



## 2+2 Program for Teachers' Performance Appraisal in China

Shuli Zhao

Foreign Language Department, Huanghuai University

Zhumadian 463000, China

E-mail: zsl7700@163.com

### Abstract

This study examined the impact of the 2+2 Alternative Teacher Performance Appraisal System that has been implemented in Shanxi province in China. A mixed research design was used to evaluate the program. Six high schools and a total of 78 teachers (13 teachers in each school) in Shanxi province were selected. Three of the schools participated in the 2+2 program while another three served as the comparison. The results showed that 2+2 program significantly improved teachers' professional performance, enhanced teachers' collaboration, and increased the feedback between the peers.

**Keywords:** Teacher performance, Teacher evaluation, Collaboration, Feedback, Teacher education

### 1. Introduction

#### 1.1 Educational system reform in China

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-90s, 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, performances, and competencies have not been changed much as has been expected by those educational reformers on all levels (Shanxi Research Center for Secondary Education, 2001).

#### 1.2 School culture and 2+2

School reform cannot succeed without changing the school culture. Researchers (Eisner, 1992; Fullan (1994) and Fullan (1996); Sarason (1995) and Sarason (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 one 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. No system exists for peer support in pursuing professional growth and instructional improvements. The 2+2 Alternative Teacher Performance Appraisal System (2+2) is designed to help change the current school culture reflected in teacher isolation, and build a positive and productive relationship among teachers (LeBlanc, 1997). The 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.

#### 1.3 Teacher performance evaluation and 2+2

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 non-discriminating that makes it non-significant in helping teachers improve their job performance (Shera, 1992).

#### 1.4 The 2+2 alternative teacher performance appraisal program in Shanxi

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 that is being pursued consistently. Serious consideration is given to issues concerning educational reform and school improvement (Shanxi Education Commission, 2000). In order to help those inexperienced 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. One major initiative of the province's educational reform package is the 2+2 Alternative Teacher Performance Appraisal Program (2+2).

##### 1.4.1 Background of the 2+2 program

The 2+2 protocol was first developed by Dwight 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, 1997).

The purpose of the 2+2 is straightforward. It is designed to maximize professional interactions, decrease teacher isolation, and increase meaningful feedback that will lead to improved instructional performance (Shanxi Research Center for Secondary Education, 2001). The essence of the 2+2 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 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. Teachers need frequent feedback to grow professionally. The 2+2 appraisal system was designed to provide more opportunities for teachers to give and receive feedback, because multiple feedback from peers will assist teacher in gaining an appreciation for innovative and diverse approaches used by other teachers (Beerens, 2000).

The 2+2 program is an experimental alternative to the province's teacher performance appraisal system, in that in most Chinese schools, 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, 2001). Based on frequent peer and administrator observation, the 2+2 program was developed to provide more frequent, less formal feedback to teachers. The protocol was designed to help reduce teacher isolation and increase feedback, hence to foster a collaborative culture that will lead to an exchange and implementation of successful instructional strategies and better performance.

##### 1.4.2 Purpose and research question

The purpose of this study was to examine the impact of the 2+2 Alternative Teacher Performance Appraisal System on the teachers' performance in classroom and teachers' collaboration. Research questions included:

- (1) How has the program impacted on teachers' professional performance?
- (2) How has the program impacted on teacher's collaboration?
- (3) What kind of feedback was provided to teachers who participate in the 2+2 program?
- (4) How did teachers compare "2+2" with the traditional teacher performance appraisal system?

## 2. Methods

### 2.1 Research design

This study employed a quasi-experimental design in which six key urban high schools were selected by the Central Office for the Program's Implementation from 43 provincial key high schools and randomly assigned to either the 2+2 (intervention) group or the comparison group. The research questions were addressed by employing both quantitative and qualitative approaches.

### 2.2 Setting

There are 9988 schools located in urban and rural areas in the Shanxi 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 3000 students with a mean of 1668 (Shanxi Education Commission, 1999). There were 43 provincially nominated key high schools because they all met the following requirements and standards set by the provincial government in 1983: (1) all teachers must have a bachelor or equivalent degree; (2) the school must have an enrolment 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 m 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).

