strengths and weaknesses of their teaching. Professional development may be enhanced because of implementation of the 2+2 program.

Last, the study will help urban senior schools in Shanxi province to establish a teacher performance appraisal model. The word “urban” means different things depending on what country is being discussed. In China, an urban area comprises one central place that has a minimum of 250,000 persons who are not involved in farming, but services and industries, and the adjacent surrounding territory which is inhabited by farmers who supply farming products for the central place (Bureau of Urban Management, 1987). In China, being urban means more advanced and better life. The six Chinese city schools where the program is conducted share some common characteristics of urban schools in the country: more resources and better teachers, but more one-child-

family students, more discipline problems, low motivation. However these six high schools belong to the best of Chinese urban schools. They have been nominated key schools by the provincial government for their remarkable academic performance, which resemble the American magnet schools. Part of their responsibility is to provide educational reform models for other schools. It is very important to identify means of developing a culture of teacher collaboration in support of improved instructional practice.

CHAPTER II

LITERATURE REVIEW

The 2+2 program is an alternative appraisal process, a teacher professional development model, a peer observation system, and a teacher collaboration mechanism, which, taken together, are expected to bring improvement of instruction and change of the school culture. As an alternative appraisal process, 2+2 shares those elements that characterize the conventional evaluation system. But in contrast to the traditional appraisal system, the 2+2 process focuses feedback and improvement. Through 2+2 teachers experience the giving and receiving of compliments and suggestions that can help them to improve their teaching practices. The professional communication among teachers may reduce teacher isolation and bring teachers into a collaborative setting (Allen et al., 1997).

The intent of this chapter is to lay the foundation for this study through theory and research. The following seven areas are included in the literature review: (1) peer coaching; (2) teacher evaluation; (3) teacher collaboration; (4) teacher isolation; (5) feedback; (6) teachers' professional development; (7) an overview of the previous evaluation of the 2+2.

Peer Coaching

The term "peer coaching" is used most often in the context of teachers appraising each other for the purpose of professional performance improvement. It is a form of collaboration that is most closely related to the 2+2 system. Peer coaching was first proposed as an on-site dimension of staff development (Joyce & Showers, 1981). It is largely regarded as a technique to develop and refine teacher behaviors (Joyce & Showers, 1982; Showers, 1985; Showers & Joyce, 1996; G. M. Sparks & Bruder, 1987). Showers (1985; 1987) investigated the hypothesis that

coaching, following initial training, would result in much greater transfer than would training alone. They assumed that the coach needed to have more expertise in the content area, and thus paired teachers with an outside consultant or an expert peer, who shared aspects of teaching, planned together, and pooled their experiences. Results showed that teachers who had a coaching relationship practiced new skills and strategies more frequently and applied them more appropriately than did their counterparts who worked alone to expand their repertoires. Members of peer-coaching groups exhibited greater long-term retention of new strategies and more appropriate use of new teaching models over time (Showers, 1984). Showers and Joyce (1996) advocated the following principles to guide a peer coaching program: (1) When working with entire faculties, all teachers must agree to be members of peer coaching study teams; (2) It is necessary and important to omit verbal feedback as a coaching component; (3) The meaning of "coach" needs to be redefined: when pairs of teachers observe each other, the one teaching is the "coach," and the one observing is the "coached." In this process teachers who are observing do so in order to learn from their colleague; (4) Teachers learn from one another while planning instruction, developing support, materials, watching one another work with students, and thinking together about the impact of their behavior on their students' learning.

The following studies provided some empirical evidence about the theory behind the system of peer coaching. Englert and Sugai (1983) used a 2 x 2 factorial design to compare the effectiveness of peer observation with well-defined observation systems and peer observation with a less well-defined observation system. The dependent variables in the study were four pupil behaviors (e.g., percent of correct/incorrect pupil responses) and six teaching/management behaviors (e.g., percent of teacher reinforcement following correct student responses). Peer coaches

using the less well-defined observation system were instructed to develop their own system of recording teacher/pupil behaviors. Twenty novice pre-service trainees (12 in the experimental group and 8 in the control group) coached one another. Peer coaches in both groups provided verbal feedback following practicum lessons; written feedback was provided after all observations were completed. Results indicated no differences between the groups in the use of behavior management outcomes. However, the group using a well-defined observation system maintained a higher level of pupil accuracy. In addition to that, teams using the well-defined system used more effective feedback strategies than peer teams using the less well-defined observation system.

Peer coaching is a simple and effective system, but requires teachers to apply various strategies in its utilization. Morgan (1992) investigated the effects of one peer coaching strategy in a multiple baseline design. They employed three coaches who had mastered the required teaching behaviors in a previous practicum provided supervision for five preservice teachers who were having difficulty learning the required teaching skills. The required teaching skills were the dependent variables in this study and included correct and incorrect use of procedures for instruction (e.g., follows a model, lead, test sequence), positive and corrective feedback, assessment, and reinforcement (e.g., scans pupils and delivers/records points). Peer coaches provided immediate, behavior specific feedback to preservice teachers during reading instruction. Results indicated that effective or correctly implemented teaching behaviors increased, and ineffective or incorrectly implemented teaching behaviors decreased for all five preservice teachers.

The effects of peer coaching were investigated by comparing other conditions. Kohler (1997) used a multiple-baseline design to compare three conditions. Four

teachers planned and conducted an instructional innovation independently during an initial baseline phase. Conversely, these tasks were conducted with a peer coach during a second phase and then alone again in a final maintenance setting. Four results were reported: (1) The four teachers made few changes or modifications in their instructional approach during the first phase. (2) Many of the changes instituted during coaching were sustained in a follow-up or maintenance condition. (3) The instructional activities were associated with a variety of different teacher and student processes. (4) The four teachers expressed varying degrees of concern and satisfaction with the innovation.

In Bowman and McCormick's (2000) study, 32 preservice elementary education majors were randomly assigned to an experimental group that employed peer coaching dyads or to a control group that experienced a traditional supervision approach. Further, they randomly placed the two groups in elementary schools in a large urban district and, as a third randomization process, they randomly assigned them to cooperating teachers within the schools. The study lasted for seven weeks. Results indicate that those who participated in peer coaching achieved greater success.

Peer coaching is practiced to serve various educational purposes. Gottesman (2000) states "Peer Coaching is a simple, nonthreatening structure designed for peers to help each other improve instruction or learning situations. The most common use is teacher-to-teacher peers working together on an almost daily basis to solve their own classroom problems" (p. 5). He advocates a model of peer coaching consists of five steps and simple rules to follow to keep the personal or judgmental aspects out of feedback and coaching. The 5 steps are: (1) the teacher Requests a Visit; (2) the Visit; (3) the Coach Reviews the Notes and Lists Some Possibilities; (4) the Talk after the Visits; (5) the Review of the Process (p. 33). The motto for his model is "No Praise,

No Blame”(p. 8). He thinks that the purpose of peer coaching is to provide for the transfer of training elements in an everyday situation so that teachers can manage them. “Its major purpose is to help implement new training or help sustain existing training so that the training will impact on student learning in the classroom. Peer coaching ensures that theory, demonstration, practice, feedback, and coaching transfer any training from the lecture room to the classroom with ongoing troubleshooting and renewal” (p. 23).

In addition to the skill levels of the coach and preservice teacher, collegiality is an important factor. Cooperativeness was considered as a willingness and desire to work together, without which other efforts at facilitating peer coaching may be undermined (DeRoche, 1987; Morgan, Menlove, Salzberg, & Hudson, 1994).

The 2+2 system resembles peer coaching in that it is also a framework designed for teachers to help one another to improve their teaching practice. But they differ in one major aspect. In peer coaching teachers are paired in order for one to perform as “coach” and the other as the “coached”, while in 2+2 every teacher can be “coach” and the “coached”. Peer coaching is a mechanism for experienced teachers to disseminate their knowledge and skills to those newcomers into the profession while 2+2 is for every teacher to obtain feedback about his or her teaching more frequently to build up improvement of instructional practice.

Teacher Evaluation

Teacher evaluation is part of teachers’ professional life. Teachers are usually evaluated by their principals or heads of departments, by school inspectors or by specially assigned evaluators. Two purposes for evaluating teachers were introduced: to improve teacher performance, and to provide a measure of accountability (DeRoche, 1987). Peterson (K. D. Peterson, 1987) listed three purposes. “Teacher

evaluation is key to understanding effective teacher practice, rewarding excellent performance, and improving training programs” (p.56). Conventional wisdom states that evaluation process is an important road leading to teacher growth, “the major positive aspects of teacher evaluation are that the process improves teacher-principal communication, it increases teacher awareness of instructional goals and classroom practices, it impacts instruction and classroom practices, it gives teachers a sense of pride and professionalism, and it increases public confidence in the schools ” (DeRoche, 1987).

Teacher evaluation has focused on two major areas, formative and summative. Formative evaluation is as ongoing process designed to improve the teacher’s performance that emphasizes promoting teacher growth. Its intent is to help teachers become more effective (Egelson, 1994). This type of evaluation builds a helping process which can provide data to teachers for making decisions about how they can best improve their own teaching techniques, styles, or strategies. While summative evaluation is a judgmental decision of the quality and worth of an individual teacher over a specified time frame. Summative evaluation is usually used for accountability and to determine if a teacher meets specified standards (Dagley & Orso, 1991). The purpose of this type of evaluation is to determine whether to dismiss a teacher, grant tenure, place on probation, or grant merit pay.

However, traditional teacher evaluation is far from being a successful story. Most evaluation instruments rely on the false assumption that accurate and objective assessment of performance is possible (Fox & Shirkey, 1997). In discussing the general nature of employee evaluations, Banks and Murphy (1985) stated that “Effective performance appraisal in organizations continues to be a compelling but unrealized goal”(p.36). Peterson (1995) held that research on teacher evaluation

conducted in the past provides compelling evidence that current practices in this area neither “improve teaching [nor] accurately tell what happens in classrooms” (p.14). Further, he said that research “reveals fundamental technical and sociological flaws with present teacher evaluation” (p.15). Schmoker (1999) wrote, “Research has finally told us what many of us suspected all along: that conventional evaluation, the kind the overwhelming majority of American teachers undergo, does not have any measurable impact on the quality of student learning”(p.45). Ellett and Garland (1987) reported the results of a national survey of school district level teacher evaluation practices conducted in 1985 in the 100 largest school districts in the United States. They conducted analysis of the quality of evaluation instruments and methodologies used by the various districts. Key findings of the study pointed to several concerns about teacher evaluation: 1) more emphasis was placed on the use of teacher evaluation data for summative rather than formative purposes; 2) policy bases of local district evaluation systems were somewhat deficient in the areas of establishing performance standards and in implementing comprehensive training programs to train evaluators to make reliable judgments about teaching and learning in classrooms; 3) few systems allowed for the use of outside evaluators or for the inclusion of peer teachers as assessors; and 4) local systems were slow to design procedures to accommodate the potential adverse effects of evaluation context variables on the reliability and credibility of evaluation data and processes.

Ten years later in 1996, a follow-up study of teacher evaluation practices in the 100 largest school districts in the United States was conducted to examine teacher evaluation practices in the country (Loup, 1996). The districts were identified and selected based upon the size of their total student populations. Subsequently, 68 surveys were completed and returned in all. Survey data were analyzed and compared

with results obtained from the prior study in an effort to identify similarities and differences in perceptions of important issues, concerns, policies, processes, and critical features of district evaluation systems. The results of the study indicated that little seemed to have changed in teacher evaluation practices at the local district level during that 10-year period (Loup, 1996).

Extensive interview studies of teachers show that they do not want to be evaluated, do not feel they need it to improve, or do not believe that it can be done. Peterson (1995) pointed out that evaluation is a threat to their livelihood and an intrusion on their time; they do not want or use the results of evaluation. No one wants to be made to look bad at doing something he or she cares about. Classrooms seem to go on well enough without it. There is little or no vision from teacher interview studies about how teacher evaluation could be changed so that it would be believable, credible, useful, and fair.

Many teachers tend to dislike the way they are evaluated (Allen et al., 1997). Boyd (1989) lists several reasons why teachers dislike the traditional evaluation: (a) Teachers do not have input into evaluation criteria and therefore distrust the process and question the validity of the results the process produces; (b) evaluators do not spend enough time on the evaluation because principals are too busy to gather quality information and provide useful feedback; (c) evaluations are done on a sporadic basis; (d) evaluators are not well trained and have little classroom experience, and evaluations are vague, subjective, and inconsistent, leading to a lack of evaluator credibility; and (e) the evaluation process is a dead end as few districts have established a clear link between teacher evaluation and teacher development.

From the perspectives of teachers, evaluation is best when it is served as an opportunity for positive interaction and professional growth. McLaughlin and Pfeifer

(1988) conducted case studies of four school districts that were engaged in implementing and sustaining meaningful teacher evaluation. In the districts that they studied, teacher evaluation was designed to serve the purpose of promoting professional improvement, as well as ensuring accountability. They found that the teachers they interviewed stressed the importance of reflection provided through evaluation for all teachers. Peterson and Comeaux (1990) had 48 high school teachers in four schools in Florida and Wisconsin analyzed two behavioral teacher evaluation systems used in Florida (Florida Performance Measurement System [FOMS] and Teacher Assessment and Development System [TADS], a Self-Evaluation Checklist and another was an Alternative System designed to assess teachers' reflection on practice. Teachers rated the alternative Assessment System most favorably because they judged that the format would be likely to encourage teachers to reflect on their teaching.

Peterson (1987) advocated a model of teacher evaluation with multiple and variable lines of evidence. He developed eight lines of evidence from which teachers could select: student report, parent survey, documentation of professionalism, and "other." Quantitative indicators were developed for five lines. Peer review, student achievement, and "other" produced qualitative or variable data. The sample of his study was from Nebo School District in central Utah during the 1985-86 school year. Of 493 educators eligible for promotion, 332 (67%) submitted dossiers containing results of multiple lines of evidence. this group include 51 "special assignment" educators such as librarians and counselors. The study was conducted on 281 regular assignment classroom teachers. The participants were asked to select a minimum of four lines for promotion in a career ladder system. Results showed that administrator reports had low variation and correlations with other measures. Student reports, parent

surveys, and teacher tests produced sufficient variance for decision-making and moderate correlations with other measures. Professional activity and years of experience showed erratic relations. Overall, the lines of evidence showed independence, suggesting that multiple measures may have tapped different constructs of quality.

The 2+2 system performs some functions of teacher evaluation. During the implementation of 2+2 program in Norfolk public schools in 1996 and 1997, participating teachers in those project schools were exempted from the traditional evaluation based on an assumption that 2+2 can be more efficient, more useful in helping teachers improve their instructional practice (LeBlanc, 1997a). 2+2 provides opportunities for teachers to observe one another teaching and help each other reflect on one's teaching by giving and receiving feedback. The system can serve as a tool for school administrators to appraise teachers' performance and help teachers to grow professionally.

The 2+2 protocol was designed as an alternative to the conventional teacher evaluation system. Redinger (1988) suggests that there is no one way to assess teachers, with each district and/or school needing to develop a method that best fits their needs.

Teacher Collaboration

In a school context, collaboration involves interaction between two or more equal parties who voluntarily share decision making in working toward a common goal (Cook & Friend, 1991). This definition of collaboration is actually interactive teaming as described by Thomas (1995) or collaborative consultation as described by Idol, Nevin, and Paolucci-Whitcomb (1994) in the second edition of their text by the same name.

Teacher collaboration has been generally applauded for its potential in improving the working lives of teachers, reducing teacher uncertainty, enhancing teachers' professional self-image and promoting collegiality and school learning (Kain, 1996).

The history of educational collaboration was viewed as a progressive movement from the prescriptive nature of consultation to the mutual parity of collaboration (L. J. Johnson, 1990). They suggest six steps for developing a more collaborative educational environment in schools:

1. Sanctioning of collaborative efforts by administration.
2. Providing assistance for teachers with clerical work and other non-instructional tasks.
3. Organizing meeting times for teachers to engage in mutual problem solving.
4. Providing opportunities for specialists and teachers to co-teach.
5. Developing common vocabulary and terminology in order to avoid specialized jargon.
6. Reserving regular faculty or in-service meetings for collaboration.

An eight-step process for collaboration was considered involving goal setting, data collection, problem identification, development of alternative solutions, action plan development, action plan implementation, evaluation, and redesign (Idol & West, 1991). 1991). Idol and West (1991) suggested thirteen principles for collaborative consultation that include establishing team member relationships, respect among the team, use of situational leadership, conflict management, information sharing, active listening, nonjudgmental responding, interviewing skills, common language, data gathering, willingness to receive as well as give feedback, giving credit where credit is due, and awareness of nonverbal messages.

Research evidence exists to show that collaboration helps improve students' achievement. Costa (1995) collected data on teacher attitudes, behavior, and efficacy; student attitudes, behavior, and achievement; and classroom characteristics in 26 elementary classrooms in British Columbia. These variables were then related to 4 teacher collaboration strategies. Those four strategies describe how the teacher volunteers worked with their teaching partners: (a) collaborative consultation between teacher dyads teaching in one double sized classrooms, (b) collaborative consultation between team teacher dyads teaching in one double sized classroom, (c) collaborative consultation without direct observation with dyads teaching in separate classrooms, and (d) collegial consultation without direct observation. One of their findings was that classroom observation is a very effective way of teacher collaboration that can increase student achievement and teacher efficacy.

Studies of teacher collaboration in schools have revealed associations between collaboration and outcomes such as collegiality (Stevenson, 1987), increased productivity and expertise (Brandt, 1987), improvement of teaching practice (Crandall & Loucks, 1983), teachers' perceptions of increased learning opportunities (Rosenholtz, 1989b; G. M. Sparks & Bruder, 1987) improvements in school climate and teachers' sense of efficacy (Leggett & Hoyle, 1987), and teachers' preference for collaborative structures (Holly, 1982).

The National Center for Education Statistics (NCES) reported that "the proportion of teachers who felt that participation in various teaching-related activities improved their teaching a lot ranged from 18 percent for mentoring another teacher in a formal relationship to 37 percent for being mentored by another teacher in a formal relationship. The proportion of teachers who felt that participating did not improve their teaching at all ranged from 2 percent for individual or collaborative research to

10 percent for mentoring another teacher in a formal relationship". (The National Center for Education Statistics, 2001, p.18)

Smith (2001) documented the collaborative relationship of three part-time English-as-a-Second-Language (ESL) teachers and the teachers' reflections on their collaborative work, within the context of their lives. Data were collected from a variety of sources: transcribed audiotapes of the teachers' weekly meetings and periodic reflection-on-collaboration interviews, teacher journals, and teacher life histories. He discovered that there were three factors that had been crucial for the teachers' collaboration. The first was their commonly held belief in collaboration as an ideal way for teachers to work together. The second factor was that they would like to talk with their co-teachers about their work to end the isolation they so often felt. Finally, they saw collaboration as a means to give and receive support as they took on challenges in the form of teaching skills they were unfamiliar with and in implementing a new project. A questionnaire was administered on their students about their perception on the teachers' collaborative relationship. The students were asked: " Think about your teachers this semester. In what ways was their work with you helpful?" " How did your teachers work together(i.e., their collaboration) help you?" typical response include: "Because they were very enthusiastic with all work they did together, my spirit got open to learning English in a wonderful way." "I think they worked together very well. They were separated in their work but it was related in work. So we learned as there were just one part." "Their collaboration was helpful for students, because they know about everyone of the students. In addition, their good relationships affected the students." Students perceived teacher collaboration positively and reportedly improved motivation and learning.

Bohlmann (1998) examined the effects of cooperative teaching in a two-teacher classroom. Three situations were clarified in the occurrence of two teachers in one classroom. The first is mentor/specialist in which the mentor is the primary classroom teacher and a specialist teacher works in the classroom for periods of time. The second is mentor/mentor that involves two teachers who are certificated and experienced and who share one classroom and the students within that classroom. The two teachers have joint responsibility for the classroom, the students, and parent relationships. The third situation is mentor/associate that involves a classroom in which one teacher is a certificated, experienced teacher acting in a mentor capacity. The other teacher may be certificated but does not have the same experience as the mentor teacher. The results of her study show that the advantages of such a collaborative system greatly outweigh the disadvantages.

Teacher collaboration does not happen in every school. Kain (1997) used the critical incident technique to study conditions that encourage or discourage teacher collaboration. It was found that time, training, and organizational support were crucial to collaboration. Lack of structured support, pressures to “cover” curriculum and conduct testing, and traditions of tracking and separate-subject instruction discouraged collaboration.

Teacher collaboration becomes a must when dealing with difficult educational situations. Gable and Manning (1997) explored ways of collaboration that can help people face the challenge of students with disabilities. They propose that teachers in all grade levels and disciplines must work with colleagues and other professionals.

