

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

<u>DATE DUE</u>	<u>DATE DUE</u>	<u>DATE DUE</u>

THREE ESSAYS ON SCHOOL CHOICE AND ACCOUNTABILITY POLICY

By

Kwanghyun Lee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Education Policy
College of Education

2004

ABSTRACT

THREE ESSAYS ON SCHOOL CHOICE AND ACCOUNTABILITY POLICY

By

Kwanghyun Lee

This dissertation provides empirical studies on the school choice and accountability policy, consisting of three independent essays. Chapter I evaluates the competitive effect of charter schools on hosting school districts using Data Envelopment Analysis (DEA), specifically a super-efficiency DEA model, which has been used in the field of operations research to measure organizational efficiency. The empirical work is based on Michigan, using district-level school finance and Michigan Educational Assessment Program score data. The results of analysis show that charter hosting districts improved their efficiency over time more than other school districts. However, this difference in efficiency is not statistically significant. Further analysis using first differencing regression also confirms that the change in efficiency in charter-hosting districts is not significantly affected by the share of local charter school enrollment.

Chapter II examines two competing perspectives regarding the impact of standards-based accountability policy on teachers' instructional practice using the 2000 National Center for Educational Statistics (NCES) School and Staffing Survey (SASS). The first perspective is derived from the behavioral predictions or "theory of action" put forth by accountability policy advocates. According to this view, a properly designed state-level accountability system will induce teachers to use state curriculum guidance and students' test results to modify their instructional practice in order to improve

a

s

a

te

m

ag

un

of

te

m

m

f

te

h

student performance. The second perspective is derived from the Cohen's cognitive theory that teachers' opportunities to learn will be the main determinant of policy implementation or change of teachers' instructional practice. Analysis of the 2000 SASS database using instrumental variables estimate and ordered probit and logit models shows that both accountability policy and teachers' opportunities to learn measured by professional development participation lead to teachers' more frequent use of state/district standards. However, the explanatory power of teachers' opportunity to learn is significantly larger than that of accountability policy. Furthermore, accountability policy produces an unintended consequence: teachers from states with strong accountability policies are more likely to group students within the classroom by achievement or ability level, which is discouraged by the state standards.

Chapter III examines the psychological effect of accountability policy on teachers. Self-determination theory (SDT) in psychology states that controlling mechanisms such as performance-contingent rewards and threat of sanction will turn agents' perceived locus of causality from inwardly to externally directed, thus undermining agents' intrinsic motivation. SDT implies that teachers under the pressure of strong accountability policy will be more likely to lose their intrinsic motivation for teaching. Analysis of the 2000 NCES SASS database using ordered probit and logit models confirms that teachers under strong accountability policy are more likely to mention that they would not become a teacher again if they were to start over, controlling for other variables. They also respond that it is a waste of time to try to do one's best as a teacher, which shows that teachers working under the strong accountability policy are losing intrinsic motivation.

ACKNOWLEDGEMENTS

Writing a dissertation is a challenging task that demands new insights and creativities. However, I could not have finished my dissertation without my great fortune: meeting my adviser, Dr. David Arsen, and dissertation committee members, Dr. David Plank, Dr. Gary Sykes, and Dr. Jeffrey Wooldridge. They all encouraged me to investigate my dissertation research topic. I will remain indebted to them throughout the remainder of my career.

My adviser, Dr. Arsen, sacrificed countless hours reading and correcting drafts, and carefully guided and inspired me through each step of my research. Without his help, it would not have been possible for me to go forward with my research topic. It was a great pleasure to discuss my research topic with Dr. Arsen. His advice inspired and motivated me so I was able to go through all dissertation steps.

Other two committee members, Dr. Sykes and Dr. Plank provided profound insights on the theoretical framework of my dissertation. Their critical comments on the theoretical framework contributed to my study and improved my dissertation.