### 2.3 Sample

Non-random sampling selection was employed. Six urban high schools were selected by the Central Office for the Program's Implementation from the 43 provincial key high schools to participate in the program. These schools were selected because they shared 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 2000, which are very much like the other provincially nominated key schools. Each of those six project schools has 13 first grade (equivalent to 10th grade in the United States) teachers including 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 faculty. Among these six schools, three were randomly assigned to the 2+2 group, which resulted in 39 first grade teachers participating in the 2+2 program. The other three schools (39 teachers) still maintained their traditional teacher evaluation and appraisal system.

### 2.4 Measures

#### 2.4.1 Teacher professional performance

Teacher professional performance was defined as a teacher's demonstration of skills or competency in class with an emphasis on teachers' ability to perform instructional tasks. In the current study, teacher performance was measured by Shanxi Teachers' Performance Measurement Scale (Shanxi Research Center for Secondary Education, 1997). 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 Lead Teachers for the 21st Century Shanxi Province Training Program (LTTP) candidates (Shanxi Research Center for Secondary Education, 1997). It has been used by most of the school districts in Shanxi since then to appraise their teachers' professional performance. 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 (Shanxi Research Center for Secondary Education, 1997).

#### 2.4.2 Teacher collaboration

Peer interaction and collaboration was measured by five open-ended questions in the 2+2 Program Response Survey. The survey was developed by the researcher based on the 2+2 survey created by LeBlanc (1997) to investigate how the 2+2 program has been implemented and how the participating teachers perceive the program. The five questions were designed to inquire the frequency of interactions and collaborations between the teacher and his/her peers in 1 month prior to the survey. Questions content included frequency of discussion regarding instruction-related topics with peers; frequency of preparing lessons with colleagues, frequency of asking colleagues for assistance; frequency of colleagues asking for assistance, and frequency of colleagues coming up to discuss instruction-related topics.

#### 2.4.3 Teachers' experience of 2+2

To gather complementary information regarding teachers' perceptions, expectations, and evaluation of the 2+2 program, structured interviews were conducted in winter 2002, with the 39 participants of the program. This information regarding to teachers' experience of 2+2 was collected by asking two of the 10 questions conducted in the interview. This two open-ended questions asked "how do you compare "2+2" with the traditional teacher performance appraisal system?" and "how did you benefit from 2+2 program?".

### 2.5 Administration of measures

This study was reviewed and approved by University's Institutional Review Board. Teacher's professional performance was assessed prior to (September 2001) and after (October 2002) the implementation of the program. The central office of the program hired five external professional evaluators to observe all the 78 participants' classroom teaching and evaluate their performance level. The evaluators were trained to use Teacher Performance Rating Scale to assess teachers' performance in classroom. Before the class began, the evaluators entered the classrooms without advance notice to the teacher. During the observation, the evaluators were required to remain quiet and as less intrusive as possible. At the end of the class, each evaluator completed the assessment individually and returned it in a sealed envelop to the principal of each school after the completion. Teacher's collaboration and interaction questions were

completed by all the teachers prior to (September 2001) and after (October 2002) the implementation of the program. The surveys along with instructions were distributed to each of the teachers in a sealed envelop either by mail or in-person. The principal of each school was responsible for collecting the completed surveys and returning them to the program manager. All the completed surveys were kept in sealed envelopes. Teachers' experience of 2+2 program was assessed by focus group interviews. Interviews were conducted by the current researcher with his assistants in the three high schools participating in the 2+2 program in October 2002. Using a semi-structured interview protocol, the researcher arranged three meetings with the participating teachers, one meeting in each of the three intervention group schools, at the conclusion of the program to discuss their experience of implementing 2+2 program. The duration of the focus group interviews ranged from 2 to 3 h. These interviews were audio-taped and transcribed.

## 2.6 Data analysis

### 2.6.1 Quantitative data

Descriptive analysis was used to examine the frequencies, distribution, central tendency, and dispersion for each of the variables. Analysis of covariance (ANCOVA) was employed to compare posttest scores of the intervention group and comparison group controlling for the pretest scores. The main independent variable was the group membership (2+2 intervention or comparison group), while the dependent variables were teacher performance scores, frequency of feedback, and frequency of teacher collaboration practice. Correlation analyses were performed to examine the relationships between number of feedback received and teacher performance.