Teacher collaboration emphasizes team decision-making (Bauwens, Hourcade, & Friend, 1989) and requires participants to share in the process of setting goals and implementing plans (Bauwens et al., 1989; Brookhart & Loadman, 1990; Cook &

Friend, 1991; Lasley, 1992). The shared planning and goal setting process helps the participants gain ownership of the instructional process and establish mutually satisfactory goals. Each party feels equally responsible for ensuring a positive outcome (Brookhart & Loadman, 1990). Collaboration allows participants to learn from one another and to establish long-lasting and trusting professional relationships (Lieberman & Miller, 1992). Teachers benefit from exposure to others' diverse philosophies, instructional experience; the stimulation of new ideas and the increased communication among professionals at all levels (Brookhart & Loadman, 1990). Teacher collaboration is predicated on voluntary participation, mutual respect, parity among participants, a shared sense of responsibility, and an equitable distribution of available resources (Cook & Friend, 1991).

The 2+2 system is a new framework for teachers to collaborate. It offers opportunities for teachers collaborate in improving their instruction by observing each other teaching, then give and receive feedback. In the context of the current study, the treatment group teachers are expected to collaborate with one another more often than the comparison group teachers.

Teacher Isolation

Starting from earlier research on teacher isolation (Forsyth & Hoy, 1978; Miskel, 1983), isolation is defined as the extent to which teachers are restricted from or restrict themselves from interactions with other individuals or groups in the school. Isolation is referred to as a situation in which a teacher is minimally influenced by and exerts minimal influence on other staff members (Bakkenes, Brabander, & Imants, 1999). "Certainly the physical isolation of teachers is a constant reminder to them of their separateness. The 'classroom as a teacher's castle,' while perhaps once serving

as a useful fortification against ignorant outsiders (both school and community representatives), no longer seems justifiable” (Knoblock & Goldstein, 1971).

Teacher isolation is well documented (Goodlad, 1984; Lieberman & Miller, 1992; Dan C Lortie, 1975), and it is clear that the daily routines of schools provide little time and few opportunities for teachers to interact and share ideas with each other (Little, 1990; Lytle & Fecho, 1991). It is not very easy for teachers to experiment and improve under conditions of isolation, thus perpetuating the status quo in education. Teachers tend to prefer to keep their ideas to themselves, to fear asking for help because they might reveal incompetence, or to fear offering help because they might be perceived as less than humble, works to “institutionalize conservatism” (M. G. Fullan & Hargreaves, 1991).

A survey of 196 elementary and secondary education teachers attending graduate programs in the University of Central Florida's College of Education provided data supporting previous research on teacher isolation (Rothberg, 1985). It was discovered that over 80 percent of the teachers felt their classrooms were private worlds entered only by themselves and their students. The teachers reported that formal and informal visits to their classrooms by observers or evaluators were rare, as were their own visits to the classrooms of other teachers. It was indicated that the teachers would welcome more informal contacts with and visits from their peers. Another finding was that high school teachers appeared more isolated than other teachers. The researcher suggested strategies for alleviating teacher isolation including developing a climate of trust within the school, sharing decision-making power, and using professional development activities to improve communication and team-building skills among teachers. Additional strategies to consider include forming quality circles or other problem-solving committees, presenting meetings focused on

the daily activities of staff members, requiring peer observation, increasing opportunities for staff social activities, encouraging attendance at professional meetings, conducting retreats for sharing values and attitudes, and planning regular informal meetings to discuss teaching techniques or new ideas gleaned from the professional literature.

The concerns about teacher isolation and its negative effect on education are also discussed by Rosenholtze (1989a) in her research on the teacher's workplace. Rosnholtz (1989a) focused her study on 78 schools in eight districts in the state of Tennessee. The results of the study indicates that schools in which teachers have a shared consensus about the goals of their organization work are more likely to incorporate new ideas directed to student learning. In contrast, those "low-consensus schools" more commonly "skirted the edge of catastrophe alone." Schools that had "shared meaning" among teachers were continually changing and improving.

Slater and Trowbridge (2000) recounted the first year of a University/School District partnership to create Master's level cohorts of inservice teachers. Neither their initial goals nor those of the public school administrators specified the forming of teams, overcoming isolation, and developing a sense of collegiality, but these factors appeared as most important to the participants. Teachers have long suffered from isolation. This condition negatively affects their feelings about school and their teaching. The teachers in the cohort appeared to be a team that was overcoming the feelings of isolation so common to the teaching profession. The teachers had a sense that they were special within the school district.

The path to reduce teachers' professional isolation is to build a collaborative work culture. Dialogues with colleagues and joint work provide the continuous pressure and support necessary for educational reform and change. It helps to relieve

teachers from classroom isolation and allows the codification and sharing of successful practices and the provision of support (M. G. Fullan & Stiegelbauer, 1991; Hargreaves & Dawe, 1990; Little, 1990). Teachers working together have the potential for higher morale and enthusiasm. It helps to open the door to experimentation and increased sense of efficacy (Cohen, 1988; Rosenholtz, 1989a).

One of the objectives of the 2+2 program was to help teachers to interact more with one another and feel less isolated in their professional environment. It was assumed that 2+2 would bring a teacher more often to another's classroom to observe, thus teachers would involve themselves more in discussing matter concerning teaching.

Feedback

"People have always operated on feedback. Sometimes they do it deliberately. Then they reflect upon their basic purposes and the values they hold dear; they ponder over how a situation had worked out to this point and the problems they have to solve to achieve their goal; finally, either slowly or in a flash, they make up their minds as to what is the best thing to do next" (Wilhelms, 1967). In terms of instructional feedback, Gil (1987) defines feedback as information provided to instructors about their performance and recommendations for future improvement, while evaluation is viewed as information provided to make judgments regarding the worth of the performance.

Anderson (1998) described four types of feedback provided to preservice elementary teachers. The feedback includes field notes, checklists, anecdotal accounts, and dialogue journals. He provided examples of each, following a coaching model. The model (a) identifies what was done well during the lesson, (b) describes some

important areas that need improvement, and (c) suggests specific ways to accomplish the improvement.

Performance feedback was demonstrated effective on modifying principal and teacher behavior (Gillat & Sulzer-Azaroff, 1994). In their first experiment, Gillat and Sulzer-Azaroff trained a principal to provide students with positive verbal and nonverbal performance feedback, and goal setting. The result of this first experiment indicated that the rate of positive feedback statements, from the principal to the students, increased as a result of the feedback from the authors to the principal. Consequently, the students in this first experiment showed gains in academic performance. In a second experiment, Gillat and Sulzer-Azaroff coached a principal of a secondary school to train the classroom teacher to implement the positive performance feedback and goal setting protocol. The principal performed as intervention trainer and provider of performance feedback. As a result of the feedback from the principal, the teacher implemented the classroom intervention with greater integrity. Similar to the previous experiment, students improved their academic performance from baseline academic scores (Gillat & Sulzer-Azaroff, 1994).

More consistent effects on treatment implementation have been observed in studies utilizing performance feedback (Green & Reid, 1991; Harchik, 1992). Performance feedback has been defined as a method of providing information or knowledge of processes and results to promote transfer or maintenance of skills and behaviors (Arco, 1991; Duncan & Bruwelheide, 1985; Fleming & Sulzer-Azaroff, 1989; Green & Reid, 1991; Hawkins, Burgio, Langford, & Engel, 1992). In order for performance feedback to be effective, the individual administering the feedback must be able to manipulate rewards and punishers related to the target person's performance has significant bearing on the effect of the feedback (Duncan & Bruwelheide, 1985).

In their thorough review of the performance feedback literature, Balcazar, Hopkins, and Suarez (1985), suggested that feedback paired with public posting, or some form of written feedback was most effective (i.e., produced positive effects in 53% of the reviewed studies), whereas isolated components of performance feedback were shown to be least effective (producing positive effects in only 28% of the reviewed studies). Fleming and Sulzer-Azaroff (1989) indicated that a "performance feedback package" including the provision of written instructions, demonstrations, and verbal feedback on-the-job, led to the greatest increases in targeted behaviors.

Some researchers have used feedback as a way of increasing the use of social skills within cooperative groups. They demonstrated that when students are taught social skills and reinforced for using them, their achievement significantly improves (D. W. Johnson & Johnson, 1989; L. J. Johnson, 1990; R. Johnson, 1976; Lew, 1986; Mesch, Lew, Johnson, & Johnson, 1986). Some other researchers have used feedback to individuals as a way to increase the use of cooperative skills within cooperative groups. Putnam, Rynders, Johnson, and Johnson (1989) demonstrated that, when students were taught social skills feedback is a very effective way to encourage them to use the skills more frequently.

The two forms of feedback are individual verses group. Archer-Kath (1994) compared the impact of individual feedback with the impact of group feedback on achievement, attitudes, and behavior in cooperative learning groups. Fifty-six American eighth-graders studying German were randomly assigned to conditions stratified for academic ability and knowledge of the German language and culture. They found that individual feedback was more effective than group feedback in increasing students' achievement motivation, actual achievement, uniformity of achievement among group members, and influence toward higher achievement within

cooperative learning groups. Individual, compared with group, feedback resulted in more positive relationships among group members and more positive attitudes toward German, the teacher, peers, and themselves (Archer-Kath, 1994).

The majority of studies of performance feedback have been conducted in organizational or institutional settings. Fewer studies have been conducted on the effectiveness of performance feedback to influence the behavior of teachers. Teachers are learners as well. Studies of learning show that feedback is essential for effective teaching and learning. Good teacher evaluation can support sound practice, influence teacher choices, and increase understanding about effective performance (Millman & Darling-Hammond, 1990). Feedback to teachers about their impact, merit, and value is important to career satisfaction (McLaughlin & Pfeifer, 1986). Cossart, Hall, and Hopkins (1973) evaluated the impact of performance feedback on the rate of praise statements delivered by teachers to students. In this study, an examiner recorded the rate of the teacher's praising statements and then provided feedback or feedback plus social praise to three elementary school teachers. The feedback and the feedback plus social praise resulted in increases in the praising behaviors of all three teachers. The authors noted that the increased rate of praising behaviors of the teachers improved overall student attention to the teacher. Moore and Schaut (1978) conducted a series of experiments to test whether or not performance feedback would influence the use and distribution of teacher attention. Feedback consisted of consultants providing teachers with verbal performance feedback on the teachers' behaviors directed toward reducing student inattention, and feedback on how to better distribute teacher attention to students considered to be more in need of assistance. The general findings indicated that performance feedback resulted in changes in teacher behavior (i.e., the

teachers became more adept at reducing the inattention of students) and that this improved the academic performance of the students in class (Moore & Schaut, 1978).

Feedback can help to improve student performance. Witt et al., (1997) demonstrated that daily performance feedback, provided to four elementary school teachers, resulted in increased treatment integrity and generally enhanced student performance. In this study, an academic intervention was task analyzed such that the completion of each step by a teacher resulted in the creation of a unique permanent product. After a baseline phase where teacher intervention implementation decreased markedly, the use of performance feedback by consultants increased intervention implementation by the teachers in the study.

In a follow-up study, Noell et al. (Noell, Witt, Gilbertson, Ranier, & Freeland, 1997)(1997) replicated the findings of Witt and his colleagues' study (Witt, Noell, LaFleur, & Mortenson, 1997) but provided the teachers with no pre-intervention training and did not provide the resources (i.e., intervention materials, reinforcers, etc.) to implement the intervention. They obtained similar results suggesting that performance feedback alone was effective in improving teacher implementation of a prereferral intervention.

2+2 is designed to provide more feedback to teachers. "Teachers receive multiple feedback on their work and are able to gain an appreciation for innovative and diverse approaches used by other teachers" (Beerens, 2000). In the context of the 2+2 teacher performance appraisal system, feedback is the two complements and the two suggestions teachers give each other following each 2+2 classroom observation, which is meant to help teachers to reflect on what they have performed in class. Lambert (1998) holds up reflective practice as a crucial aspect of building high capacity into schools and emphasizes the need to give appropriate time for such

activities. "Learning is a consequence of thinking. Retention, understanding, and the active use of knowledge can be brought about only by learning experiences in which learners think about and think with what they are learning" (Perkins, 1992).

People need feedback to learn. Teachers need 2+2 to have more opportunities for feedback. Feedback is most valuable when people have the opportunity to use it to revise their thinking as they are working on a unit or project. They learn to value opportunities to revise (Black & Wiliam, 1998).

Teachers' Professional Development

Professional development is considered as a sign of a quality teacher since it reflects an ongoing effort to improve practice and to keep teachers current (Pelletier, 1995). Such practices can include participation in teacher workshops, special training, additional college course or advanced degrees, frequent participation in inservice meetings, as well as being a member of teachers organizations, networks, or unions (Pelletier, 1995).

However, the existence of the variables mentioned above does not guarantee that a teacher will be successful. A deficiency in one area can not always be compensated by strength in another area (Mehrens & Green, 1986). For example, if a teacher has all of the other fine characteristics except for adequate classroom management skills, all the variables mentioned above will be ineffective (Papanastasiou, 1999).

Traditional approach to teachers' professional development has formal courses and inservice seminars as the central components which are like a voice coach giving advice to a singer whom he or she has never heard sing (Eisner, 1992). Teachers are not often consulted on what type of assistance they need, adding to perceptions that professional development is a waste of time.

A number of descriptions of professional development have emerged during the last decade or so as generic guides for teacher change processes. Some of the definitions include: “professional development is defined as those processes and activities designed to enhance the professional knowledge, skills, and attitudes of educators so that they might, in turn, improve the learning of students. In some cases, it also involves learning how to redesign educational structures and cultures” (Guskey, 2000). “Professional development signifies any activity that develops an individual’s skills, knowledge, expertise and other characteristics as a teacher. These include personal study and reflection as well as formal courses” (Center for Educational Research and Innovation, 1998). Hargreaves and Fullan (1992) see teacher professional development as knowledge and skill development, as development in self-understanding, and also as social change. They emphasize the influential and determining dimensions of the person, the social group and the context in the process of teacher change. Some others, such as Joyce and Showers (Joyce & Showers, 1988) suggest that teachers’ professional development is most effective when it is examined in terms of individual needs, the needs of schools and systems, the particular learning programs in place, and the students, their needs, abilities and characteristics. In Guskey’s model specific elements which encourage and support teacher change were focused (Guskey, 1998). The model proposes that teacher change will occur if teachers are encouraged to alter their classroom practice, and see and experience improvement in student learning outcomes. Still others have attempted to isolate key characteristics that can be components of effective teachers’ professional programs. Loucks-Horsley (1987) listed ten features including collaboration and collegiality, time to work on staff development and assimilate new learning, and the incorporation of experimentation and risk taking.

Horn (1999) examined the experiences of mid-career teachers who were selected as Fellows in the Stokely Institute for Liberal Arts Education, an interdisciplinary residential program in Louisiana State University. He did a qualitative study to understand how and what changes occurred in the professional and personal lives of Stokely Fellows. Specifically, the research asked these essential questions: (1) what changes were derived by participants of the Stokely experience; (2) was there anything about the Stokely experience that facilitated the emergence of such change; and (3) what personal and professional effects did the Stokely experience have in the lives of the participants? 20 participants were interviewed, and they all reported that the experience was supportive, challenging, open, and integrative. Some participants began with a sense of adventure; others came with a sense of uncertainty that often became one of intimidation. The participants were impressed by the breadth of seminar content. The initial uncertainty was transformed to personal challenge and commitment through collegiality, collaborative interactions, and supportive conditions. Those who experienced the transformation to openness attained levels of interactions necessary for further transitions to occur. The participants viewed themselves as being treated as valued professionals within a safe, diversion-free environment in which they were provided with the latitude to explore personal meanings related to many novel concepts and ideas.

Clement and Vandenberghe (2001) analyzed the impact of two workplace conditions, autonomy and collegiality, on elementary school teachers' professional development. They developed a substantial theory accounting for the relation between primary school teachers' autonomy and collegiality and its impact on their professional development. They found that tension between autonomy and collegiality creates different kinds of learning opportunities that are essential for teachers'

professional development. Learning opportunities can have a formal or informal character. They can originate during in-service activities, but also during a school day when for instance an experienced teacher counsels a novice or when a teacher tries to find a solution for a problem he or she is confronting (Clement & Vandenberghe, 2001).

Constructivism also transforms the ways in which professional development activities are structured and facilitated in many preservice and inservice settings (Rock & Levin, 2002). Constructivists hold that teachers should actively pursue their own questions, building upon their own knowledge base, and interactions within a social environment. Darling-Hammond and McLaughlin (1995) think that effective professional development means that teachers must be provided with opportunities to reflect critically on their practice to construct new knowledge and beliefs about content, pedagogy, and learners. "Teachers must be given ample opportunities to learn in constructivist settings and construct for themselves educational visions through which they can reflect on educational practices" (Brooks & Brooks, 1999).

Professional development should help teachers to think and reason about their teaching role (Shulman, 1987). Sparks and Hirsh (1997) recommend "activities such as action research, conversations with peers about the beliefs and assumptions that guide their instruction, and reflective practices such as journal keeping". Teachers should become an active part of professional development strategies in field settings. Lieberman (1995) asserts that if teachers are given opportunities to discuss, think about, try out, and hone new practices, their new role as a teacher action researcher will become not just a professional development activity with a life span of one or two days, but a part of their role and vision of what they do as a professional.

A number of underlying themes are identifiable across all of these models and descriptions. First, the success of teachers' professional development depends very much on the extent to which teachers' personal change is developed in their skills, knowledge, expertise and other characteristics as a teacher. Second, it is very crucial for teachers to review and reflect upon their own practices and beliefs. Third, opportunities to engage in professional dialogue with colleagues in similar situations, facing similar challenges, can provide encouragement, support and a new school culture of collaboration. Fourth, it is both simple and effective to have teachers visit each other's class and share with each other experiences of teaching. Finally, teachers should be provided with time, space and opportunity to experiment with new ideas and to reflect upon their experiences.

2+2 was proposed to become part of the in-service professional development of schools. The ultimate aim of the 2+2 program is to help teachers to grow professionally and improve their instructional practices.

Previous Evaluation of the 2+2 Program

Since 1981, most schools in China have gone through changes and attempted reform efforts in some way to improve their service for students. As the country has been striving to open up to the outside world socially and economically, educators in the country look to the developed countries for new ideas and methods that can be applied to the Chinese context (Shanxi Research Center for Secondary Education, 1997). The 2+2 program is one system that some educators in Shanxi learned from the then chief technical adviser of the World Bank educational project in China and found it useful to the country's educational system.

An evaluation conducted by Alyce LeBlanc employing a utilization focused evaluation design with a formative emphasis. The evaluation revealed that the 2+2

program implementation achieved almost undisputed recognition among participating teachers. In spite of minor implementation concerns and minimal active administrative support, teachers' enthusiasm and numbers of 2+2 observations increased over the course of the school year (LeBlanc, 1997a). Developed by Patton (1987), "Utilized evaluation is not a formal model or recipe for how to conduct evaluative research. Rather, it is an approach, an orientation, and a set of options. The active-reactive-adaptive evaluator chooses from among these options as he or she works with decision makers and information users throughout the evaluation process" (p. 284). In evaluating the 2+2 program implemented in Norfolk public schools, the "Teacher Performance Evaluation Attitude Questionnaire" was administered to measure attitudes toward the 2+2 protocol. Interviews were conducted "...to gather information regarding teachers' expectations of the program, and their reasons for participation. They were also asked how they felt about the Norfolk Public Schools teacher appraisal system, how they felt about being observed, how they anticipated 2+2 would impact their instruction, and if they intended to ask for student 2+2 feedback" (p. 54). Focus groups were utilized to investigate program implementation. All completed 2+2 observation forms were collected and analyzed (LeBlanc, 1997a).

Evaluation data indicated that 2 + 2 was viewed very favorably in many positive ways. Teachers felt overwhelmingly that 2 + 2 helped them share expertise, overcome isolation and expand their organizational perspective, and that it led to professional growth. Teachers preferred participation in the 2 + 2 program to evaluation under the traditional Norfolk Public School (NPS) appraisal system. Most teachers also thought the 2 + 2 appraisal program was superior to the NPS appraisal system (LeBlanc, 1997a).

Improvement of instructional practice was reported by participating teachers while being interviewed. In answering the question “Why did you opt for 2+2 this year instead of the Norfolk Public Schools’ appraisal system?” Seventy-seven percent responded positively indicating 2+2 “will improve instruction,” “expect it to be interesting, easy, convenient,” and “a positive way to grow professionally. ” “However, teachers cited improvement of instruction many times as evidence of the value of 2+2 and were overwhelmingly positive about the usefulness of 2+2 feedback in relation to improvement of instruction. Teachers did perceive that gaining multiple perspectives on instruction led to improvement in their own teaching” (LeBlanc, 1997a).