Dr. Wooldridge helped me to understand and improve the research methodology of econometrics. I am deeply indebted to him for his invaluable comments on the methods I used for my dissertation. I enhanced my knowledge on econometrics because of him. I would never forget the help of my adviser and my dissertation committee members.

Also, I am very grateful for my colleagues, Marisa Burian-Fitzgerald and Debbi Harris, who helped me to manage the NCES SASS database. They helped me to obtain some very useful variables.

I would never have completed my doctoral study without the support of my parents, In-Chae Lee who is my father and Sun-Jin Kim who is my mother. Without their support, I would never think about starting my doctoral study. This small achievement of mine is truly theirs and dedicated to them for their pride, pleasure and happiness.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

DO CHARTER SCHOOLS SPUR IMPROVED EFFICIENCY IN TRADITIONAL PUBLIC SCHOOLS? AN EVALUATION OF MICHIGAN SCHOOLS USING DATA ENVELOPMENT ANALYSIS

1. Introduction	1
2. Michigan Charter School Studies	3
3. Methodology	4
4. Data	11
5. Results of Analysis	13
5.1 Basic Results and Chi-Square Tests	13
5.2 Difference-in-Difference Estimate and Robust Regression Analysis	17
6. Discussion and Concluding Remarks	21
APPENDIX	
1. Number of Charter School by Locale Code in Michigan	25
2. Scatter Plot, Change in Log of Efficiency by Change of Charter School Enrollment Share in District	25
3. Scatter Plot, Change in Efficiency by Change of Charter School Enrollment Share in District	26
REFERENCES	27

CHAPTER II

ACCOUNTABILITY POLICY OR OPPORTUNITY TO LEARN? THE IMPACT SIZE OF ACCOUNTABILITY AND PROFESSIONAL DEVELOPMENT ON INSTRUCTIONAL PRACTICE – EVIDENCE FROM 2000 NCES SASS DATABASE

1. Introduction	29
2. The Principal-Agent Model and Theory of Action	32
3. Pedagogy of Policy and Capacity of Teachers and Schools	35
4. Current Literature on the Effect of Accountability Policy	37
4.1 Effect of Accountability Policy on Student Academic Achievement	37
4.2 Effect of Accountability Policy on Classroom Instruction	43
5. Research Design	49
6. Data	53
7. Methods	55
7.1 Regression Model and Independent Variables	55
7.2 Self-Selection Problem and Quality of Professional Development Program	61
7.3 Dependent Variables and Summary Statistics	65
8. Results	69

8.1 The Effect of the Teachers' Use of State/District Standards	69
8.1.1 Specification and Change of the Coefficient Size	70
8.1.2 Increase of Sample Size by Dropping a Few Control Variables	73
8.1.3 Checking Instrumental Variables Estimation	79
8.2 The Effect on the Teachers' Use of Information from State or Local Achievement Tests	85
8.2.1 Teachers' Use of Student Test Score to Strengthen Their Content Knowledge and Teaching Practice	85
8.2.2 Teachers' Use of Student Test Score to Adjust Their Curriculum in Areas Where Their Students Encountered Problems	89
8.2.3 Teachers' Use of Student Test Score to Group Students into Different Instructional Groups by Achievement or Ability	92
9. Conclusion	95
APPENDIX	
1. Accountability Index, by State, 1999-2000	97
2. Number of Sample on Full-Time Teachers by State	100
3. Questionnaires of Public Teacher Survey on the Use of State or District Standards and Student Test Score, Which are Employed as Dependent Variables in Analysis	100
4. Ordered Probit Regression and Ordered Logit Regression Model	101
5. Professional Development Program Participation Rates by State	102
6. Questionnaires Used to Make the Number of Professional Development Programs for School or District Administrators	103
7. Questionnaire on the Usefulness of Professional Development Programs	104
REFERENCES	105