### 2.6.2 Qualitative data

Content analysis was employed to analyze the compliments and suggestions the teachers had provided on the 2+2 observation forms. Purposive sampling was used to draw sample from the 3314 collected forms. Altogether 350 forms were selected by teachers' teaching major, year of teaching and gender. A process of categorizing and/or labeling of the 2+2 compliments and suggestions across cases were utilized. Compliments and suggestions were analyzed separately.

Compliments and suggestions were tentatively assigned to a category. As compliments/suggestions were found unfit in a category, a new category or subcategory was 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 focus group interview. 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.

Insert Table 1 Here

## 3. Results

### 3.1 Teachers' characteristics

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 40 years old or younger and 12 (28.2%) aged 41 years or older. Forty (51.3%) had 3 years or less teaching experience, 19 (24.4%) had 4–10 years teaching experience, and another 19 (24.4%) had 11 years or more teaching experience. No statistically significant differences in terms of the change of teachers' performance and collaboration prior to and after the program were observed among gender, age, and teaching experience groups (all  $p > 0.05$ ).

### 3.2 Program impact on professional performance

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 ( $p < 0.001$ , Table 1). The pretest total score of professional performance for the 2+2 group was 154.41 (SD=23.78) and the posttest score 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 ( $p < 0.05$ ). 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  $0.05/9 = 0.0056$ ). After the adjustment, the differences remained statistically significant ( $p < 0.0056$ ). The mean scores on each of the nine functions were obtained by dividing the total scale score by the number of items of the scale. The descriptive statistics by subscale are presented in Table 1. 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.

### 3.2.1 2+2 visitations and professional performance

Results show that 2+2 classroom 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 ( $r^2=0.35$ ,  $r=0.592$ ,  $p<0.01$ ). The more visitations a teacher had made, the more improvement had been found in his/her teacher performance.

### 3.2.2 Teachers' perceived benefit of 2+2 on performance

The teachers mentioned various benefits that they perceived from the 2+2 program for their performance. A majority of the 38 teachers participated in 2+2 indicated that 2+2 program benefited their performance by providing more chance to observe other teachers' performance (90% of the teachers), more opportunities to learn from other teachers (80%), and more opportunities to discuss instructional affairs with colleagues (60%).

### 3.3 Program impact on teacher collaboration

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 2). 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  $0.05/5=0.01$ ). After the adjustment, the scores of all the collaboration categories remain significantly higher in 2+2 group than in the comparison group ( $p<0.001$ ).

### 3.4 Feedback provided to teachers who participate in the 2+2 program

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 from the 350 forms were available for analysis, each of which was assigned to a category and recorded on a coding form. Aggregate results were calculated and are represented in Table 3. 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%).

Insert Table 2 Here

Insert Table 3 Here

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 as 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.

### 3.5 Teachers' comparison of "2+2" with traditional teacher performance appraisal system

The majority (60%) of the teachers in 2+2 group expressed their strong preference of the 2+2 to the traditional teacher performance evaluation system. A typical response is like a math teacher's statement below: 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. One teacher stated: 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 evaluate 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. It was depicted that 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.

#### **4. Discussion**

##### *4.1 Professional development*

The findings 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. 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. 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. 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; 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". (Gordon, 2004). The 2+2 program shares many of the characteristics identified as for 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, and 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 result of observation of each other's work, and the collaboration of peers. However, note that the performance improvement was observed right after the completion of the program, whether the change will sustain in the long run is still a question. Moreover, other factors such as knowledge, beliefs, attitudes, and intentions should be taken under consideration if 2+2 intends to serve more than an appraisal system and help teachers to improve their professional performance.

##### *4.2 Teachers' collaboration*

The teachers in the 2+2 group experienced collaboration much more than the comparison group across all of the pertaining categories after the program implementation. The implementation of 2+2 represented a fundamental change in the way the teachers interacted with colleagues.

Teamwork develops through observation and communication (LeBlanc, 1997). 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). 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 (Hudak, Hogg-Johnson, Bombardier, McKeever, & Wright, 2004). 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, 1989), 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's teaching, then giving and receiving feedback.