The evaluation data indicated that teacher isolation was alleviated. Seventy-two of the teachers indicated the quality of interaction among their colleagues had improved as a result of implementing 2+2. “As indicated by the evaluation data, the 2+2 program provided a powerful and effective mechanism to alleviate teacher isolation. The reduction of teacher isolation, one of the stated objectives of the 2+2 program, is critical to teacher learning and collaboration. As teacher isolation dissipates, more and different kinds of dialogue among teachers can emerge and lead to increased teacher collaboration” (LeBlanc, 1997a).

Teachers’ reported anxiety about being observed and evaluated was reduced. Teachers who experienced the 2+2 process revealed less discomfort with or distrust of an evaluation situation. Teachers felt that 2+2 was designed to help them focus and improve their teaching performance (LeBlanc, 1997a).

More administrative support was proposed. Administrators must perform a crucial role in the development of the 2+2 program. A school administrator must make more frequent visits, and department heads at least four 2+2 visits each

semester. School administrators and department heads must work with the building coordinator to support the 2+2 program. Teachers must be supported in risk-taking behaviors, in creating change, and in accepting increased accountability (LeBlanc, 1997a).

2+2 was designed to provide more opportunities for teachers to offer and receive instructional feedback to one another and help them to combat isolation in their workplace. The system features many of the characteristics and components of peer coaching, conventional teacher evaluation, and teacher collaboration which have been studied by many researchers.

However, very limited literature was discovered related with its application. Only one empirical evaluation has been conducted based entirely on self-report perceptions. Consequently, it is difficult to draw any conclusion about its effectiveness. The current study investigated how the program has been implemented in Shanxi province and would provide more objective data about the effectiveness or ineffectiveness of the program and contribute to the related literature.

CHAPTER III

METHODOLOGY

This chapter presents the procedures utilized to evaluate the implementation and outcomes of the 2+2 program in Shanxi Province, China. It also addresses the research plan, research questions, research design, data collection, data analysis, the role of the researcher, and limitations.

Research Plan

The purpose of this study was to examine the effectiveness of the 2+2 Alternative Teacher Performance Appraisal System that has been implemented as one of the educational reform efforts in Shanxi province in China from its inception through to its completion. The whole study went through three stages.

The first stage of the evaluation effort that took place September 2001 was to help the participating schools produce a working plan to implement the 2+2 program according to the program guidelines at the school level. Activities were designed for individual teachers to apply the 2+2 system in their classrooms. A full review of the program documents was constructed and the goals and objectives of the program were clarified. All participants and stakeholders were expected to discuss what would be achieved as results of the 2+2 program. Interviews were conducted to inquire about how the participating teachers perceive the 2+2 system and how they can improve their instructional practices and change the school culture by implementing the program.

The second stage of the evaluation process was a pretest that was conducted in September 2001 to investigate the participating teachers' pedagogical beliefs, attitudes and their professional performance level, and to collect the baseline information about the

program. A Teachers' Pedagogical Belief and Attitude Scale was administered as a pretest, and a Professional Performance Measurement Scale was utilized to evaluate their professional performance before the program implementation.

In the third stage of the evaluation, qualitative and quantitative methods were used to address the research questions. Outcome data were collected to investigate the effectiveness of the program. The treatment group was compared with the comparison group in terms of their scores on the Pedagogical Belief and Attitude Scale and on the Professional Performance Measurement Scale.

Research Questions

Seven research questions were generated for the current study:

1. How effective is the program in helping teachers to improve their professional performance?
 - a. How does the number of 2+2 visitations predict change in teachers' professional performance?
 - b. How does the amount of 2+2 feedback predict the change in teacher's professional performance?
2. How has the program influenced teachers' pedagogical belief and attitude?
 - a. How does the number of 2+2 visitations predict change in teachers' beliefs and attitudes?
 - b. How does the amount of 2+2 feedback received predict the change in teacher's beliefs and attitudes?
3. Does the frequency of feedback provided differ between teachers in the treatment group and those in the comparison group?

4. Do teachers in the treatment group collaborate more in their teaching practice than those in the comparison group?
5. What kind of feedback is provided to teachers who participate in the 2+2 program?
6. What are teachers' perceptions of the advantages and disadvantages of the 2+2 system?
7. What suggestions do teachers provide for improving the 2+2 program?

There was insufficient empirical support for formal hypotheses; however, if the 2+2 program is effective, it was expected that the teachers participating in the 2+2 program would be more likely to have better professional performance. It also was expected that the implementation of the 2+2 program would contribute to the change of the teachers' pedagogical beliefs and attitudes. The participating 2+2 teachers were expected to provide more feedback and to collaborate more in their teaching practice than before participating in the program. Moreover, it was expected that the participating 2+2 teachers would get adequate feedback from their peers that they perceive as helpful for improvement of their instructional practice, and have positive perceptions of the 2+2 system. Teachers in the 2+2 group would list more advantages than disadvantages of the system and provide a variety of suggestions for improving the 2+2 program.

Research Design

The evaluation of the 2+2 Teacher Performance Appraisal System was designed as a utilization-focused program evaluation. According to Patton (1997): "Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming. Utilization-focused

program evaluation (as opposed to program evaluation in general) is evaluation done for and with specific, intended primary users for specific, intended uses (p.2).” The methods utilized to investigate the research questions included both quantitative and qualitative inquiries. The data collection methods varied according to different research questions (Table 1).

Table 1: Research Questions and Research Design Table

Research questions	Research design	Indicators	Data collection
1. How effective is the program in helping teachers to improve their professional performance?	Quasi-experimental Non-equivalent control group design	Teacher Professional Performance scores	Teacher Performance Scale Questionnaire: Pretest and posttest on both treatment group and comparison group
2. How has the program influenced teachers' pedagogical beliefs and attitudes?	Quasi-experimental Non-equivalent control group design	Teacher's Pedagogical Belief and Attitude scores	Teacher's Pedagogical Belief and Attitude Test: Pretest and posttest on both treatment group and comparison group
3. Does the frequency of feedback provided differ between teachers in the treatment group and those in the comparison group?	Quasi-experimental Non-equivalent control group design	Amount of feedback that provided by teachers	2+2 Program Response Survey: Pretest and posttest on both treatment group and comparison group
4. Do teachers in the treatment group collaborate more in their teaching practice than those in the comparison group?	Quasi-experimental Non-equivalent control group design	Teacher's peer interaction and collaboration scale score	2+2 Program Response Survey: Pretest and posttest on both treatment group and comparison group
5. What kind of feedback is provided to teachers who participate in the 2+2 program?	Cross-sectional study	Content of the 2+2 feedback form	2+2 Feedback Form
6. What are teachers' perceptions of the advantages and disadvantages of the 2+2 system?	Cross-sectional study	Teachers' perceptions	Interview
7. What suggestions do teachers provide for improving the 2+2 program?	Cross-sectional study	Teachers' suggestions	Interview

For research questions comparing the teachers' professional performance, pedagogical beliefs and attitudes, frequency of feedback and teacher collaboration before and after participating the 2+2 program, a quasi-experimental non-equivalent groups design is employed, in which six key urban high schools were selected and randomly assigned to either the treatment group or the comparison group. The non-equivalent groups design (NEGD) is probably the most frequently used design in social research (Compbell and Standley, 1960), which was structured like a pretest-posttest randomized experiment.

For study on the type of feedback, teachers' perceptions of and suggestions for the 2+2 program, there is no comparison group and pretest involved, therefore descriptive cross-sectional analysis was employed and qualitative research is preferred. A qualitative approach is well suited for examining the process of the program and the meaning of how people make sense of their experiences (Denzin & Lincoln, 1998). According to Patton (Patton, 1987), three methods of qualitative research, detailed but open-ended interviews, direct observation, written documents (work with words and visual data, not numbers) will complement the quantitative data in the following ways. First, it can provide the depth and detail, which is not provided in a standardized questionnaire. Second, it can generate new theories and recognize phenomena ignored by most or all previous researchers and literature. Third, it helps people see the worldview of those studied--their categories, rather than imposing the researchers' preconceived categories. Last, it attempts to avoid pre-judgments for its goal is to try to capture what is happening without being judgmental, and try to represent participants' perspectives so reader can see their views.

Data Collection

Setting

The setting of this study was Shanxi province, People's Republic of China. Shanxi province, located in the northwestern part of China, has a population of about 30 million people, of which over 5 million are receiving primary and secondary education. There are 9,988 schools located in urban and rural areas in the province (Shanxi Education Commission, 1999). Five hundred and fifty-six of them are senior high schools. Currently, about 200,000 teachers are in service of the secondary education, of which about 10,000 are high school grade one teachers (Shanxi Education Commission, 1999). The high school sizes range from 300 students to 3,000 students with a mean of 1,668 (Shanxi Education Commission, 1999). There were 43 provincially nominated key high schools. In Shanxi those 43 provincially were nominated key schools because they all met the following requirements and standards set by the provincial government in 1983: (1) All teachers must have a bachelor degree; (2) The school must have an enrollment of about 600-800 students; (3) The school must have a decent school building that can provide enough room for its students; (4) The school must have standard science laboratories for all of its students; (5) There must be a sports ground in the school which includes a 400 meter track; (6) The achievement level of the students in the school must be the best among the schools in the county or city (Shanxi Education Commission, 1999).

Sample

Non-random sampling selection was employed. The School-Based 2+2 Program was designed in the year 2000 as a follow-up of the LTTP project. Six urban high schools

were selected by the Central Office for the Program's Implementation from among the 43 provincial key high schools to participate in the program: The Provincial Experimental High School in the city of Taiyuan, the No. 1 High School in the city of Xinzhou, The No.1 High School in the city of Jincheng, the No.1 High School in the city of Yuci, the No. 1 High School in the city of Datong, and Quwo High School in the city of Quwo. These schools were selected because they share some common characteristics in terms of their size, students' achievement level, and teachers' educational background. All of these six schools have a student population of about 2,000, which are very much like the other provincially nominated key schools.

The teachers in the program schools all had obtained a bachelor or equivalent degree that was required as a must by all of the key schools for their teacher position applicants. These six schools in this sample were representative of the 43 key schools in the province. Each of those six project schools has 13 first grade (equivalent to 10th grade in the United States) teachers -four Chinese language teachers, two math teachers, three English teachers, two physics teachers, one chemistry teacher, and one social science teacher. Each of these schools has one lead teacher on the first grade teaching staff. Among these six schools, the No. 1 High School in the city of Xinzhou, No.1 High School in the city of Jincheng and Quwo High School in the city of Quwo were randomly assigned to the treatment group to participate in the 2+2 program. There were 39 first grade teachers participating in the program. The Provincial Experimental High School in the city of Taiyuan, the No.1 High School in the city of Yuci, and the No. 1 High School in the city of Datong still maintain their traditional teacher evaluation and appraisal

system. There were additional 39 teachers participating as the comparison group.

Altogether there were 78 teacher participants in the current study.

There were 25 (32.1%) male teachers and 53 (67.9%) female teachers. Fifty six (71.8%) of them were forty years old or younger and 12 (28.2%) aged forty-one years or older. Forty (51.3%) had three years or less teaching experience, 19 (24.4%) had four to 10 years teaching experience, and another 19 (24.4%) had eleven years or more teaching experience.

Instrumentation

Three questionnaires were employed in this study. They were the Teacher Professional Performance Measurement Scale, Teachers' Pedagogical Belief and Attitude Scale and 2+2 Program Response Survey to measure teachers' classroom teaching performance, teachers' pedagogical beliefs and attitudes and teachers' responses to participating 2+2 program.

Teacher Professional Performance Measurement Scale. Teacher professional performance was defined as a teacher's demonstration of skills or competency in class. In the current evaluation teacher performance refers to what is measured by Shanxi Teachers' Performance Measurement Scale (Shanxi Research Center for Secondary Education, 1997). Shanxi Teacher Performance Measurement Scale was employed to evaluate the teacher professional performance. The scale was developed by a panel consisting of 10 educational experts from three teacher education institutions in Shanxi province in 1997 to determine the professional performance level of the LTTP candidates. It has been used by most of the school districts in Shanxi since then to appraise their teachers' professional competence. The emphasis is on teachers' ability to perform

instructional tasks. Based on the pilot use of the scale, a review meeting of the same 10 educational experts who developed the scale was held in summer 1997, and several minor modifications were made to address its content validity considering the relevance of the elements measured in the scale. This scale comprises thirty nine items and provides assessment of nine major functions of teaching: preparedness for instruction (2 items), management of instruction time (3 items), management of student behavior (5 items), instructional presentation (11 items), monitoring of student performance (3 items), providing reinforcement and feedback (5 items), facilitating instruction (5 items), communicating with students (2 items), and chalk board skill (3 items) (Shanxi Research Center for Secondary Education, 1997). Evaluators rate all the functions on 6 points scale: unsatisfactory (1), below standard (2), at standard (3), above standard (4), well above standard (5), and superior (6). The evaluators were also encouraged to add comments at the end of each major function. This scale is now used in Shanxi as a standard to measure secondary educational teachers' professional performance. The complete Shanxi Teacher Professional Performance Measurement Scale is presented in Appendix A.

The scale was first utilized in five different high schools in May 1997. Then in July 1997, the scale was administered in the same group of schools for the second time. The internal consistency of sub-scales in Teacher Professional Performance Measurement Scale has been also calculated. The Cronbach's Alphas of reliabilities established among the sections of the scale were: preparedness for instruction, 0.91; management of instruction time, 0.91; management of student behavior, 0.90; instructional presentation, 0.89; monitoring of student performance, 0.90; providing reinforcement and feedback,

0.91; facilitating instruction, 0.88; communicating with students, 0.92; and chalk board skill, 0.90.

Teachers' Pedagogical Belief and Attitude Scale: Teacher's pedagogical belief and attitude was measured by Shanxi Teachers' Pedagogical Belief and Attitude Scale (Shanxi Research Center for Secondary Education, 1997). The scale was developed by the same panel who produced Shanxi Teacher Professional Performance Measurement Scale. The content of these items were evolved from Teacher Attitude Inventory developed by Glascock (1996), and Teaching Behaviors Questionnaire created by Marchant and Bowers (1988). One hundred and fifty teachers from 5 schools were tested on the scale. The Cronbach's Alpha of internal consistency is 0.86.

Shanxi Teachers' Pedagogical Belief and Attitude Scale has been used by most of the school districts in the province since then to test job applicants for teaching positions. The development and modification process was the same as that of Shanxi Teachers' Performance Measurement Scale. This scale is now in use as a standard to measure secondary educational teachers' pedagogical belief and attitude. The scale comprises thirty Likert type questions about pedagogical belief and attitude. Respondents rate all the functions on a 5-point scale: strongly disagree (1), disagree (2), undecided (3), agree (4), and strongly agree (5).

Chinese evaluators had different considerations when inventing those assessment instruments. They tended to include measurements of attitudes in some of the pedagogical belief and attitude scales. Thus in the current pedagogical belief and attitude scale, ten of those attitude related items were excluded in analyzing the participants' pedagogical belief and attitude, while twenty other items that were closely related to

pedagogical belief and attitude were selected for study. The scale with the 20 items used to assess pedagogical belief and attitude can be found in Appendix B.

2+2 Program Response Survey: the 2+2 Program Response Survey was developed by the researcher based on the 2+2 survey created by LeBlanc in 1997. This instrument was developed to investigate how the 2+2 program has been implemented and how the participating teachers perceive the program. The 2+2 Program Responses survey is distributed to gather participant responses on five aspects of 2 + 2: (a) information about observation and feedback (8 open-ended questions), (c) assessment of the 2+2 program observation feedback form (2 items scale), (d) perception about participating in the 2+2 program observation (6 items, yes/ no questions), and (e) peer interaction and collaboration (5 open-ended question). A revised version of the 2+2 Response Survey was administered to the teachers in the comparison group, with items concerning 2+2 assessment of feedback forms and perception about participating 2+2 observation excluded. Information regarding observation and feedback consisted of eight open-ended questions about the frequency of teacher observing or being observed, and giving feedback or receiving feedback. Assessment of the 2+2 program observation feedback form consisted of two 5-point Likert scale questions to inquire about the quality of the compliments and the suggestions. Perception about participating in the 2+2 program consisted of six “yes/no” questions to solicit teacher’s feeling of participating in the 2+2 program. The last category of the 2+2 Program Response Survey was to explore teacher collaboration and isolation. Five open-ended questions were employed to inquire about the frequency of interaction and collaboration activities experienced by the teachers.

Reliability of this survey was determined by administering the instrument in six in-service teacher-training sessions that took place in the summer of 2000 and 2001 in Shanxi province. All these sessions included the 2+2 system as one of the training components. An internal consistency coefficient (Cronbach's Alpha) was computed for each scale: assessment of the 2+2 program observation feedback form, 0.92; and perception about participating 2+2 program observation, 0.94. Minor modifications to the survey were made by a panel of eight Chinese secondary education experts so that the survey would be more suitable for the situation in Shanxi province. The survey is shown as Appendix C.

Interview

Structured interviews were conducted in winter, 2002, with 39 participants of the 2 + 2 program. The purpose of the interview was to gather complementary information regarding teachers' perceptions, expectations and evaluation of the program. The interview protocol consisted of ten questions. The participants were asked how they experienced and thought about Shanxi teacher appraisal system, how they react when being observed, how they anticipate 2 + 2 would impact their instruction, and their perception of collaboration and isolation. The following questions were developed for the interview: (1) How do you describe the implementation of the 2+2 program? (2) How do you compare "2+2" with the traditional teacher performance appraisal system? (3) What were the strengths of "2+2"? (4) What were the weaknesses of "2+2"? (5) What progress have you made with "2+2"? (6) How did you benefit from it? (7) What were your reactions to being observed? (8) What were your reactions to being an observer? (9)

What is the next step in your plan for “2+2”? (10) What else can you tell me about 2+2 that would help me understand its impact?

Procedure

Implementation of the 2+2 Program

Prior to September 15, 2001, the 39 participating teachers in the treatment group were given an opportunity to sign up for 2+2 in lieu of the conventional evaluation procedure, and all of them agreed to participate. The school administration arranged for a school-wide system to track observations by the treatment group teachers. The direct supervisor or designee periodically reviewed the records to ensure that each subject was receiving and giving an appropriate number of 2+2 visitation. The treatment group teachers visited another teacher in the same group at any time when instruction is given, usually without prior notice. A visit may take place any time during the class period. Without acknowledgement from the observee, the observer took a seat in the classroom and observed for as long as it took to formulate two compliments and two suggestions for improvement. The visit lasted a minimum of ten minutes, usually more. The forms were left in a pre-determined location in the classroom at the conclusion of the visit or no later than the end of that school day. The observer was instructed to keep a copy of each form and a copy was given to the direct supervisor. A minimum of 40 visitations were to be conducted per semester. Post-visitation conferences were not a requirement of the 2+2 system. The teachers were encouraged, however, to discuss any questions resulting from 2+2 visitation or feedback and to focus different aspects of his or her colleague's teaching in different perspectives and aspects each time. The number of 2+2 visitation was recorded reported by the teacher and recorded by the program manager.

A 2+2 portfolio was maintained and included all of the completed feedback forms received by the teachers. For the current study, the 2+2 feedback form (see Appendix D) was the record of the feedback from the teachers and administrators on classroom teaching during the 2+2 program process. Whenever any administrator or resource person visited a classroom instructed by a 2+2 teacher during instructional time for more than ten minutes, the visitor left a feedback form at the conclusion of the visit or no later than the end of that school day. The observer kept a copy of each form. The program manager was responsible for keeping all of the original records. The 2+2 feedback forms had been distributed to the teachers and administrators at the three high schools of the treatment group since September 2001, and collected by the one lead teacher in each school. Overall, 3314 feedback forms have been collected by the end of program implementation.

Administration of Measures

The pedagogical belief and attitude pretest was conducted in September 2001, and the posttest was administered in October 2002. The participants' professional performance was measured both prior to and after the implementation of the program. The central office of the program hired five professional evaluators to observe the participants' classroom teaching and evaluate their performance level. The questionnaires were distributed to the 78 teachers in the six schools by mail. The 2+2 Response Survey was only completed by teachers who were in the treatment group, i.e. who participated 2+2 program. The principal of each school was responsible for collecting the completed questionnaires and return them to the program manager. All the questionnaires were kept in sealed envelopes.

Interviews were conducted by the current researcher with his assistants in the three high schools in the treatment group when the researcher of the current study returned to Shanxi in Fall 2002. The researcher arranged three meetings with the participating teachers, one meeting in each of the three treatment group schools at the conclusion of the program. The teachers in each school discussed their experience of implementing 2+2 program, using a semi-structured interview protocol (Appendix E). The duration of the focus group interviews ranged from two to three hours. These interviews were audio-taped and transcribed.

Data Analysis

Quantitative Data

Non-Equivalent Groups Design (NEGD) was employed to address the research questions of whether the 2+2 program improved teachers' professional performance, pedagogical belief and attitude, frequency of feedback and the teachers' collaboration practices. NEGD involves a treatment group, a comparison group, and each variable is measured before and after implementation of the 2+2 system. The statistical model would involve pretest variables, posttest variables, and a group variable that designates group membership. Descriptive analysis was used to examine the frequencies, distribution, central tendency and dispersion for each of the variables.

Correlation analysis showed that there was a non-significant weak correlation ($r^2=0.02$, $p>.05$) between belief and attitude and performance, therefore MANCOVA is not the best test for this study because it would violate the linearity assumption for using MANCOVA (Tabachnick & Fidell, 2001). Analysis of Covariance (ANCOVA) was employed to compare posttest scores of the treatment group and comparison group

controlling for the pretest scores. The main independent variable was the group membership at the nominal measurement level (treatment or comparison group), while the dependent variables were teacher performance scores (interval), teacher pedagogical belief and attitude scores (interval), frequency of feedback (ratio), and frequency of teacher collaboration practice (ratio). A data analysis matrix for variables by measurement level and statistical test used for above research questions is shown in Table 2.

Qualitative Data

Content analysis was employed to analyze the compliments and suggestions the teachers had provided on the 2+2 observation forms. Purposive sampling is used to draw sample from the 3314 collected forms. Altogether 350 forms were selected by teachers' teaching major, year of teaching and gender. Patton (1987) defined content analysis as "the process of identifying, coding, and categorizing the primary patterns in the data" (p.381). Classifying the data is a preliminary step in analyzing content, and facilitates the search for patterns and themes within a particular setting or across cases (Patton, 1987). According to Patton (1987), establishing a classification system can be a "simple filing system," a way to index the data by devising appropriate labels for different ideas represented in the data.

For the analysis of the 2 + 2 feedback forms, a process of categorizing, or labeling, 2 + 2 compliments and suggestions across cases were utilized. Compliments and suggestions were analyzed separately. Teachers' 2 + 2 feedback forms were examined, and compliments/suggestions were tentatively assigned to a category. As compliments/suggestions were found unfit in a category, a new category or sub-category

were created. Categories were revised, as compliments/suggestions were reviewed and assigned to categories in an iterative back and forth process.

Content analysis was also used to analyze the open-ended interview and questionnaire data. The interview and questionnaire provided supplemental information on various aspects of 2+2 program and teachers' perceptions of collaboration and isolation. Individual responses to each interview question were examined, compared, and coded. The coding process itself was a "cut and paste" iterative process whereby conceptually similar responses were grouped into categories. Thus, responses from different teachers to each question were grouped together under categories that emerged from the distribution of the responses themselves after thorough reviews of the data.

Table 2: Data Analysis Matrix

Research questions	Research design	Independent Variables	Dependent Variable/indicator	Data Analysis
1. How effective is the program in helping teachers to improve their professional performance?	Quasi-experimental Non-equivalent control group design	Group (treatment Vs. Control) Nominal level	Teacher Performance Scale score	Total score- ANCOVA controlling for pretest score
2. How has the program influenced teachers' pedagogical beliefs and attitudes?	Quasi-experimental Non-equivalent control group design	Group (treatment Vs. Control) Nominal level	Teacher's Pedagogical Belief and Attitude Test score	Total score- ANCOVA controlling for pretest scores
3. Does the frequency of feedback provided differ between teachers in the treatment group and those in the comparison group?	Quasi-experimental Non-equivalent control group design	Group (treatment Vs. Control) Nominal level	# and type of the feedback and provided by teachers	# of feedback- ANCOVA
4. Do teachers in the treatment group collaborate more in their teaching practice than those in the comparison group?	Quasi-experimental Non-equivalent control group design	Group (treatment Vs. Control) Nominal level	Teacher's Peer Interaction and Collaboration Scale score	Total score- ANCOVA controlling for pretest scores
5. What kind of feedback is provided to teachers who participate in the 2+2 program?	Cross-sectional study		Content of the 2+2 Feedback Form	Content analysis
6. What are teachers' perceptions of the advantages and disadvantages of the 2+2 system?	Cross-sectional study		Teachers' perceptions from survey responses	Content analysis
7. What suggestions do teachers provide for improving the 2+2 program?	Cross-sectional study		Teachers' suggestions from survey responses	Content analysis

Role of the Researcher

The involvement of the researcher in the program can be positive. The researcher has been a part of the LTTP/2+2 project since it began in 1997, and had good working relationships with many of the teachers participating in the 2 + 2 program. At various times the researcher facilitated seminars on transdisciplinary curriculum where issues of collaboration, teacher isolation and overload were frequently raised. With regard to the 2 + 2 program, the researcher developed the evaluability assessment, conducted orientation and training sessions, and monitored the implementation of 2 + 2. Because of this history, the researcher was very familiar with the school environments and with various positive and negative change forces operating within the context of the LTTP project and within the 2 + 2 program. Access to 2 + 2 participants and 2 + 2 observation forms by the researcher was freely given by participating schools.

CHAPTER IV

RESULTS

This chapter presents the results of the study. It consists of eight sections, corresponding to the research questions. Each section starts with a restatement of the research question followed by a description of the variables including independent variables, dependent variables and the data sources. Then the data analysis procedure is described. Results of inferential statistics to address the research question are then presented along with a description of descriptive results to help interpret findings.

Program Impact on Professional Performance

The first research question addressed whether teachers in the 2+2 program had more improvement in their professional development scores than did teachers in the comparison group. To better explore how 2+2 influences teacher's professional performance, two sub-questions were generated to address the impact of the major components of the 2+2 program, visitation and feedback. Only the data of teachers in the 2+2 group were analyzed to answer the two sub-questions. The questions are listed below:

- a. How does the number of 2+2 visitations predict change in teachers' professional performance?
- b. How does the amount of 2+2 feedback predict the change in teacher's professional performance?

Professional performance, the dependent variable was operationally defined as posttest scores on the teachers' professional performance scale, with the pretest scores on this same measure serving as the covariate. The independent variable was group membership (2+2 or comparison) for the general question. The number of classroom

visitations the teachers had made and the amount of feedback they had received were the independent variables in the follow-up analyses to examine the impact of the program components. The number of classroom visitations was measured by counting number of the visitations the teacher made in the 2+2 program, which was recorded by the program manager. The feedback was measured by counting the amount of feedback teachers received from their colleagues, which was a question in the 2+2 response survey.

Overall program Impact of 2+2

ANCOVA was used to analyze the data. The results revealed a significant difference between groups on the posttest total performance score while controlling for the pretest total performance scores ($p < 0.001$). The pretest total score of professional performance for the 2+2 group was 154.41 (SD=23.78) and the posttest score for this group was 185.14 (SD=25.28). The pretest score of professional performance for the comparison groups was 152.57 (SD=30.73) and the posttest score was 147.85 (SD=31.30).

ANCOVA tests on each of the nine functions also revealed that 2+2 group teachers had significantly higher posttest scores for most of the functions except the chalkboard skill while controlling for the pretest scores ($p < .05$). The Bonferroni adjustment was used to adjust the probability level for families of hypotheses, (i.e. the probability level for the nine comparisons on teachers' professional performance is $.05/9 = .0056$). After the adjustment, the differences remained statistically significant ($P < .0056$).

The mean scores on each of the nine functions were obtained by dividing the total scale score by the number of items on the scale. The descriptive statistics by subscale are

presented in Table 3. As is shown, the professional performance of the teachers in the 2+2 group had improved from “at standard” (3.97) to “above standard” (4.75) while that of the comparison group remained at “at standard” (3.91 to 3.79). The top three functions of teachers’ performance on the improvement list were monitoring of student performance, communicating with students, and facilitating instruction.

Impact of 2+2 program component on the teachers in the 2+2 Program

2+2 visitations and professional performance

Results show that 2+2 visitations were positively related to professional performance improvement for the teachers in the 2+2 group. The improvement of the teachers’ performance for the 2+2 group was measured by calculating the difference between the pretest and posttest total scores. The improvement ranged from -20 to 98 with a mean of 27.71 (SD=28.22). The total visitations completed by each individual of the 2+2 group teachers ranged from 80 to 118 with a mean of 84.97 (SD=7.48). Pearson’s correlation showed that there was a significant positive relationship between the improvement of teachers’ performance and the number of 2+2 visitations ($r^2=.35$, $r=.592$, $p<.01$). The more visitations a teacher had made, the more improvement had been found in his/her teacher performance.

2+2 feedback and professional performance

The frequency of feedback was positively related to the improvement of 2+2 group teachers’ professional performance. The total amount of feedback received by each individual of the 2+2 group teachers from their colleagues in the previous month ranged from 14 to 20 with a mean of 16.10 (SD=2.21). Pearson’s correlation showed that there was a significant but weak positive relationship between the improvement of teachers’

performance and the amount of 2+2 feedback ($r^2=.26$, $r=.512$, $p<.01$). The more feedback a teacher had received, the more improvement had been found in his/her professional performance.

Table 4, 5 and 6 further illuminate the descriptive findings pertaining teachers professional performance. After the program implementation, the scores for the 2+2 group's professional performance increased on all of the tested items except "rules and procedures for student participation". The percentage of those who performed above standard increased dramatically for teachers in the 2+2 group. An increase of 30% or more was discovered on many items.

Table 3: Comparison of the Teachers' Professional Performance

Function Total Score	Items	2+2 group			Comparison group		
		Pretest	Posttest	Change	Pretest	Posttest	Change
Preparedness for instruction**	2	8.15	9.48	1.33	7.66	7.02	-0.64
Mean Score		4.08	4.74	0.67	3.83	3.51	-0.32
Management of instruction time**	3	13.10	14.61	1.51	11.95	11.66	-0.29
Mean Score		4.37	4.87	0.50	3.98	3.89	-0.10
Management of student behavior**	5	19.84	23.00	3.16	18.41	17.84	-0.57
Mean Score		3.97	4.60	0.63	3.68	3.57	-0.11
Instructional presentation*	11	43.77	52.08	8.31	44.56	43.59	-0.97
Mean Score		3.98	4.73	0.76	4.05	3.96	-0.09
Monitoring of student performance**	3	11.33	14.43	3.10	11.46	10.64	-0.82
Mean Score		3.78	4.81	1.03	3.82	3.55	-0.27
Providing reinforcement and feedback**	5	18.97	23.10	4.13	19.23	18.39	-0.84
Mean Score		3.79	4.62	0.83	3.85	3.68	-0.17
Facilitating instruction**	5	18.23	22.47	4.24	18.12	17.56	-0.56
Mean Score		3.65	4.49	0.85	3.62	3.51	-0.11
Communicating with students**	2	8.02	9.89	1.87	8.28	7.97	-0.31
Mean Score		4.01	4.95	0.94	4.14	3.99	-0.16
Chalk board skill	3	11.94	14.29	2.35	13.10	13.15	0.05
Mean Score		3.98	4.76	0.78	4.37	4.38	0.02
Total **	39	154.41	185.14	30.74	152.57	147.85	-4.72
Mean Score ^a		3.96	4.75	0.79	3.91	3.79	-0.12

^a Average score (based on the Professional Performance Scale Criteria):

- | | | |
|-------------------|------------------------|----------------|
| 1. Unsatisfactory | 2. Below Standard | 3. At Standard |
| 4. Above Standard | 5. Well Above Standard | 6. Superior |

*p<.01

**p<.001

The score for “starting class on time” in “management of instruction time” increased most dramatically (46.1%). Items on “management of student behaviors” that increased 30% were “having rules and procedures for student movement” (30.8%), and “stopping inappropriate student behaviors” (32.7%). The only item that has showed a decrease of all the professional performance was “having rules and procedures for student participation”, but the magnitude was very small (-0.6%). Among “the instructional presentation items”, appropriate presentation pacing (38.5%), clear assignment (35.9%), and appropriate transitions (30.8%) were the three functions that increased most. “Monitoring the student performance” was the function that has the highest increase. Two of the three items, “having appropriate performance standard” (56.4%) and “routinely checking students’ progress” (41.1%), increased more than 40%. “Having prompt feedback” (38%) was the only item on “providing reinforcement and feedback” that increased more than 30%. “Use diagnostic information” (56.4%) and “encouraging independent thinking and working” (41%) stood out of the five items on “facilitating instruction”, while “maintaining accurate records” increased 30.9%. Both of the two items on “communication with students, fair and equitable” (30.7%) and “effective interaction” (33.4%) were among those that increased 30% or more. “Making main points on board” increased the highest (34.2%) on “the chalkboard skill” (See Tables 4-6).

Table 4: Comparison of Preparedness for Instruction, Management of Instruction Time, and Management of Student Behavior

Function	2+2 group			Comparison group		
	Pretest	Posttest	Change	Pretest	Posttest	Change
Preparedness for instruction						
Prepare materials	59.0	79.5	20.5	56.4	35.9	-20.5
Instructional plan	69.5	89.7	20.2	56.4	38.5	-17.9
Management of instruction time						
Start class on time	43.6	89.7	46.1	33.3	38.5	5.2
Get students on task quickly	82.1	89.7	7.6	69.2	53.8	-15.4
Maintain a high level of students time-on-task	79.5	84.6	5.1	64.1	30.8	-33.3
Management of student behavior						
Rules and procedures for routine administer matters	59.0	87.2	28.2	48.7	38.5	-10.2
Rules and procedures for student participation	76.9	76.3	-0.6	61.5	53.8	-7.7
Rules and procedures for student movement	53.8	84.6	30.8	43.6	30.8	-12.8
Frequently monitor the student behaviors	66.7	87.2	20.5	51.3	41.0	-10.3
Stop inappropriate behaviors	43.6	76.3	32.7	30.8	41.0	10.2

* Mean Score >4

Table 5: Comparison of Instructional Presentation and Monitoring of Student Performance

Function	2+2 group			Comparison group		
	Pretest	Posttest	Change	Pretest	Posttest	Change
Instructional presentation						
Attract students' attention	69.2	94.9	25.7	61.5	59.0	-2.5
Appropriate introductions	59.0	82.1	23.1	56.4	69.2	12.8
Speak fluently and precisely	66.7	94.9	28.2	76.9	74.4	-2.5
Appropriate concept and language	69.2	89.7	20.5	71.1	74.4	3.3
Relevant examples and demonstrations	71.8	84.6	12.8	66.7	48.7	-18
Appropriate task assignment	74.4	87.2	12.8	61.5	43.6	-17.9
Appropriate questions	71.8	87.2	15.4	71.8	61.5	-10.3
Appropriate presentation pacing	56.4	94.9	38.5	64.1	57.9	-6.2
Appropriate transitions	53.8	84.6	30.8	46.2	43.6	-2.6
Clear assignment	56.4	92.3	35.9	41.0	35.9	-5.1
Appropriate summary	48.7	76.9	28.2	46.2	48.7	2.5
Monitoring of student performance						
Appropriate performance standard	41.0	97.4	56.4	64.1	51.3	-12.8
Move around to check students' performance	66.7	87.2	20.5	48.7	33.3	-15.4
Routinely check students' progress	53.8	94.9	41.1	48.7	41.0	-7.7

* Mean Score >4

Table 6: Comparison of Providing Reinforcement and Feedback, Facilitating Instruction, Communicating with Students and Chalk Board Skill

Function	2+2 group			Comparison group		
	Pretest	Posttest	Change	Pretest	Posttest	Change
Providing reinforcement and feedback						
Providing feedback	59.0	82.1	23.1	56.4	53.8	-2.6
Prompt feedback	46.2	84.2	38	41.0	43.6	2.6
Appropriate affirmation	69.2	94.9	25.7	64.1	59.0	-5.1
Sustaining feedback	53.8	82.1	28.3	56.4	53.8	-2.6
Positive feedback	59.0	84.6	25.6	66.7	51.3	-15.4
Facilitating instruction						
Encourage independent thinking and working	51.3	92.3	41	69.2	38.5	-30.7
Use diagnostic information	35.9	92.3	56.4	51.3	51.3	0
Maintain accurate records	38.5	69.4	30.9	28.2	21.1	-7.1
Match students needs	48.7	69.2	20.5	30.8	23.1	-7.7
Use human and material resources	64.1	92.1	28	35.9	30.8	-5.1
Communicating with students						
Fair and equitable manner	59.0	89.7	30.7	66.7	69.2	2.5
Effective interaction	61.5	94.9	33.4	69.2	69.2	0
Chalk board skill						
Write clear and correct	65.8	81.6	15.8	78.4	84.6	6.2
Main points on board	50.0	84.2	34.2	73.0	87.2	14.2
Design of layout	60.5	84.2	23.7	83.8	71.8	-12

* Mean Score >4

Program Impact on Teachers' Pedagogical Beliefs and Attitudes

The second research question addressed whether the 2+2 program had any influence on teachers' pedagogical beliefs and attitudes. Two questions were generated to address the impact of the major components of the program. The questions are listed below:

1. How does the number of 2+2 visitations predict change in teachers' beliefs and attitudes?
2. How does the amount of 2+2 feedback received predict the change in teacher's beliefs and attitudes?

Pedagogical beliefs and attitudes were operationally defined as posttest scores on a scale to investigate the teachers' pedagogical beliefs and attitudes with the pretest scores as the covariate. The pedagogical beliefs and attitudes served as the dependent variable, while whether or not participating in the 2+2 group was the independent variable. The number of classroom visitations the teachers had made and the amount of feedback they had received also served as independent variables

Analysis of covariance (ANCOVA) was employed to determine if the 2+2 program influenced teachers' pedagogical beliefs and attitudes. The "yes"s on the positively stated items and "no"s on the negatively stated items scored "1" point, while the alternative choice scored "0" points. The sum score of the teachers' pedagogical beliefs and attitudes was calculated by adding all the items' scores together. The results showed that there was not a significant difference between the posttest scores while controlling for the pretest scores ($F=.023, p>0.05$). The mean pretest score for the 2+2 group is 15.23 (SD=2.63) and the posttest sum score is 15.28 (SD=2.07). The pretest

score for the comparison group was 15.64 (SD=1.88) and the posttest score was 15.17 (SD=2.44). This suggests that the 2+2 program did not significantly change teachers' pedagogical beliefs and attitudes. Since there is no change on teachers' pedagogical beliefs and attitudes, no further analyses were performed to test the relationship between 2+2 visitation/feedback and the beliefs and attitudes. Therefore, the relationships between visitations or feedback and teacher's beliefs and attitudes (i.e. the two sub-questions) were not explored in the current study.

Program Impact on the Frequency of Classroom Observation

The third research question addressed whether teachers in the 2+2 program conducted more classroom observations than did teachers in the comparison group. The frequency of classroom observation served as the dependent variable and the group membership (2+2 or comparison) was the independent variable.

ANCOVA showed that the teachers in 2+2 group had far more observations than the comparison group across most of the observation categories ($p < .05$) after the program implementation except for parent observation ($p > 0.05$). The Bonferroni procedure was used to adjust the probability level for families of hypotheses, (i.e. the probability level for the four comparisons on observation is $.05/4 = .0125$). After the adjustment, the number of observation from/on colleagues, and from principals among the 2+2 group teachers remain significantly higher than that in the comparison group ($P < .0125$, Table 7).

Table 7: Comparison of Observation between 2+2 and Comparison Group

Observation	2+2 group (n=39)		Control group (n=39)		ANCOVA	
	Pre	Post	Pre	Post	F	P
	Mean SD	Mean SD	Mean SD	Mean SD		
1. How often do you observe other teachers teaching in one month?	1.49 .72	8.51 .88	1.74 .82	2.26 1.31	584.27	.000
2. How often do other teachers observe your teaching in one month?	1.59 .64	9.10 1.17	1.48 .79	1.72 1.17	797.93	.000
3. How often does the principal observe your teaching in one month?	1.15 .84	5.90 1.70	1.18 .56	1.31 .77	236.30	.000
4. How often do you have parents visit your classroom in one month?	.87 1.20	0.77 1.68	1.00 .23	0.97 .28	0.717	.400

Prior to the implementation of the program, 5.1 % of the 2+2 group teachers conducted more than three observations per month in other teachers' classroom while that of the comparison group teachers was 10.2%. Eight percent of the 2+2 group teachers had other teachers observe their teaching more than three times one month while 10.3 % of the comparison group teachers had other teachers observe their teaching more than three times one month. Three percent of the 2+2 school principals conducted more than three observations in a month, while none from the comparison schools had more than three observations, although the difference between the groups was not statistically significant.