#### 4.3 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 responses. It is recommended that the program should develop certain component to address the age/experience difference between teachers.

### 5. 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. It calls attention to the teachers' collaboration, peer visitations, and feedback and their influences on teachers' professional performance.

### References

- D.R. Beerens. (2000). *Evaluating teachers for professional growth: Creating a culture of motivation and learning*, Corwin Press, Thousand Oaks, CA.
- M. Fullan. (1994). *Change forces: Probing the depths of educational reform*, Falmer Press, Taylor & Francis Inc., Bristol, PA.
- Gordon S. Gordon. (2004). *Professional development for school improvement: Empowering learning communities*, Pearson Education, Inc.
- Guskey T.R. Guskey. (1998). The age of our accountability, *Journal of Staff Development* **19** (4) pp. 36–44.
- Holly M.L. Holly. (1982). Teachers' views on inservice training, *Phi Delta Kappan* **63** (417–418) .
- Kain D.L. Kain. (1996). Looking beneath the surface: Teacher collaboration through the lens of grading practices, *Teachers College Record* **97** (4), pp. 569–587.
- LeBlanc, A. C. (1997). *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 & Hoyle. (1987). D. Leggett and S. Hoyle, Preparing teachers for collaboration, *Educational Leadership* **45** (3), pp. 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.
- J.W. Little. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations, *Teachers College Record* **91** (4), pp. 509–536.
- Lytle & Fecho. (1991). S.L. Lytle and R. Fecho, Meeting strangers in familiar places: Teacher collaboration by cross-visitation, *English Education* **23** (1), pp. 5–28.
- M.B. Miles. (1995). Foreword. In: T.R. Gusket and M. Huberman, Editors, *Professional development in education: New paradigms and practices*, Teachers College Press, New York, pp. vii–ix. motivation and learning, Corwin Press, Thousand Oaks, CA. (2000).

J. Norton. (2001). Grounded in research, *Journal of Staff Development* **22** (3), pp. 30–32.

C.M. Pelletier. (1995). A handbook of techniques and strategies for coaching student teachers. A guide for cooperating teachers, mentors, college supervisors, and teacher educators, Longwood Division, Allyn and Bacon, Needham Heights, MA.

Shanxi Education Commission Shanxi Education Commission. (1999). *The commissioners' report on educational reform*. Taiyuan, China.

Shanxi Research Center for Secondary Education Shanxi Research Center for Secondary Education. (2001). *Handbook on LTP implementation*. Taiyuan, China.

D. Sparks and S. Hirsh. (2000). A national plan for improving professional development (VIEWPOINTS (Opinion Papers, Position Papers, Essays, etc.)), National Staff Development Council, Oxford, OH [BBB21681].

Table 1. Comparison of the teachers' professional performance

	Items	2+2 group			Comparison group		
		Pretest Change	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	30.7	152.57	147.85	-4.72
Mean score <sup>a</sup>		3.96	4.75	0.79	3.91	3.79	-0.12

\* $p < 0.01$ . \*\* $p < 0.001$ . Mean 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.

Table 2. Comparison of collaboration in 1 month ( $n=39$ )

In 1 month	2+2 group		Comparison	
	group			
	Pre	Post	Pre	
	Post		Mean	
	Mean	Mean	Mean	Mean
	SD	SD	SD	SD
I discuss instruction-related topics with my peers.*	1.81	7.62	2.10	2.31
	1.09	1.25	0.60	0.61
I prepare lessons with my colleagues.	1.75	10.86	1.92	2.05
	0.65	1.32	0.74	0.83
I ask my colleagues for assistance.*	1.69	10.95	1.38	1.51
	1.51	1.62	0.49	0.56
My colleagues ask me for assistance. *	1.44	4.56	1.54	1.56
	0.84	0.99	0.55	0.60
My colleagues come up to discuss instruction-related topics with me.	2.03	8.82	2.28	3.00
	0.56	1.27	0.60	0.92

\*ANCOVA ( $p<0.001$ )

Table 3. Responses in compliments and suggestions categories

Categories	Compliments		Suggestions	
	<i>n</i>	%	<i>n</i>	%
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