A number of 2+2 group teachers (7.8%) reported they had three parent observations, whereas not one teacher in the comparison group observed parents.

At the end of the implementation of the program, 100 % of the 2+2 group teachers conducted eight or more observations one month in other teachers' classroom while none of the comparison group teachers achieved that rate. All (100%) of the 2+2 group teachers had other teachers observe their teaching more than eight times one month while none from the comparison group had the same (0%). Ninety-five percent of the 2+2 school principals conducted more than three observations in one month, while only 2.6% from the comparison schools had more than three observations. About 13% of the 2+2 group teachers reported that they had three parent observations while not one teacher in the comparison group had parents to observe them teaching; however the difference was not statistically different due to the small sample size.

Program Impact on the Frequency of Feedback

The fourth research question addressed whether teachers in the 2+2 program gave and received more feedback than did teachers in the comparison group. The frequency of feedback served as the dependent variable and the group membership (2+2 or comparison) was the independent variable.

ANCOVA showed that the teachers in 2+2 group received far more feedback than the comparison group across most of the observation categories after the program implementation ($p < 0.05$). The Bonferroni procedure was used to adjust the probability level for families of hypotheses, (i.e. the probability level for the four comparisons on feedback is $.05/4 = .0125$). After the adjustment, the feedback from administrators, colleagues and parents were significantly higher in 2+2 group ($p < .001$) (Table 8).

Table 8: Comparison of Feedback between 2+2 and Comparison Group

Feedback	2+2 group (n=39)		Control group (n=39)		ANCOVA F
	Pre	Post	Pre	Post	
	Mean	Mean	Mean	Mean	
	SD	SD	SD	SD	
1. How much feedback did you get from your administrators last month	1.44 1.98	9.72 2.08	1.18 .51	1.13 .52	642.21*
2. How much feedback did you get from your colleagues last month?	1.74 2.12	16.10 2.21	1.26 .55	1.26 .55	1737.19*
3. How often do you have students give you feedback on your teaching?	1.26 .55	2.36 1.66	1.21 .41	1.13 .66	4.83**
4. How often do you have parents give you feedback on your teaching?	1.72 1.76	2.03 2.01	1.31 .52	1.21 .73	18.10*

*ANCOVA ($p < .001$)

** ANCOVA ($p > .0125$)

These findings are further illuminated by the descriptive data presented below.

Prior to the program, 64.1% of the 2+2 group and 82.1 % of the comparison group teachers received less than two instances of feedback from administration per month.

Nearly fifty-four percent teachers in the 2+2 group and 74.4% of the comparison group teachers received less than two instances of feedback from peer teachers in one month.

About three fourths of the teachers in the 2+2 group and 79.5 % of the comparison group teachers received less than two instances of feedback from students in one month. Fifty-

nine percent of the teachers in the 2+2 group and 71.8 % of the comparison group teachers received less than two instances of feedback from parents in one month.

At the end of the program, none of the teachers in the 2+2 group and 84.6 % of the comparison group teachers received less than two instances of feedback from administration in one month. None of the 2+2 group and 74.4% of the comparison group teachers received less than two instances of feedback from peer teachers in one month. Thirty-six percent of the 2+2 group and 94.9 % of the comparison group teachers received less than two instances of feedback from students in one month. Forty-nine percent of the 2+2 group and 89.7 % of the comparison group teachers received less than two instances of feedback from parents in one month.

Program Impact on Teacher Collaboration

The fifth research question addressed whether teachers in the 2+2 program collaborated more in their teaching practice than did teachers in the comparison group. The collaboration was measured by the questions in the 2+2 response survey which asked the frequency of collaborative work. The frequency of teacher collaboration activities served as the dependent variable and the group membership (2+2 or comparison) was the independent variable.

ANCOVA showed that the teachers in the 2+2 group experienced far more collaboration than the comparison group across all of the pertaining items ($p < 0.05$) after program implementation (Table 9). The Bonferroni procedure was used to adjust the probability level for families of hypotheses, (i.e. the probability level for the four comparisons on feedback is $.05/5 = .01$). After the adjustment, the scores of all the

collaboration categories remain significantly higher in 2+2 group than in the comparison group ($P<0.001$).

Table 9: Comparison of Collaboration in One Month

In one month	2+2 group (n=39)		Comparison group (n=39)		ANCOV A
	Pre	Post	Pre	Post	F
	Mean SD	Mean SD	Mean SD	Mean SD	
1. I discuss instruction-related topics with my peers.	1.81 1.09	7.62 1.25	2.10 .60	2.31 .61	629.57*
2. I prepare lessons with my colleagues	1.75 .65	10.86 1.32	1.92 .74	2.05 .83	1403.915 *
3. I ask my colleagues for assistance.	1.69 1.51	10.95 1.62	1.38 .49	1.51 .56	1316.128 *
4. My colleagues ask me for assistance.	1.44 .84	4.56 .99	1.54 .55	1.56 .60	248.103*
5. My colleagues come up to discuss instruction-related topics with me.	2.03 .56	8.82 1.27	2.28 .60	3.00 .92	530.533*

* ANCOVA ($p<0.001$)

Prior to the program, 47.2 % of the 2+2 group and 12.8 % of the comparison group teachers involved in discussions on instruction-related topics with their peers less than twice a month. Thirty-six percent of the 2+2 group and 30.8 % of the comparison group teachers prepared lessons with their colleagues less than twice a month. Fifty-eight percent of the 2+2 group and 61.5 % of the comparison group teachers asked their

colleagues for assistance less than twice a month. Fifty-six percent of the 2+2 group and 48.7 % of the comparison group teachers had colleagues to ask them for assistance less than twice a month. Fourteen percent of the 2+2 group and 7.7 % of the comparison group teachers had colleagues to come up to discuss instruction-related topics with them less than twice a month.

At the end of the program, none of the 2+2 group and 7.7 % of the comparison group teachers discussed instruction-related topics with their peers less than twice a month. None of the 2+2 group and 30.8 % of the comparison group teachers prepared lessons with their colleagues less than twice a month. None of the 2+2 group and 51.3 % of the comparison group teachers asked their colleagues for assistance less than twice a month. None of the 2+2 group and 48.7 % of the comparison group teachers had colleagues to ask them for assistance less than twice a month. None of the 2+2 group and 5.1 % of the comparison group teachers had colleagues to come up to discuss instruction-related topics with them less than twice a month.

Program Impact on the Kind of Feedback Received

The sixth research question was raised to investigate the kinds of feedback that is provided to teachers who participated in the 2+2 program. Content analysis was employed to analyze the compliments and suggestions the teachers have provided on the 2+2 feedback forms. A total of 3314 forms were collected. Purposive sampling was used to draw the sample from the 3314 collected forms. A sample size of 350 was selected for the final data analysis according to Krejcie and Morgan's sample size determination method (Krejcie & Morgan, 1970). Altogether 350 forms were selected based on

teachers' demographics characteristics. Feedback provided to participating teachers was represented in the 2+2 feedback forms collected from the three 2+2 schools.

Even though all the teachers in the 2+2 group filled out the form, not all respondents were able to generate two compliments with two suggestions on each form. Altogether 688 compliments and 616 suggestions were available for analysis, each of which was assigned to a category and recorded on a coding form. Compliments and suggestions were categorized. The nine functions of the teachers' performance questionnaire were used as established categories. Compliments and suggestions were analyzed separately, and were tentatively assigned to one of the nine established categories in an iterative back and forth process. Aggregate results were calculated and are represented in Table 10. The top three categories that the teachers' compliments focused on were facilitating instruction (30.4%), instructional presentation (17.8%), and providing reinforcement and feedback (15.3%). The top three categories of teachers' suggestions focused on were facilitating instruction (30.6%), instructional presentation (14.9%), and communicating with students (12.7%). It is interesting to notice that facilitating instruction and instructional presentation were listed as both the top two compliment categories and the top two suggestion categories. Teachers' compliments and suggestions were quite scattered on the other categories, however, they had their attention focused on their colleagues' performance on facilitating instruction and instructional presentation.

Being considered highly related to improvement of teachers' performance, suggestions caught more attention of the researcher than did the compliments and most of them were productive. The suggestions given on facilitating instruction focused on using

more modern technology such video and audio, and computer-assisted activities.

Suggestions about instructional presentation addressed the oral presentation ability of some teachers and called for more training on this skill. Suggestions pertaining to communication with students reflected concerns on how to meet all the students' needs and encourage them participate in the communication, especially among inactive students.

Compliments and suggestions that could not be categorized under the pre-established categories were coded into the additional categories that were constructed for the analysis (Table 11 and 12). Most of those compliments and suggestions were general and non-specific comments which were categorized as "Non-productive" by LeBlanc (1997, p.121).

Table 10: Responses in Compliments and Suggestions Categories

Categories	Compliments		Suggestions	
	n	%	n	%
Preparedness for instruction	47	7.1	34	5.8
Management of instructional time	38	5.8	39	6.6
Management of student behavior	23	3.5	21	3.6
Instructional presentation	117	17.8	88	14.9
Monitoring of student performance	26	4.0	65	11.0
Providing reinforcement and feedback	101	15.3	59	10.0
Facilitating instruction	200	30.4	181	30.6

Communicating with students	78	11.9	75	12.7
Chalkboard skills	28	4.3%	29	4.9
Total	658	100.0%	591	100.0

Table 11: Responses of Non-Productive Compliments by Category

Compliments	n	%
Good job	3	10.0
The students enjoyed	4	13.3
I learned a lot	9	30.0
Can be a model for others	1	3.3
Has paid attention to students' moral well-being	9	30.0
A qualified teacher	3	10.0
I like the way you smiled to the students	1	3.3
Total	30	100.0

Table 12: Responses of Non-Productive Suggestions by Category

Suggestions	n	%
I am not qualified to comment	5	20.0
Be brave and more open	8	32.0
Change your accent	2	8.0
Read some more educational theories	1	4.0
More attention on female students	6	24.0
Should not depend too much on the textbook	3	12.0
Total	25	100.0

Teachers' Perceptions of the Advantages and Disadvantages of the 2+2 system

The seventh research question explored teachers' perceptions of the advantages and disadvantages of the 2+2 system. The researcher interviewed 30 participating 2+2 teachers in September 2002 prior to the closure of the program in focus groups. The focus group interviews were conducted around ten interview questions (Appendix E). The first eight interview questions (question A-H) were used to address teachers' perceptions of the advantages and disadvantages of the 2+2 system. The last two interview questions (question J and K) addressed teachers' suggestions for improving the 2+2 Program, which is presented in the next section. The focus group interviews were analyzed on a cross-case basis. Similarities of responses were noted and categories were developed around similar responses which were clustered. The results of the focus group meetings are described below. The results are presented separately by question. For each question, a table shows the frequency and percentage of teacher responses by category. To complement the tables and retain the richness of the data, excerpts of transcribed responses are noted under each category.

A. How do you describe the implementation of the 2+2 program?

The responses to this question were broadly categorized as either "went well" (e.g. positive responses such as "excellent", "successful", and "great"), "fairly" (e.g. moderate responses such as "OK", "half and half" and "it depends") or "not well" (negative responses such as "unsuccessful", "un-smooth" and "poor") (Table 13). Twenty-four respondents (80%) indicated that they had no difficulty implementing the program by the time they were interviewed. They perceived the implementation of the program as

successful and efficient. An example of this type of comment was, “2+2 is easy to do. It has worked fine for me.” A teacher named Li put it this way:

The theory about 2+2 is quite simple. We came to understand it as soon as the definition was delivered. We know that the purpose of 2+2 is for us to do more classroom observations and provide more feedback to one another, promote collaboration, and improve instruction. It took very little time for us to learn about the procedure of application. It was not a very difficult task for us to use 2+2 in classroom observation. We come into another teacher's classroom to stay long enough to come up with two compliments and two suggestions. It is not a big deal. The only thing that we must remember is that we should at least conduct two observations each week.

Three teachers (10%) evaluated the implementation of the 2+2 program as going “fairly well.” One teacher said, “I have been doing what we were asked to. I think everything goes in the way exactly as was planned.” A Chinese language teacher illustrated her experience with the 2+2 program implementation as:

We understand that 2+2 has become part of our job. We followed every step and procedure that was specified in program plan. I did the required numbers of visitations and provided feedback as was instructed. The procedure itself was not difficult.

Three attendants (10%) considered the implementation of the program as unsuccessful.

One of the teachers put it in this way, “It is useless but disturbing. I try so hard to think of what to say. I do not think it works.” A teacher whose name is Wang described the program implementation as:

2+2 sounds very simple. But it is not very easy to compose the feedback. The program is an unnecessary add-on to my schedule, and I found it hard to do so many visitations each week. It is not a good idea to have so many teachers running around to do observations every day. Instruction is frequently disrupted with teachers coming and leaving the classroom. I can hardly believe that this program will last.

Table 13: Teachers' Description of the Implementation of 2+2 Program

Response	N	%
Went well	24	80
Fairly	3	10
Not well	3	10
Total	30	100

B. How do you compare “2+2” with the traditional teacher performance appraisal system?

The majority (60%) of the respondents expressed their strong preference of the 2+2 to the traditional teacher performance evaluation system (Table 14). A math teacher compared “2+2” with the traditional teacher performance appraisal system in this way:

We finally have found an appraisal system that is not so complicated and threatening. Before, seldom would you have a colleague come in and observe. Administrators and outsiders occasionally came to watch us teaching. They were always very critical and picky. They would give us a long list of things that we should improve on which were very often too confusing to handle with. 2+2 is simple and effective. It is meant for us ordinary classroom teachers. You do not have to know a lot of theories before you practice it.

An additional six (20%) agreed that “2+2 is a better alternative than the traditional teacher performance appraisal system. They proposed that 2+2 stand side by side with the traditional teacher performance appraisal system to help teachers to improve instruction:

2+2 can be a substitute of the traditional teacher performance appraisal system. It is easier to practice and less time consuming. It is especially a better tool for teachers to appraise each other's performance. Not a lot of training is required

before you can come into a classroom to do 2+2. It is better to evaluated teachers with traditional system as well as 2+2.

Five (17%) teachers indicated that “2+2 is quite another thing. It is a mistake to compare it with other teacher evaluation systems”. They proclaimed that 2+2 does not share those characteristics of an appraisal system:

2+2 is not a system to appraise teachers’ professional performance. It can never indicate how well a teacher performs in class. No matter how well you do things or how badly you teach your students, the feedback is set to be two compliments vs. two suggestions.

Only one teacher regarded 2+2 as worse. She complained that:

2+2 distracts students’ attention and waste teachers’ time. It is another new method that carries a fancy name, but with no positive effect. It is so hard to focus on real teaching when you have to pop in and out of other’s classroom so often. Your own teaching is frequently disrupted. You can never expect to do serious observation with 2+2.

Table 14: Teachers’ Comparison of “2+2” with the Traditional Appraisal System

Response	N	%
Much better	18	60
Better	6	20
Worse	1	3
A mistake to compare	5	17
Total	30	100

C. What were the strengths of “2+2”?

The respondents listed various strengths of the 2+2 system (Table 15). Those who gave positive responses to the implementation of the 2+2 program described the strength of 2+2 as being “not threatening, time-saving, promoting cooperation among teachers, easy to use, specific and focused.” One teacher commented:

I had thought of giving up the job because every day seemed to be a repetition of the previous ones. I was burned up. Now we have 2+2. My colleagues come to help bring changes to the classroom. You know very well that those sitting in the classroom watching you teaching are to help you to improve. You do not have to be embarrassed with a long list of things that you have not done well. When time comes for you to observe others, you do not have to stay there until the students are released.

Another teacher considered it a great of experience of learning:

I think I am learning and growing everyday. It is quite encouraging to have your colleagues to tell you your strength in teaching. 2+2 gives us opportunity to learn from each other in a comfortable way. I now begin to treat my colleagues as friends and allies in helping our students, not only as competitors. For the first time I am not feeling threatened by an observer sitting in class watching me.

One young teacher was delighted by the opportunity of interaction with other teachers:

It is such a good thing that I can now go to another teacher's classroom whenever I feel that I would like to. I have more opportunities to learn from others. All my colleagues are my natural mentors. And it is so nice that I can even offer my feedback to those senior teachers. Before, I always tried to say nice things about their teaching. Now I must give them suggestions as well. I am growing much faster.

Table 15: Teachers' Perceptions of the Strengths of 2+2 System

Response	N	%
Not time consuming	24	80
More helpful for improvement of instruction	24	80
Strengthening cooperation	21	70
Not threatening	18	60
Encouraging	18	60
Easy to handle	15	50
Bringing changes to classroom	12	40
No comment	3	10

D. What were the weaknesses of “2+2”?

Those who had negative stance toward 2+2 implementation pointed out the weakness of the system as being “disruptive to instruction, time-consuming, giving only unthoughtful, unstructured and random feedback” (Table 16). One teacher commenting on 2+2 said:

Students' attention is distracted whenever someone comes in during the classroom session. When they hear something funny, they turned to look at the observers to see how they respond; when their teacher makes a careless error, they tried to find out how the observers take it. Some observers even grab a textbook from a student to get an idea about what teacher is talking about. Some may ask a student to inquire about what is going on in class. It is such a stressful thing to have people watch you everyday.

One teacher was irritated by receiving suggestions that she thought as improper:

I cannot accept those irresponsible and carelessly made suggestions. They came in just in the middle of a session, and stayed there just for some minutes trying to

critique how the teacher was conducting teaching. The feedback provided is nothing as reliable.

Some teachers viewed the system as something too simple, nothing theoretical. They argued:

Why 2+2, not 3+3 or other? What is the theory behind the 2+2 system? All those ideas and assumptions about the system seem to be groundless. The system is arbitrary, and with no evidence to support it. It is something that simply jumped out of unrealistic mind.

Table 16: Teachers' Perception of the Weaknesses of 2+2 System

Response	N	%
Disruptive to instruction	1	3
Time-consuming	3	10
Giving only unthoughtful, unstructured and random feedback	3	10
Not theoretical	6	20
No comment	24	80

E. What progress have you made with “2+2”?

When the teachers were queried about the progress they had made with the 2+2 program, 18 (60%) of them reported that they had completed the required visitations. Twelve (40%) said they had made more visitations than required (Table 17).

Table 17: Teachers’ Perceived Progress with 2+2 Program

Response	N	%
Have done the required visitations	18	60
Have done more visitations than required	12	40
Have done less visitations than required	0	0

F. How did you benefit from it?

The teachers mentioned various benefits that they had from the 2+2 program (Table 18). One teacher declared:

I think the 2+2 program has done wonders to improve my performance in class. It gives me a real sense of growing each day with compliments and suggestions from my colleagues. The program gave me more chances to observe others teach and learn from them.

Others commented that the 2 + 2 program help to promote colleague interaction and collaboration:

My colleagues have become my resources for instructional improvement. Before we met each other to say hello or involve ourselves in some small talks. Whenever there was some serious discussion, most likely it would be on some topics unrelated with what take place in classrooms. Exchange of ideas about teaching only occurred at those to meetings organized by the authority. Now out way of life at school is different. We visit each other's classroom frequently and try to help one another to improve.

Table 18: Teachers' Description of the Benefits from 2+2 Program

Response	N	%
Have more chance to observe others teach	28	93
Have more opportunity to discuss instructional affairs with colleagues	18	60
Have more opportunities to learn from others	24	80
More willing to bring changes to class	24	80
No comment	2	7

G. What were your reactions to being observed?

A majority of teachers felt very comfortable about being observed by peers (Table 19). Comments included: "It is expected that we observe each other teach. I like it." "I welcome it. I know someone is there to help me." "It does not bother me that someone is watching me teach." "I feel observers are being like part of the environment." Ten teachers accepted having observers in the class, though they are very conscious of their presence. A typical response of this category was, "I do not feel uncomfortable." Six teachers felt varying degree of anxiety. "I felt uneasy when I know that I am being observed." "I felt that my instruction is disturbed."

Table 19: Teachers' Reaction to Being Observed.

Response	N	%
Comfortable	14	47
Fairly comfortable	10	33
Uncomfortable	3	10
Very uncomfortable	3	10

H. What were your reactions to being an observer?

Almost all of the respondents found no difficulty in observing others teach (Table 20). Most expressed that they were quite comfortable while observing others. They saw the opportunity as precious. One teacher remarked, "Every time we observe someone else it helps us to look at ourselves in another light." Younger teachers were happy that they were not assigned to observe or have to ask permission to observe when they want to, as one teacher commented, "I am so happy that I no longer have to ask a permission to sit in another teacher's class".

Table 20: Teachers' Reaction to Being an Observer

Response	N	%
Happy	6	20
Comfortable	18	60
Fairly comfortable	3	10
No comment	3	10

Suggestions Teachers Provided for Improving the 2+2 Program

During the focus group interview, the teachers provided a variety of suggestions for improvement of the 2+2 program. A number of modifications to the 2+2 program were proposed by the teachers (Table 21).

Table 21: Suggestions Provided for Improving 2+2 Program

Suggestions	N	%
Make 2+2 compulsory for all teachers	24	80
2+2 become part of school administrators' job on daily basis, each of them should make at least three 2+2 visitations each day	28	92
Give teachers a guideline handbook on 2+2	21	70
A summative evaluation conference should be held each week within the teaching research groups	12	40
More training on 2+2 is needed	9	30
2+2 can be oral. We do not have to write the complements and suggestions on paper.	6	20
No comment	2	6

Almost all of the teachers (92%) recommended that 2+2 become part of school administrators' job on daily basis, and each of them should make at least three 2+2

visitations each day. The teachers' comments include: "They now only come to observe our teaching when it is evaluation time. They have no idea about our daily work." "They try very hard to find out what went wrong in the class. They think it is their natural duty to point to our weaknesses. It is so discouraging to hear their comments after the evaluation." "It is good for administrators to make balanced feedback as is required by 2+2." "The administrators should come to classrooms more often. Some of them seldom teach that they have no idea about what it is like in the classroom." The majority of the teachers suggested that 2+2 should be made compulsory for all teachers. "If it is voluntary, many teachers will not buy it even if it proves helpful." "Teachers need some pushing". More training was proposed. Some teachers requested a guideline handbook on 2+2 in order to make it easier for them to wrap up the two complements and two suggestions more adequately.

I. What is the next step in your plan for "2+2"?

The majority of the teachers (60%) declared that they would employ 2+2 to help each other improve instruction only if there is continuous administrative support (Table 22). They believed that 2+2 could not last if there is no active leadership from the principal. "Nothing can happen without the principal's command." Only three teachers determined to use 2+2 under any circumstances. One teacher said: "I see no point giving up such a wonderful tool for evaluation."

Six teachers (20%) felt varying degrees of anxiety when they were inquired about what they would do for the next step. One teacher was very concerned that "My colleagues would expect me to do a very good job every time. I am afraid I will be burned out." Another teacher worried that his list of compliments and suggestions would

soon run out and “I will have difficulty offering the two compliments and two suggestions.” Three teachers (10%) reported that they would only continue to conduct 2+2 only if the leadership asks them to.

Table 22: Plans for the Next Steps for “2+2”

Response	N	%
Would continue no matter what	3	10
Continue if there is administrative support	18	60
Felt varying degree of anxiety	6	20
Would only continue if the leadership ask them to	3	10

J. What else can you tell me about 2+2 that would help me understand its impact?

In response to this question, various themes emerged throughout the discussion. Half of the teachers acknowledged that the 2+2 system changed their way of thinking about evaluation, which should be encouraging and balanced with compliments and suggestions. One teacher said:

I used to think that evaluation is to find out a teacher's pitfalls. Only those senior and lead teachers can enjoy the sunshine of positive comments. I now know that the first step to the road of improvement is to know your strength as well as your weaknesses.

Another reported to understand that the improvement of teaching is an ever-continuing step-by-step process:

I understand very well that you can never expect to turn yourself into a perfect teacher in one day for one shot. The road to success and perfection is through a process of gradual improving. You learn only one or two things a day, but day by day you build on what you have already achieved and keep on growing. The 2+2 system can serve as a powerful tool to their assistance on their life-long journey for perfection.

Summary of the Findings

The analysis of results presented in this chapter provided a foundation for interpretation of the data that addressed the research questions. The ultimate question the study intended to answer was whether the 2+2 program was successful and efficient as was defined by the program objectives. Data were analyzed to investigate whether the 2+2 program in Shanxi, China contributed to improvement of teachers' professional performance, changes in teachers' beliefs and attitudes, the frequency of observations they conduct and receive, and the quality and type of the feedback they provide and obtain. The results revealed that the 2+2 program had a significant impact on teacher's professional performance, though no similar impact was found on teacher's beliefs and attitudes. The results also showed that the 2+2 program significantly increased the frequency of teachers' classroom observation, feedback and collaboration.

Data obtained from the survey and interview revealed that teachers in the 2+2 group collaborated more in their teaching practice than those in the comparison group. Content analysis was utilized to obtain information about the pattern of comments being made as a result of 2+2 observations. Most of the teachers' perceptions on the 2+2 program were positive and many of them expressed that they would use 2+2 again in

their future teaching work. The quality of feedback was perceived as high quality by the teachers themselves and the researcher.

CHAPTER V

IMPLICATIONS AND CONCLUSIONS

The purpose of this study is to examine the effectiveness of the 2+2 Alternative Teacher Performance Appraisal System that has been implemented as one of the educational reform efforts in Shanxi province in China. Information was gathered and analyzed to reveal the program implementation and to evaluate its effectiveness. Data triangulation was achieved through varied data sources including focus group meetings, 2 + 2 feedback forms, and questionnaires. Results will be presented to the stakeholders, including Shanxi Educational Commission, the local school districts, and the participating school, to help them to determine whether or not the program should be further implemented. In this chapter, implications and conclusions drawn from the findings are reported, and recommendations for future 2 + 2 implementation offered. The discussion is divided into seven sections that reflect findings related to the outcome measures: teacher professional performance, observation, 2+2 feedback, teacher's collaboration, teachers' pedagogical beliefs and attitudes, and the future direction for the 2+2 program.

Professional Development

Evaluation data indicated that the 2+2 program made a significant positive difference in the way how teachers perform in class. After exposure to the program, the teacher in the 2+2 group performed better in all of the nine functions that were measured by the evaluators. Teachers felt overwhelmingly that 2+2 helped them with their professional growth. In answering the question "what were the strengths of 2+2?", 80% participating teachers reported that 2+2 was helpful for improvement of instruction. This

result adds new knowledge to conventional wisdom on teachers' professional performance.

Conventional wisdom holds that improvement of teachers' professional development relies on practices such as participation in teacher workshops, special training, additional college course or advanced degrees, frequent participation in in-service meetings, as well as being a member of teachers' organizations, networks, or unions (Pelletier, 1995). Traditional approach to teachers' professional development has formal courses and in-service seminars as the central components which are considered like a voice coach giving advice to a singer whom he or she has never heard sing (Eisner, 1992). Teachers are not often consulted on what type of assistance they need, adding to perceptions that professional development is a waste of time (Guskey & Huberman, 1995). Although the need for professional development is apparent to those who study school improvement, effective professional development is not taking place in most schools. The current state of professional development is aptly described by Miles (1995):

A good deal of what passes for "professional development" in schools is a joke—one that we'd laugh at if we weren't trying to keep from crying. It's everything that a learning environment shouldn't be: radically under-sourced, brief, not sustained, designed for "one size fits all," imposed rather than owned, lacking any intellectual coherence, treated as a special add-on event rather than as part of a natural process, and trapped in the constraints of the bureaucratic system we have come to call "school." In short, it's and no more knowledgeable, skilled, or committed than before. (p. vii)

Reasons for the failure of many teacher professional development activities to produce long term change are well documented (Goertz, Floden, & O'Day, 1996). Summarizing these reasons, Miles (1995) strongly criticized traditional one-shot professional development courses, characterizing them as opportunities for active engagement, being able to demonstrate a link between theory and practice, including time

for reflection, and modeling exemplary practice. The facilitation theme included ensuring that the materials were high quality and user friendly. The application theme included ensuring that activities were followed up with support from the planners and that teachers were motivated and empowered as a result of the experience (Miles, 1995).

Over the last several years, Gordon (2004) has conducted a national study on outstanding school-focused professional development programs. He found that even though each of the professional development programs had a different focus, the programs shared several common characteristics. These characteristics are similar to those identified in a long line of research and literature on effective professional development (Birman, Desimone, Porter, & Garet, 2000; Guskey, 1998; Norton, 2001; Richardson, 2000; D. Sparks & Hirsh, 2000; Wood, 1993) The characteristics are strong leadership and support, collegiality and collaboration, data-based development, program integration, a developmental perspective, relevant learning activities, and professional development as “a way of life.” The 2+2 program shares many of the characteristics identified as effective professional development. Evidence documented and analyzed in this study points to the conclusion that 2+2 helped teachers to improve their professional performance. Not limited to the traditional approaches, the 2+2 program addresses the interaction between teachers, teachers and the administrations. The key components of the 2+2 program, two suggestions and two compliments, come from observation and require collaboration. The improvement on the performance is the results of observation of each other’s work, and the collaboration of peers.

Observation

The current study showed that 2+2 program significantly increased the frequency of the teacher observation. This result was expected because the 2+2 visitation was one of the mandated components of the program. The teachers in the 2+2 group had far more observations than the comparison group. At pre-test, the teachers in the 2+2 group conducted observations on an average of 1.5 times per month, while at post-testing, they reported observing others teaching about nine times a month while the comparison group teachers did that only about twice a month. Addressing observation makes 2+2 distinguished from other teacher education programs.

Observation is the basis for the improvement of professional performance. However, the craft of traditional teaching is still largely a private affair in many classrooms where observation is not a popular form. Some teachers spend 40 years in the classroom, teaching maybe 50,000 lessons or more, of which only a tiny number are witnessed by other adults (Lytle & Fecho, 1991). It is clear that the daily routines of schools provide little time and few opportunities for teachers to interact and share ideas with each other (Lytle & Fecho, 1991). It is often difficult to obtain detailed accounts of lessons, because teachers are so busy with the running of the lesson there is little time for them to observe each other teaching.

How can we frame the characteristics of quality observations? This is an important question for those who are engaged in teacher education. Authors such as Stiggins (1986) and Airasian (1991) stress the importance of classroom realities. They emphasize that in order to really assist teachers, researchers must understand that the demands of the classroom are different from what is relevant in research. Along with this

realization, Stiggins (1986) still believes that teachers need help from researchers. The kind of help needed remains unspecified. Airasian (1991) also believes that better teacher education is important for improving the quality of observations. Because teachers are often unaware of the extent to which they rely on informal, unsystematic assessment for decision making, more information on sizing up students and on instructional assessment should be included in courses and texts.

Teachers' need for valid and reliable help in conducting high quality observation was addressed at the beginning of the program implementation. Training workshops was organized for the 2+2 teachers to learn about the principles of observation and master the skills and techniques required for classroom observation. Observation is becoming an essential part of teaching. Above 80% of the 2+2 teachers felt comfortable and happy about being an observer. An equal number of teachers felt comfortable about being observed. All 2+2 group teachers conducted more than two visitations to others' classroom each week.

One of the problems faced by both experienced and inexperienced classroom observers is the matter of deciding what should be the focus of attention. So much happens in classrooms that any task or event, even apparently simple ones, could be the subject of pages of notes and hours of discussion. The ecology of many classrooms can be extremely rich and full. However, 2+2 teachers do not have to concern about the quality of observation. The system was designed that teachers observe teaching just to come up with two compliments and two suggestions that he or she feels appropriate and important. It is feasible for teachers to conduct more observations because the system

makes it easier to handle and not time consuming as 80% of the respondents reported.

The 2+2 program provides a model on how to conduct more observation on teaching.

Feedback

In the 2+2 program, feedback is the crucial bridge of observation and collaboration. It requires two suggestions and two compliments instead of any negative remarks. It is interesting to notice that teachers produced more compliments than suggestions, which may indicate that unlike professional evaluators, teachers tend to be more willing to offer compliments rather than suggestions in the 2+2 setting. Above 60% of the respondents characterized 2+2 feedback as encouraging. This result indicates that teachers would like to help and encourage their peers' work, and the 2+2 feedback provides a way for them to share the experience. The productive and constructive feedback made the collaboration in the 2+2 program more frequent and successful.

The literature in the area of teacher professional development identifies feedback as a critical aspect of the learning process for teachers (Lehman, 1989). Information about performance-strengths and weaknesses-is essential to continued learning and professional growth. Nevertheless, teachers in practice seldom receive feedback about their performance and even less frequently meaningful information that actually leads to professional development (Bridges, 1986; Natriello, 1984).

In traditional evaluation, teachers are provided with a lengthy list of comments on his or her teaching at the end of one observation. He or she has too many things to consider for improvement. With 2+2, teachers are observed for a short period of time, just enough for the observer to compile two compliments and two suggestions. The teachers are expected to benefit from numerous, regular visitations. Multiple perspectives

on instruction are not gained just from observation of one classroom session, but a series of visitations by several colleagues and administrators.

The goal of traditional evaluation is to provide both positive and negative feedback to teachers. But very often, teachers receive far more negative comments from administrators and students at the time of evaluations (Munro & Elliott, 1987). The 2+2 program becomes a remediation of this pit-fall with the concept of balance as a valued component of the system. People need feedback to learn. Teachers need 2+2 to have more opportunities for feedback. Feedback is most valuable when people have the opportunity to use it to revise their thinking as they are working on a unit or project.

Teachers' Collaboration

The teachers in the 2+2 group conducted far more observations. As a result they experienced collaboration much more than the comparison group across all of the pertaining categories after the program implementation. In the interviews, all of the teachers responded that they involved themselves more often in discussions of instruction-related topics. Prior to the program they prepared lessons with their colleagues less than twice a month, while after the program they did teamwork in lesson preparation about 11 times a month. They asked their colleagues for assistance less than twice a month before the program implementation, they did that about 11 times. The implementation of 2+2 represented a fundamental change in the way they interacted with colleagues.

Teamwork develops through observation and communication (LeBlanc, 1997b). In the fields of education, no one opposes sharing information, developing common goals, collaborating in planning and implementing programs, and sharing responsibility for the

achievement of quality services for students. Collaboration is compatible and congruent with the goals of all organizations devoted to educating students, helping people, and facilitating change. Teacher collaboration has been generally applauded for its potential in improving the working lives of teachers, reducing teacher uncertainty, enhancing teachers' professional self-image and promoting collegiality and school learning (Kain, 1996). And yet in spite of the rhetoric in support of the "team concept," many teachers still work in schools where isolation, not cooperation exists more as a reality, where protecting one's turf is more important than providing coordinated services, and where teachers feel trapped and frustrated instead of creative and powerful. In his landmark study of teacher isolation, Lortie (1975) identified "the persistence of separation and low task interdependence" (p. 15) in "egg-crate" schools as the distinguishing organizational factor of schooling. As long as teachers were not interdependent, what each did in his or her own classroom did not affect colleagues' work. With the exception of a brief and limited foray into team teaching during the latter 1960s and early 1970s which never caught on nationwide, teaching has been marked by the absence of collaboration (Bohlmann, 1998).

The idea that teachers should cooperate, communicate effectively, and be "team players" has been discussed, advocated, and accepted by educators and human services professionals for a long time (R. Johnson, 1976). Not until the major reform efforts beginning in the 1980s did collaboration begin to be seen as one of the critical goals of educational reform (Legters, 1999). Teacher collaboration then has been generally applauded for its potential in improving the working lives of teachers, reducing teacher uncertainty, enhancing teachers' professional self-image and promoting collegiality and

school learning (Kain, 1996). Studies of teacher collaboration in schools have revealed associations between collaboration and outcomes such as collegiality (Stevenson, 1987), increased productivity and expertise (Brandt, 1987), improvement of teaching practice (Crandall & Loucks, 1983), teachers' perceptions of increased learning opportunities (Rosenholtz, 1989b) improvements in school climate and teachers' sense of efficacy (Leggett & Hoyle, 1987), and teachers' preference for collaborative structures (Holly, 1982).

The 2+2 program supports the contention that collaboration is a critical part of education. The 2+2 system is a new framework for teachers to collaborate. It offers opportunities for teachers collaborate in improving their instruction by observing each other teaching, then give and receive feedback.

Teachers' Pedagogical Beliefs and Attitudes

The current findings suggest that the 2+2 program did not significantly change teachers' pedagogical belief or attitude. This contradicted the researcher's expectation that teachers' belief and attitude would change after observing others' teaching. However, the result of the current study found consensus in previous studies.

Beliefs and attitude has been claimed as "thoughtful analysis of the nature of the relationship between beliefs and practice suggest that belief systems are dynamic, permeable mental structures, susceptible to change in light of experience" (Thompson, 1992). Empirical research findings of teachers' beliefs and attitude appear to support claim that teacher beliefs and belief systems are grounded in their personal experiences and, thus, are highly resistant to change (Pajares, 1992).

There are other reasons that the 2+2 program had no influence on teachers' beliefs and attitude. First of all, as the results showed, there were high agreements on certain items of the teacher belief and attitude survey, which suggests lack of variation in responses. Measures on beliefs and attitudes need to be improved and the questions need to be constructed based upon developed theoretical model. Second, the previous teachers' professional development programs have focused on changing teachers' beliefs and attitudes, therefore most teachers may have responded in socially desirable ways.

For the development of 2+2 program, beliefs and attitudes should be addressed because the role of beliefs as a determinant of teacher change has received increasing attention among researchers over the last several years, and a substantial body of evidence has emerged during this time suggesting that teacher beliefs drive instructional pedagogy (Pajares, 1992; Richardson, 1996; Thompson, 1992). To change teaching practices, teachers' beliefs should be taken into consideration (Hart, 2002). These beliefs are then either challenged or nurtured during what Lortie (1975) termed as a period of apprenticeship of observation, which occurs throughout the teacher training program. Thus, as recommended by Hart (2002) "It seems imperative that teacher education programs assess their effectiveness, at least in part, on how well they nurture beliefs that are consistent with the program's philosophy of learning and teaching." (p.4)

Recommendations for Future 2+2 Practice

Strong leadership and administrative support contributed to the success of the program. The participating teachers expressed satisfaction with the principal and administrators for their role in organizing program activities. Leaders established an atmosphere of support and trust, offered incentives and rewards for program participation,

and provided sustained moral and material support. It is a common reality in most Chinese schools that the principal has so many other priorities that he or she spends little time in classroom observation. However, it is recommended, as the teachers indicated, that the leadership should conduct 2+2 themselves to serve as role models by participating fully in the program.

One of the major complaints the teachers had about the program implementation is that the orientation period was too short. A lack of full understanding of the 2+2 system was felt by a number of participating teachers. They experienced difficulty in composing the two compliments and the two suggestions. They felt that they were thrown into the water before they could learn to swim. It is recommended that longer and more systematical orientation training should be conducted prior to the implementation.

Variations in the age, gender, teaching experience and subject area of the teachers may have an effect on the program implementation and outcome. During the interview sessions, more enthusiasm was exhibited by the younger teachers. Senior and experienced teachers tended to give more and detailed answers. It is recommended that the program should develop certain component address the age/experience difference between teachers.

Limitations and Recommendations to Further Research

Limitations

The current study was designed to develop an evaluation model to measure the effectiveness of the 2+2 program on teachers' growth in terms of their pedagogical beliefs and attitudes, teachers' professional performance in classroom, teachers' responses to the 2+2 program, teachers' feedback and teachers' collaborative practice.

Because the unit of the analysis of this study was the teacher, six schools were selected as similar as possible in order to fairly compare the teachers in the 2+2 schools with the comparison schools, however there was no guarantee that the teachers in the 2+2 group and comparison group were equivalent. Therefore, this study was considered as quasi-experimental design even though random assignment was used to assign the six schools into treatment group and comparison group. As a result, the groups may be different prior to the study. That is, the NEGD is especially susceptible to the internal validity threat of selection. Any prior differences between the groups may affect the outcome of the study. Under the worst circumstances, this can lead to the conclusion that the program didn't make a difference when in fact it did, or that it did make a difference when in fact it didn't (Compbell & Standley, 1960). Nonetheless, in an education study, random design is not feasible for most occasions, and NEGD is the best choice given the practical considerations. The conditions of the investigation necessitated a quasi-experimental design. Therefore caution should be exercised in interpreting the findings due to the possibility of selection bias.

Another threat to the internal validity was the commitment of the participants, especially those of the administration toward program. While most administrators believed that the 2+2 program had merit, they sometimes were overtaken by their daily routines. There was a huge difference among the project schools in the level of administrative involvement.

As noted, there was also a concern about the social desirability bias of the self-report measures. Teachers might respond the survey in a "good" but untrue way to impress the administration.

History may also be a threat to the internal validity of the study. Many events occurred during the program implementation. These events included change of curriculum and standards. Another potential confounding variables of history was that 2+2 was a follow-up of the LTTP project. The two may take effect on school culture at the same time besides many other factors influencing a teacher's profession growth. Therefore, caution must be taken in drawing conclusions or forming extrapolations regarding educational progress that may have taken place in those project schools.

The evaluator's bias was another threat to the internal validity. The researcher's close relationship with stakeholders may lead to overly favorable reporting and a lack of honest negative feedback (Patton, 1997). While easy access to information facilitates data collection, the researcher was, as a proponent of systemic reform and the LTTP initiatives, also subject to bias in evaluating the 2 + 2 program.

The main threat to external validity was the sampling method and sample size. The sample was selected through a convenient approach from a few key schools with a small number of teachers. Only first grade teachers were selected to participate so that the results of this study might not be generalizable to the other two grades in high schools in Shanxi. In addition, the characteristics of key schools limited the generalizability of the results to the non-key schools.

Although the design of this study has limitations, the findings generated will provide valuable information to the limited body of knowledge regarding the 2+2 alternative teacher performance appraisal system. As Patton (1997) puts it, "new participatory, collaborative, intervention-oriented, and developmental approaches were appropriate and helping intended users make informed decisions about their

appropriateness” (p. 112-113). It is a challenge to ensure that evaluation results are reported with the needs of the intended users in mind.

Recommendations to Future Research

There exists a great potential for further research in the 2+2 program. Many more questions could be asked regarding components of the program and its implementation. For example, in what way did the orientation of the program influence the program outcomes? What role did the administration play in the implementation process? How do variations in the age, gender, teaching experience, and subject area of the teachers affect the program implementation and outcomes? To what extent did the 2+2 program have an impact on students’ learning? These questions can inevitably create many future research opportunities. Teachers’ expectation about the 2+2 program focused on improvement of instruction from the outset (LeBlanc, 1997b). But the ultimate objective of the 2+2 program lies in improvement of students learning. The current study sacrificed the variation of students’ learning for the sake of simplification. It remains a research topic for future studies. In addition to the potential research questions, better measurements are recommended for the further research. More sensitive instrument that can better measure changing the beliefs and attitudes of teachers may help understand the impact of program on the teachers.

Further research need to address the limitations of the current study to eliminate the threat to internal validity and external validity. First of all, further research need more non-program-involved researcher for eliminating the researcher’ bias. However, the potential for reporting exaggerated positive results is offset by a genuine desire to determine the impact of the 2+2 program and how to improve it. That is, if the program

in its present form shows no positive impact on teachers' perceptions of isolation, collegiality, or enhanced instructional repertoires, this evaluation may point the way to major modifications needed in the program. The program must be congruent with the needs and values of participants to support systemic reform. The researcher is also open to issues as yet unknown that may emerge from in-depth feedback from participants. Information and communication, without hidden agendas, are keys to any talk of empowerment and organizational change, including programs such as 2 + 2. The researcher's unique knowledge of the LTTP project from the outset leads naturally to an informed perspective, but also a perspective aware of the pitfalls of bias. Because the values of an evaluator always influence his or her perspective, regardless of the evaluator's role as internal or external, results of any evaluation must be examined in the context of the needs of the stakeholders and the purpose of the evaluation.

Secondly, individual interviews are recommended for future research for diminishing potential social desirability bias. Individual interviews might provide in-depth information which the participants might not be willing to share in the focus group or in the written survey.

Last but not least, further research needs to be done in those schools have not experience similar program such as LTTP to decrease the influences of history events. The inclusion of regular schools, not only the key schools, is necessary to make the sample more generalizable. Random selection and larger sample size is recommended for future research when possible.

Conclusions

Although this study has limitations, the findings generated provide valuable information to the limited body of knowledge regarding the 2+2 alternative teacher performance appraisal system. Overall the program was successful and effective in helping teachers improve their professional performance. It might not be a startling result that 2+2 increased the frequency of teachers' classroom visitations and the amount of feedback they had provided and received, for the program mandated that teachers using 2+2 should conduct at least 2 visitations each week and come up with feedback consisting of 2 compliments vs. 2 suggestions at the end of the visitation. But the research has produced some encouraging findings pertaining the positive relationship between 2+2 and teacher collaboration. Among other things, the whole experience of participating in the program was a learning process for the 2+2 teachers and administrators. As Patton (1987) puts it, "new participatory, collaborative, intervention-oriented, and developmental approaches are appropriate and helping intended users make informed decisions about their appropriateness" (p. 112-113).

BIBLIOGRAPHY

- Airasian, P. W. (1991). Perspectives on measurement instruction. *Educational Measurement: Issues and Practice*, 10(1), 13-16,26.
- Allen, D. W., Nichols, R. D., & Leblanc, A. C. (1997). The prime teacher appraisal program: 2+2 for teachers. *The High School Magazine*, 4, 30-35.
- Anderson, N. A. (1998). Providing feedback to preservice teachers of reading in field settings. *Reading Research and Instruction*, 37(2), 123-136.
- Archer-Kath, J. (1994). Individual versus Group Feedback in Cooperative Groups. *Journal of Social Psychology*, 134(5), 681-694.
- Arco, L. (1991). Effects of outcome performance feedback on maintenance of client and staff behavior in a residential setting. *Behavioral Residential Treatment*, 6(4), 231-247.
- Bakkenes, I., Brabander, d. C., & Imants, J. (1999). Teacher isolation and communication network analysis in primary schools. *Educational Administration Quarterly*, 35(2), 166-202.
- Balcazar, F. E., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, 7(3-4), 65-89.
- Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. *Personnel Psychology*, 38(2), 335-345.
- Bauwens, J., Hourcade, J. J., & Friend, M. (1989). Cooperative teaching: A model for general and special education integration. *Rase: Remedial & Special Education*, 10(2), 17-22.

- Beerens, D. R. (2000). *Evaluating teachers for professional growth: Creating a culture of motivation and learning*. Thousand Oaks, CA: Corwin Press.
- Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational Leadership*, 57(8), 28-33.
- Black, P., & Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-144.
- Bohlmann, N. L. (1998). Cooperative teaching: a model for teacher collaboration. *Teaching & Change*, 5(3-4), 199-224.
- Bowman, C. L., & McCormick, S. (2000). Comparison of peer coaching versus traditional supervision effects. *The Journal of Educational Research*, 93(4), 256-258.
- Boyd, R. T. C. (1989). *Improving teacher evaluations. ERIC Digest No. 111* (No. RI-88062003). Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation, American Institutes for Research.
- Brandt, R. (1987). On cooperation in schools: A conversation with David and Roger Johnson. *Education Leadership*, 45(3), 14-19.
- Bridges. (1986). Collaborative research: the case of the incompetent teacher. *Teacher Education Quarterly*, 13(2), 59-66.
- Brookhart, S. M., & Loadman, W. E. (1990). School-University Collaboration: Different Workplace Cultures. *Contemporary Education*, 61(3), 125-128.
- Brooks, J. G., & Brooks, M. G. (1999). *In Search of Understanding. The Case for Constructivist Classrooms. [Revised.]*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Center for Educational Research and Innovation. (1998). *Staying ahead: in-service training and teacher professional development*. Paris: OECD.
- China Education and Research Network. (2001). *Education Evolution in China*. Retrieved Feb. 2, 2003, from <http://www.edu.cn/20010906/3000468.shtml>
- Clement, M., & Vandenberghe, R. (2001). How School Leaders Can Promote Teachers' Professional Development: An Account from the Field. *School Leadership & Management*, 21(1), 43-57.
- Cohen, M. (1988). Designing State Assessment Systems. *Phi Delta Kappan*, 69(8), 583-588.
- Cook, L., & Friend, M. (1991). Collaboration in Special Education: Coming of Age in the 1990s. *Preventing School Failure*, 35(2), 24-27.
- Cossairt, A., Hall, R., & Hopkins, B. (1973). The effects of experimenter's instructions, feedback, and praise on teacher praise and student attending behavior. *Journal of Applied Behavior Analysis*, 6(1), 89-100.
- Costa, J. L. d. (1995). Teacher Collaboration: A Comparison of Four Strategies. *Alberta Journal of Educational Research*, 41(4), 407-420.
- CPC Central Committee. (1985). *Reform of China's educational structure: Decision of the CPC Central committee*. Beijing: CPC Central Committee.
- Crandall, D. P., & Loucks, S. F. (1983). *A Roadmap for School Improvement. Executive Summary of the Study of Dissemination Efforts Supporting School Improvement. People, Policies, and Practices: Examining the Chain of School Improvement, Volume X (Research/Technical VIEWPOINTS)*. Andover, MA: Network of Innovative Schools, Inc.

- Dagley, D. L., & Orso, J. K. (1991). Integrating formative, summative modes of evaluation. *National Association of Secondary School Principals Bulletin*, 53(6), 72-82.
- Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies That Support Professional Development in an Era of Reform. *Phi Delta Kappan*, 76(8), 597-604.
- Denzin, N., & Lincoln, Y. (1998). *Collecting and interpreting qualitative materials*: Sage Publications.
- DeRoche. (1987). *An Administrator's Guide for Evaluating Programs and Personnel: An Effective Schools Approach* (Second ed.). Newton, MA: Allyn and Bacon.
- Duncan, P. K., & Bruwelheide, L. R. (1985). Feedback: Use and possible behavioral functions. *Journal of Organizational Behavior Management*, 7(3-4), 91-114.
- Egelson, P. (1994). *Teacher Evaluation Plans That Support Professional Growth*. Paper presented at the Center for Research in Educational Accountability and Teacher Evaluation (CREATE), Kalamazoo, MI.
- Eisner, E. W. (1992). Educational reform and the ecology of schooling. *Teachers College Record*, 93, 610-627.
- Ellett, C. D., & Garland, J. S. (1987). Teacher Evaluation Practices in Our Largest School Districts: Are They Measuring Up to "State-of-the-Art" Systems? *Journal of Personal Evaluation in Education*, 1(1), 69-92.
- Englert, C. S., & Sugai, G. (1983). Teacher Training: Improving Trainee Performance through Peer Observation and Observation System Technology. *Teacher Education and Special Education*, 6(1), 7-17.

- Fleming, R. K., & Sulzer-Azaroff, B. (1989). Enhancing quality of teaching by direct care staff through performance feedback on the job. *Behavioral Residential Treatment*, 4(4), 377-395.
- Forsyth, P. B., & Hoy, W. K. (1978). Isolation and Alienation in Educational Organizations. *Educational Administration Quarterly*, 14(1), 80-96.
- Fox, C. J., & Shirkey, K. A. (1997). Employee Performance Appraisal: The Keystone Made of Clay. In C. Ban & N. M. Riccucci (Eds.), *Public Management* (pp. 205-218). New York: Longman.
- Fullan, M. (1994). *Change Forces: Probing the Depths of Educational Reform*. Bristol, PA: Falmer Press,
Taylor & Francis Inc.
- Fullan, M. (1996). Professional Culture and Educational Change. *School Psychology Review*, 25(4), 496-500.
- Fullan, M. G., & Hargreaves, A. (1991). *What's Worth Fighting For? Working Together for Your School* (Non-Classroom Use No. RP-91-002-008). Andover, MA: Regional Laboratory for Educational Improvement of the Northeast & Islands.
- Fullan, M. G., & Stiegelbauer, S. (1991). *The New Meaning of Educational Change* (Second ed.). New York, NY: Teachers College Press.
- Gable, R. A., & Manning, M. L. (1997). The Role of Teacher Collaboration in School Reform. *Childhood Education*, 73(4), 219-223.
- Gil, D. H. (1987). Instructional Evaluation as a Feedback Process. *New Directions for Teaching and Learning*, 57-64.

- Gillat, A., & Sulzer-Azaroff, B. (1994). Promoting Principals' Managerial Involvement in Instructional Improvement. *Journal of Applied Behavior Analysis*, 27(1), 115-129.
- Goertz, M. E., Floden, R. F., & O'Day, J. (1996). *The bumpy road to education reform (CPRE Policy Brief No. RB-20)*. Paper presented at the Consortium for Policy Research in Education, Philadelphia.
- Goodlad, J. I. (1984). *A Place Called School. Prospects for the Future*. New York, NY: McGraw-Hill Book Company.
- Gordon, S. (2004). *Professional development for school improvement: empowering learning communities*: Pearson Education, Inc.
- Gottesman, B. (2000). *Peer Coaching for Educators*. Lanham, MD: Scarecrow Press/Technomic Books.
- Green, C. W., & Reid, D. H. (1991). Reinforcing Staff Performance in Residential Facilities: A Survey of Common Managerial Practices. *Mental Retardation*, 29(4), 195-200.
- Guskey, T. R. (1998). The Age of Our Accountability. *Journal of Staff Development*, 19(4), 36-44.
- Guskey, T. R. (2000). *Evaluating Professional Development*. Thousand Oaks, CA: Corwin Press, Sage Publications Company.
- Guskey, T. R., & Huberman, M. (Eds.). (1995). *Professional Development in Education: New Paradigms and Practices*. New York: Teachers College Press.
- Harchik, A. E. (1992). Ongoing Consultation as a Method of Improving Performance of Staff Members in a Group Home. *Journal of Applied Behavior Analysis*, 25(3), 599-610.

- Hargreaves, A., & Dawe, R. (1990). Paths of Professional Development: Contrived Collegiality, Collaborative Culture, and the Case of Peer Coaching. *Teaching and Teacher Education*, 6(3), 227-241.
- Hargreaves, A., & Fullan, M. G. (1992). *Understanding Teacher Development*. New York, NY: Teachers College Press.
- Hart, L. (2002). Preservice teachers' beliefs and practice after participating in an integrated content/methods courses. *School Science & Mathematics*(102), 4-14.
- Hawkins, A. M., Burgio, L. D., Langford, A., & Engel, B. T. (1992). The effects of verbal and written supervisory feedback on staff compliance with assigned prompted voiding in a nursing home. *Journal of Organizational Behavior Management*, 13(1), 137-150.
- Holly, M. L. (1982). Teachers' views on inservice training. *Phi Delta Kappan*, 63(417-418.).
- Horn, J. (1999). Personal Renewal and Professional Growth for Teachers: a study of meaningful learning in an interdisciplinary environment. *Teacher Development*, 3(2).
- Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1994). *Collaborative consultation (2nd ed.)*: (1994). xiii, 303pp.
- Idol, L., & West, J. F. (1991). Educational Collaboration: A Catalyst for Effective Schooling. *Intervention in School and Clinic*, 27(2), 70-78.
- Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*: (1989). viii, 253pp.

- Johnson, L. J. (1990). Professional Collaboration: Challenges of the Next Decade. *Teaching Exceptional Children*, 22(2), 9-11.
- Johnson, R. (1976). *Teacher Collaboration, Principal Influence, and Decision Making in Elementary Schools. Technical Report No. 48* (Research/Technical No. SCRDT-TR-48 NE-C-00-3-0062). CA Stanford Center for Research and Development in Teaching: Stanford Univ.
- Joyce, B., & Showers, B. (1981). Transfer of Training: The Contribution of "Coaching." *Journal of Education*, 163(2), 163-172.
- Joyce, B., & Showers, B. (1982). The Coaching of Teaching. *Educational Leadership*, 40(1), 4-8,10.
- Joyce, B., & Showers, B. (1988). *Student Achievement through Staff Development*. White Plains, NY: Longman Inc.
- Kain, D. L. (1996). Looking beneath the Surface: Teacher Collaboration through the Lens of Grading Practices. *Teachers College Record*, 97(4), 569-587.
- Kain, D. L. (1997). Critical Incidents in Teacher Collaboration on Interdisciplinary Teams. *Research in Middle Level Education Quarterly*, 21(1), 1-29.
- Knoblock, & Goldstein. (1971). *The lonely teacher*: Allyn and Bacon, Inc.
- Kohler, F. W. (1997). Effects of Peer Coaching on Teacher and Student Outcomes. *Journal of Educational Research*, 90(4), 240-250.
- Krejcie, R. V., & Morgan, D. W. (1970). Determining Sample Size for Research Activities. *Educational and Psychological Measurement*, 30, 607-610.

- Lambert, L. (1998). How to Build Leadership Capacity. *Educational Leadership*, 55(7), 17-19.
- Lasley, T. J. (1992). Collaborative and Noncollaborative Partnership Structures in Teacher Education. *Journal of Teacher Education*, 43(4), 257-261.
- LeBlanc, A. C. (1997a). *An evaluation of the 2+2 for teachers: alternative performance appraisal program in the Norfolk public schools prime projec*. Old Dominion University, Norfolk, VA.
- LeBlanc, A. C. (1997b). *An evaluation of the 2+2 for teachers: alternative performance appraisal program in the Norfolk public schools prime project*. Unpublished doctoral dissertation, Old Dominion University, Norfolk VA.
- Leggett, D., & Hoyle, S. (1987). Preparing Teachers for Collaboration. *Educational Leadership*, 45(3), 58-63.
- Legters, N. E. (1999). *Teacher Collaboration in a Restructuring Urban High School* (Research/Technical No. 37 R-117-D40005, RED-9452861). Baltimore, MD: Center for Research on the Education of Students Placed At Risk.
- Lehman, L. E. (1989). Practical Motivational Strategies for Teacher Performance and Growth. *Nassp Bulletin*, 73(520), 76-80.
- Lew, M. (1986). Components of Cooperative Learning: Effects of Collaborative Skills and Academic Group Contingencies on Achievement and Mainstreaming. *Contemporary Educational Psychology*, 11(3), 229-239.
- Lieberman, A. (1995). Practices That Support Teacher Development. *Phi Delta Kappan*, 76(8), 591-596.

- Lieberman, A., & Miller, L. (1992). *Teachers--Their World and Their Work: Implications for School Improvement*. Teachers College Press, Teachers College, Columbia University, 1234 Amsterdam Avenue, New York, NY 10027.
- Little, J. W. (1990). The Persistence of Privacy: Autonomy and Initiative in Teachers' Professional Relations. *Teachers College Record*, 91(4), 509-536.
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- Lortie, D. C. (1975). *Schoolteacher; a sociological study*. Chicago, IL: University of Chicago Press.
- Loucks-Horsley, S. (1987). *Continuing to Learn: A Guidebook for Teacher Development* (Descriptive No. 400-86-0005). Oxford, OH: National Staff Development Council.
- Loup, K. S. (1996). Ten Years Later: Findings from a Replication of a Study of Teacher Evaluation Practices in Our 100 Largest School Districts. *Journal of Personnel Evaluation in Education*, 10(3), 203-226.
- Lytle, S. L., & Fecho, R. (1991). Meeting Strangers in Familiar Places: Teacher Collaboration by Cross-Visitation. *English Education*, 23(1), 5-28.
- Marchant, G. J., & Bowers, N. D. (1988). *The Development and Uses of the Teaching Behaviors Questionnaire*.
- McLaughlin, M. W., & Pfeifer, R. S. (1986). *Teacher Evaluation: Learning for Improvement and Accountability and Case Studies* (No. SEPI-86-5, SEPI-86-5A). Stanford, CA: Stanford Education Policy Inst.

- McLaughlin, M. W., & Pfeifer, R. S. (1988). *Teacher Evaluation: Improvement, Accountability, and Effective Learning*. New York, NY: Teachers College Press, Teachers College, Columbia University.
- Mehrens, W. A., & Green, D. R. (1986). *Standardized Tests and School Curricula* (Evaluative/Feasibility).
- Mesch, D., Lew, M., Johnson, D. W., & Johnson, R. (1986). Isolated teenagers, cooperative learning, and the training of social skills. *Journal of Psychology*, 120(4), 323-334.
- Miles, M. B. (1995). Foreword. In M. Huberman (Ed.), *Professional development in education: New paradigms and practices* (pp. vii-ix). New York: Teachers College Press.
- Millman, J., & Darling-Hammond, L. (Eds.). (1990). *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers*. Thousand Oaks, CA: Corwin Press, Sage Publications.
- Miskel, C. (1983). Structural and Expectancy Linkages within Schools and Organizational Effectiveness. *Educational Administration Quarterly*, 19(1), 49-82.
- Moore, J. W., & Schaut, J. A. (1978). Increasing Instructional Effectiveness through the Use of a Problem-Solving Approach to the Design of Instructional Systems. *Journal of Experimental Education*, 47(2), 156-161.
- Morgan, R. L. (1992). Peer Coaching in a Preservice Special Education Program. *Teacher Education and Special Education*, 15(4), 249-258.

- Morgan, R. L., Menlove, R., Salzberg, C. L., & Hudson, P. (1994). Effects of peer coaching on the acquisition of direct instruction skills by low-performing preservice teachers. *Journal of Special Education*, 28(1), 59-76.
- Munro, P., & Elliott, J. (1987). Instructional Growth through Peer Coaching. *Journal of Staff Development*, 8(1), 25-28.
- Natriello, G. (1984). Teachers' Perceptions of the Frequency of Evaluation and Assessments of Their Effort and Effectiveness. *American Educational Research Journal*, 21(3), 579-595.
- Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing Teacher Intervention Implementation in General Education Settings through Consultation and Performance Feedback. *School Psychology Quarterly*, 12(1), 77-88.
- Norton, J. (2001). Grounded in Research. *Journal of Staff Development*, 22(3), 30-32.
- Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, 62(3), 307-332.
- Papanastasiou, E. C. (1999). *Teacher Evaluation: Theories and Practice* (INFORMATION ANALYSES (State-of-the-Art Papers, Research Summaries, Reviews of the Literature on a Topic)).
- Patton, M. Q. (1987). *How To Use Qualitative Methods in Evaluation* (CSE Program Evaluation Kit, Volume 4 Second ed.). Thousand Oaks CA: SAGE Publications Inc.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation: The New Century Text* (3 ed.). Thousand Oaks, CA: SAGE Publications.

- Pelletier, C. M. (1995). *A Handbook of Techniques and Strategies for Coaching Student Teachers. A Guide for Cooperating Teachers, Mentors, College Supervisors, and Teacher Educators*. Needham Heights, MA: Longwood Division, Allyn and Bacon.
- Perkins, D. (1992). *Smart schools : from training memories to educating minds*. New York: Free Press.
- Peterson, K. D. (1987). Teacher Evaluation with Multiple and Variable Lines of Evidence. *American Educational Research Journal*, 24(2), 311-317.
- Peterson, K. D. (1995). *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*. Thousand Oaks, CA: Corwin Press.
- Peterson, P. L., & Comeaux, M. A. (1990). Evaluating the Systems: Teachers' Perspectives on Teacher Evaluation. *Educational Evaluation and Policy Analysis*, 12(1), 3-24.
- Putnam, J. W., Rynders, J. E., Johnson, R. T., & Johnson, D. W. (1989). Collaborative skill instructions for promoting positive interactions between mentally handicapped and nonhandicapped children. *Exceptional Children*, 55(6), 550-557.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of research on teacher education* (pp. 102-119). New York: Simon & Schuster.
- Richardson, V. (2000). *Alexis De Toqueville and the Dilemmas of Professional Development (Descriptive)*: Center for the Improvement of Early Reading Achievement, Ann Arbor, MI.[BBB34978].

- Rock, T. C., & Levin, B. B. (2002). Collaborative Action Research Projects: Enhancing Preservice Teacher Development in Professional Development Schools. *Teacher Education Quarterly*, 29(1), 7-21.
- Rosenholtz, S. J. (1989a). *Teachers' workplace : the social organization of schools*. New York: Longman.
- Rosenholtz, S. J. (1989b). Workplace Conditions that Affect Teacher Quality and Commitment: Implications for Teacher Induction Programs. *Elementary School Journal*, 89(4), 421-439.
- Rothberg, R. A. (1985, November 22-26). *Improving School Climate and Reducing Teacher Isolation*. Paper presented at the 10th Annual Meeting of the National Council of States on Inservice Education, Denver, CO.
- Sarason, S. B. (1995). Some Reactions to What We Have Learned. *Phi Delta Kappan*, 77(1), 84-85.
- Sarason, S. B. (1996). *Revisiting "The Culture of the School and the Problem of Change."* New York, NY: Teachers College Press.
- Schmoker, M. (1999). *Results: The Key to Continuous School Improvement* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Shanxi Education Commission. (1999). *The Commissioners' Report on Educational Reform*. Taiyuan, China.
- Shanxi Education Commission. (2000). *The Commissioners' Report on Educational Reform*. Taiyuan, China.

- Shanxi Research Center for Secondary Education. (1997). *Handbook on Lead Teachers for the 21st Century Shanxi Province Training Program (LTTP) Implementation*. Taiyuan, China.
- Shanxi Research Center for Secondary Education. (2001a). *The 2+2 Program Implementation Handbook*. Taiyuan, China.
- Shanxi Research Center for Secondary Education. (2001b). *Handbook on LTTP Implementation*. Taiyuan, China.
- Shera, W. (1992). Educational Evaluation in China: An Analysis of Current Practices. *Evaluation and Program Planning*, 15(1), 45-53.
- Showers, B. (1984). *Peer Coaching: A Strategy for Facilitating Transfer of Training*. A CEPM R&D Report (No. BBB09086). Washington, DC: Center for Educational Policy and Management at Oregon University.
- Showers, B. (1985). Teachers Coaching Teachers. *Educational Leadership*, 42(7), 43-48.
- Showers, B. (1987). The Role of Coaching in the Implementation of Innovations. *Teacher Education Quarterly*, 14(3), 59-70.
- Showers, B., & Joyce, B. (1996). The Evolution of Peer Coaching. *Educational Leadership*, 53(6), 12-16.
- Shulman, L. S. (1987). Assessment for Teaching: An Initiative for the Profession. *Phi Delta Kappan*, 69(1), 38-44.
- Slater, C. L., & Trowbridge, S. (2000). Master's Level Cohorts Combat Teacher Isolation: University/School District Collaboration. *Action in Teacher Education*, 22(1), 15-22.

- Smith, L. C. (2001). Life history as a key factor in understanding teacher collaboration and classroom practice. *Teacher Education Quarterly*, 28(3), 111-125.
- Sparks, D., & Hirsh, S. (1997). *A New Vision for Staff Development*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Sparks, D., & Hirsh, S. (2000). *A National Plan for Improving Professional Development* (VIEWPOINTS (Opinion Papers, Position Papers, Essays, etc)): National Staff Development Council, Oxford, OH.[BBB21681].
- Sparks, G. M., & Bruder, S. (1987). Before and After Peer Coaching. *Educational Leadership*, 45(3), 54-57.
- State Education Commission. (1996). *Teacher education in China*. Shanghai, China: Shanghai Scientific and Educational Press.
- Stevenson, R. B. (1987). Staff Development for Effective Secondary Schools: A Synthesis of Research. *Teaching and Teacher Education*, 3(3), 233-248.
- Stiggins, R. J. (1986). Classroom Assessment: A Key to Effective Education. *Educational Measurement: Issues and Practice*, 5(2), 5-17.
- Tabachnick, B. G., & Fidell, L. S. (2001). Multivariate analysis of variance and covariance. In *Using multivariate statistics* (2 ed.). Boston: Allyn and Bacon, c2001.
- Thomas, C. C. (1995). *Interactive Teaming: Consultation and Collaboration in Special Programs* (Second ed.): Merrill, Prentice-Hall, Inc., Englewood Cliffs, NJ 07632.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*. New York: Macmillan.

- Wholey, J. S. (1987). Evaluability Assessment: Developing Program Theory. *New Directions for Program Evaluation*(33), 77-92.
- Wilhelms, F. T. (1967). *Evaluation as Feedback and Guide. 1967 Yearbook* (No. FGK04685). Washington, DC: Association for Supervision and Curriculum Development.
- Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis*, 30(4), 693-696.
- Wood, F. (1993). *How To Organize a School-Based Staff Development Program* (Non-Classroom Use): Association for Supervision and Curriculum Development, Alexandria, VA.[BBB30657].
- Yang, D., & Wu, J. (1999). Some issues in the reform and development of teacher education and training in China. *Teacher Development*, 3(2), 157.
- Zhang, D. (1994). The achievement in Chinese teacher education and its future development. *Studies of Higher Teacher Education*, 3, 3-6.

APPENDICES

APPENDIX A: Teacher Professional Performance Scale

- | | | |
|-------------------|------------------------|----------------|
| 1. Unsatisfactory | 2. Below Standard | 3. At Standard |
| 4. Above Standard | 5. Well Above Standard | 6. Superior |

1. PREPAREDNESS FOR INSTRUCTION 1 2 3 4 5 6

- 1.1. Teacher has materials, supplies and equipment ready at the start of the lesson or instructional activity.
- 1.2. Teacher has an instructional plan which is compatible with the school and system-wide curricular goals.

COMMENTS: My comment.

2. MANAGEMENT OF INSTRUCTIONAL TIME 1 2 3 4 5 6

- 2.1. Teacher starts the class on time.
- 2.2. Teacher gets students on task quickly at the beginning of each lesson or instructional activity.
- 2.3. Teacher maintains a high level of student time-on-task.

COMMENTS: My comment.

3. MANAGEMENT OF STUDENT BEHAVIOR 1 2 3 4 5 6

- 3.1. Teacher has a set of rules and procedures that govern the handling of routine administrative matters.
- 3.2. Teacher has a set of rules and procedures that govern student verbal participation and talk during different types of activities--whole class instruction, small group instruction, etc.
- 3.3. Teacher has a set of rules and procedures that govern student movement in the classroom during different types of instructional activities.
- 3.4. Teacher frequently monitors the behavior of all students during whole-class, small group, and seat work activities and during transitions between instructional activities.
- 3.5. Teacher stops inappropriate behavior promptly and consistently, yet maintains the dignity of the student.

COMMENTS: More comments.

4. INSTRUCTIONAL PRESENTATION 1 2 3 4 5 6

- 4.1. Teacher attracts students attention.
- 4.2. Teacher introduces the lesson or instructional activity and specifies learning objectives when appropriate.
- 4.3. Teacher speaks fluently and precisely.
- 4.4. Teacher presents the lesson or instructional activity using concepts and language understandable to the students.
- 4.5. Teacher provides relevant examples and demonstrations to illustrate concepts and skills.
- 4.6. Teacher assigns tasks that students handle with a high rate of success.
- 4.7. Teacher asks appropriate levels of questions that students handle with a high rate of success.
- 4.8. Teacher conducts lesson or instructional activity at a brisk pace, slowing presentations when necessary for student understanding but avoiding unnecessary slowdowns.
- 4.9. Teacher makes transitions between lessons and between instructional activities within lessons efficiently and smoothly.
- 4.10. Teacher makes sure that the assignment is clear.
- 4.11. Teacher summarizes the main point(s) of the lesson at the end of the lesson or instructional activity.

COMMENTS:

5. MONITORING OF STUDENT PERFORMANCE 1 2 3 4 5 6

- 5.1. Teacher maintains clear, firm and reasonable work standards.
- 5.2. Teacher moves around during classwork to check all students' performance.
- 5.3. Teacher routinely checks student progress, and knows what students have learned.

COMMENTS:

6. PROVIDING REINFORCEMENT AND FEEDBACK 1 2 3 4 5 6

- 6.1. Teacher provides feedback on the correctness or incorrectness of in-class work to encourage student growth.
- 6.2. Teacher regularly provides prompt feedback on assigned out-of-class work.
- 6.3. Teacher affirms a correct oral response appropriately, and moves on.
- 6.4. Teacher provides sustaining feedback after an incorrect response or no-response by probing, repeating or rephrasing the question, giving a clue, or allowing more time.
- 6.5. Teacher gives positive feedback to reinforce whenever necessary.

COMMENTS:

7. FACILITATING INSTRUCTION 1 2 3 4 5 6

- 7.1. Teacher encourages students to think and work independently
- 7.2. Teacher uses diagnostic information obtained from tests and other assessment procedures to develop and revise objectives and/or tasks.
- 7.3. Teacher maintains accurate records to document student performances.
- 7.4. Teacher has instructional plan that matches/aligns objectives, learning strategies, assessment and student needs at the appropriate level of difficulty.
- 7.5. Teacher uses available human and material resources to support the instructional program.

COMMENTS:

8. COMMUNICATING WITH STUDENTS 1 2 3 4 5 6

- 8.1. Teacher treats all students in a fair and equitable manner.
- 8.2. Teacher interacts effectively with students.

COMMENTS:

9. CHALK BOARD SKILL 1 2 3 4 5 6

- 9.1. Teacher writes clearly and correctly on the board.
- 9.2. Teacher has the main points on the board.
- 9.3. Teacher has a design of chalk-board layout.

COMMENTS:

SUMMARY COMMENTS

Evaluator's Summary Comments:

Evaluator's Signature	Title	Date
Teacher's Signature		Date

Signature indicates that the written evaluation has been seen and discussed.
Levels of Performance

6. SUPERIOR

Performance within this function area is consistently outstanding.
Teaching practices are demonstrated at the highest level of performance.

Teacher continuously seeks to expand scope of competencies and constantly undertakes additional, appropriate responsibilities.

5. WELL ABOVE STANDARD

Performance within this function area is frequently outstanding.

Some teaching practices are demonstrated at the highest level while others are at a consistently high level. Teacher frequently seeks to expand scope of competencies and often undertakes additional, appropriate responsibilities.

4. ABOVE STANDARD

Performance within this function area is frequently high. Some teaching practices are demonstrated at a high level while others are at a consistently adequate/acceptable level. Teacher sometimes seeks to expand scope of competencies and occasionally undertakes additional, appropriate responsibilities.

3. AT STANDARD

Performance within this function area is consistently adequate/acceptable. Teaching practices fully meet all performance expectations at an acceptable level. Teacher maintains an adequate scope of competencies, and performs additional responsibilities as assigned.

2. BELOW STANDARD

Performance within this function area is sometimes inadequate/unacceptable and needs improvement. Teacher requires supervision and assistance to maintain an adequate scope of competencies, and sometimes fails to perform additional responsibilities as assigned.

1. UNSATISFACTORY

Performance within this function area is consistently inadequate/unacceptable and most practices require considerable improvement to fully meet minimum performance expectations. Teacher requires close and frequent supervision in the performance of all responsibilities.

APPENDIX B: Teachers' Pedagogical Belief and Attitude Scale

Scoring: 5 = Strongly Agree; 4 = Agree; 3 = Undecided; 2 = Disagree; 1 = Strongly Disagree

1. Teachers are life long learners.
2. A teacher should be active in participating educational reform initiatives and experiments.
3. A teacher's instructional practice can always be improved.
4. Teachers should keep in touch with the students' parents and establish a close working relationship with them.
5. Teachers should teach thinking skills and interpersonal skills along with content knowledge.
6. The teacher should monitor the progress of each student daily.
7. The teacher should spend most of the class time talking to the students.
8. The teacher should wait at least 3 seconds after asking a question in class before calling on a student.
9. The teacher should expect that every student is to succeed.
10. The teacher should move around the classroom during the time that the students are working independently at their seats.
11. The teacher should overtly acknowledge correct answers that a student gives to a question asked in class.
12. The teacher should criticize the student when he/she gives a wrong answer to a question asked in class?

13. The teacher should spend time at the beginning of the school year teaching classroom rules and procedures.
14. The teacher should read the grades aloud in class after an exam and make it know in class how each student is ranked?
15. Moral education is none of the teacher's business?
16. It is a shame for the teacher if he/she does not know something that the students know?
17. Students are too young to make any decisions for themselves?
18. Your colleagues are always your competitors?
19. It is ideal if the teacher can always be left alone to do the job?
20. The teacher should not ask the students to comment the effectiveness of his/her teaching?

APPENDIX C: 2+2 Program Response Survey

Dear 2+2 project participants:

As you may have been experienced, the essence of the 2+2 Alternative Teacher Performance Appraisal Protocol is a series of regular classroom observations by teachers and administrators. The observer visits a classroom and makes two compliments and two suggestions for improvement or change. The purpose of the 2+2 protocol is straightforward. It is set up to maximize professional interactions, decrease teacher isolation, and increase meaningful feedback that will lead to improved instruction. Thanks.

Background Information

1. Gender: ☐ Male ☐ Female
2. Education Background: ☐ College diploma ☐ Bachelor's ☐ other
3. Age: ☐ 21-26 ☐ 27-31 ☐ 32-40 ☐ 41-49 ☐ 50+

Teaching History

Hefei8

1. Years of Teaching:
 - ☐ 0-3 years
 - ☐ 4-10 years
 - ☐ 11 years or more
 2. Subject Area:
 - ☐ English
 - ☐ Math
 - ☐ Physics
 - ☐ Chemistry
 - ☐ Social Studies
-

Information about observation and feedback:

The following questions are about your experience of observing or being observed. Please note that 5 indicates more frequent. Please check the one you think appropriate.

How often do you observe other teachers teaching in one month? _____

How often do other teachers observe your teaching in one month? _____

How often does the principal observe your teaching in one month? _____

How often do you have students give you feedback on your teaching in one month?

How often do you have parents visit your classroom in a month? _____

How often do you have parents give you feedback on your teaching in one month?

How much feedback did you get from your administrators last month? _____

How much feedback did you get from your colleagues last month? _____

Assessment of 2+2 Program observation feedback form.

The following questions are about your assessment of the 2+2 observation, especially the quality of the compliments and the suggestions. Please note that 5 indicates a high quality. Please circle the one you think appropriate.

1. How would you rate the quality of 2+2 compliments you have received?

(High) (Low)

5 4 3 2 1

2. How would you rate the quality of 2+2 suggestions you have received?

(High) (Low)
5 4 3 2 1

Perception about participating 2+2 program observation:

The following questions are about you feeling of participating 2+2 program. Please check the one you think appropriate.

1. Did you ever hear of the “2+2” appraisal system before the project?

Yes no

2. Did you ever do anything similar to this?

Yes no

3. Do you feel more comfortable to be observed?

Yes no

4. Do you feel more comfortable to observe?

Yes no

5. Do you think “2+2” too simple to undertake?

Yes no

6. Do you think “2+2” too complicated to undertake

Yes no

Peer interaction and collaboration

The following questions are about the frequency of interactions and collaborations between you and your peers in one month. Please fill in the blank with the number you think appropriate.

1. I discuss instruction-related topics with my peers. _____
2. I prepare lessons with my colleagues . _____
3. I ask my colleagues for assistance. _____
4. My colleagues ask me for assistance. _____
5. My colleagues come up to discuss instruction-related topics with me. _____

Thank you for your participating!

APPENDIX D: 2+2 Feedback Form

Teacher _____ School _____
Class _____ Subject _____ # of Students _____
Observer _____ Date _____

Compliment

1. _____

2. _____

Suggestions

1. _____

2. _____

Reflections

APPENDIX E: Focus Group Question Protocol

Introduction: Thank you everyone for being here and participating. This is an opportunity for us as a group to reflect on the process we're going through in implementation of 2+2 program. I have some broad discussion questions for you, but I encourage you to share all your insights on this process and what you're learning, even if it seems tangential to the issues. Hopefully some new insights will emerge from discussing these issues together. The interview will be audio-taped and transcribed later, and I will be analyzing it to learn about your views on these issues.

- 1) How do you describe the implementation of the 2+2 program?
 - a) Easy part? Hardship?
- 2) How do you compare "2+2" with the traditional teacher performance appraisal system?
 - a) Which one do you think it better?
- 3) What were the strengths of "2+2"?
 - a) Strength of compliments?
 - b) Strength of suggestions?
- 4) What were the weaknesses of "2+2"?
 - a) Weakness of compliments
 - b) Weakness of suggestions
- 5) What progress have you made with "2+2"?
 - a) Pedagogical progress
 - b) Performance progress
 - c) Personal progress

- 6) How did you benefit from it?
 - a) What kind of benefits?
- 7) What were your reactions to being observed?
 - a) First reaction?
 - b) Get used of it?
- 8) What were your reactions to being an observer?
 - a) First reaction?
 - b) Later on? At the closure?
- 9) What is the next step in your plan for “2+2”?
 - a) Do you have any plan?
 - b) What are they?
 - c) Is there anything other than 2+2?
- 10) What else can you tell me about 2+2 that would help me understand its impact?
 - a) Any comment is welcome.

VITA

Weiping Wang

Education

- 2005** PhD Urban Services/ Education, Old Dominion University
ECI Department, Rm. 145, Darden College of Education, Old Dominion
University, Norfolk, VA, 23529
- 1990** MA Linguistics
Beijing China Beijing Foreign Studies University, Beijing China
- 1985** BA English Language
Beijing Foreign Studies University, Beijing China

Professional Experience

- 2003-Date** Executive Vice President, Shanxi Insitutite of Technology, Jincheng, China
- 1990 - 2003** Associate Professor and Director of Research Center for Secondary School
Education, Shanxi Institute of Education, Taiyuan, China
- 1997 - 1998** Principal, Dongda High School, Taiyuan, China

Publications

Books:

- Creativity and Synthesis (2001), Shanxi Shuhai Publishing House, P.R. China
- Modern Educational Theories (1998), Beijing Normal University Publishing House, P. R. China
- Microteaching (1996, co-authored with Dwight W. Allen), Xinhua Publishing House, P. R. China
- English Reading Comprehension for Trainee Teachers (1995), Shanxi Higher Education
Publishing House, P. R.. China
- English test Questions for Junior High School Students (1988), Shanxi Provincial Publishing
House, and P. R. China

Articles and Papers:

- Instructional Fore-play: A New Model for School-based Professional Development and Teaching
Research, Theory and Practice of Education (November, 2004), P. R. China
- Education Reform Waves in 20th Century American Education, Instruction and Administration
(July, 1997), P. R. China
- One Way to Improve Teacher Training, Teacher Training (April, 1996), P. R. China
- Cooperative Learning in Classroom, Theory and Practice (June, 1995), P. R. China