CHAPTER III

DOES ACCOUNTABILITY POLICY DIMINISH TEACHERS' INTRINSIC MOTIVATION? EVIDENCE FROM THE 2000 SASS DATABASE

1. Introduction	109
2. Accountability Framed by the Principal-Agent Model	111
3. Self-Determination Theory	113
4. Literature Review	117
5. Study Hypotheses	119
6. Data	120
7. Method	126
8. Results	127
8.1 Would Teachers Under Strong Accountability Policy Become a Teacher Again, If They Were to Start Over?	127
8.2 Do Teachers Under Strong Accountability Policy Become More Likely to Think that It is a Waste of Time to Try to Do Best as a Teacher?	131
8.3 Do Teachers Become More Dissatisfied When Working Under Strong Accountability Policy?	135

9. Conclusion	138
APPENDIX	
1. Questionnaires Used as Dependent Measures	141
2. Two groups by the intensity of accountability in 1999-2000	141
REFERENCES	143

LIST OF TABLES

Table 1.1: Example of One Input and One Output	6
Table 1.2: Change of District Efficiency of Michigan Public School Districts, 1995-2000	13
Table 1.3: Change in District Efficiency by Charter-Hosting Status	14
Table 1.4: Location of Charter-Hosting Districts by Efficiency Change	15
Table 1.5: Change of Efficiency by Districts with At Least 6% Charter Enrollment	16
Table 1.6: Difference-in-Difference Estimates of District Efficiency Change	17
Table 1.7: FD and Robust Regression. Effects of Change in Charter School Share of Enrollment on Change in District Efficiency	19
Table 1.8: Median FD Regression. Effects of Change in Charter School Share of Enrollment on Change in District Efficiency.....	20
Table 1.9: Robust and Median Regression. Difference-in-Difference Estimates.....	21
Table 2.1: Underlying Theories of Accountability and Professional Development	37
Table 2.2: Two Groups by the Intensity of Accountability in 1999-2000	51
Table 2.3: Definitions of Dependent and Independent Variables Used in the Analysis...	57
Table 2.4: Average Usefulness of Five Professional Development Programs Rated by Teachers	65
Table 2.5: Summary of Statistics for Variables	67
Table 2.6: Four Ordered Probit Models. The Effect on the Use of State Standards	70
Table 2.7: Change of Coefficient Size on Accountability in Ordered Probit Model	72
Table 2.8: Linear Regression, Ordered Probit, and Ordered Logit Estimates of Teachers' Use of State/District Standards for Instruction	74
Table 2.9: Linear Regression, Ordered Probit, and Ordered Logit Estimates of Teachers' Use of State/District Standards for Instruction After Dropping Three Variables, Age, HRSmath, and HRSread.....	77

Table 2.10: The Partial Effect of Expenditure of Instructional Staff Support on Teachers' Standards-related Professional Development Program Participation in Michigan (Probit Model)	79
Table 2.11: The Partial Effect of the Number of PD for Administrators on Teachers' Standards-related Professional Development Program Participation (Probit Model)	82
Table 2.12: IV Estimation for the Effect on Teachers' Use of State/District Standards ..	83
Table 2.13: Four Ordered Probit Modes of the Effect on Teachers' Use of Test to Strengthen Their Content Knowledge and Teaching Practice	86
Table 2.14: Effect on the Teachers' Use of Information from State or Local Test Scores to Strengthen Subject Area and Practice	87
Table 2.15: Four Ordered Probit Models of the Effect on Teachers' Use of Test to Adjust Their Curriculum in Areas Where Their Students Encountered Problems ...	89
Table 2.16: Effect on the Teachers' Use of Information from State or Local Test Scores to Adjust Their Curriculum in Areas Where Their Students Encountered Problems.....	91
Table 2.17: Three Ordered Probit Models of the Effect on Teachers' Use of Test to Group Students by Achievement or Ability	92
Table 2.18: Effect on the Teachers' Use of Information from State or Local Test Scores to Group Students into Different Instructional Groups by Achievement or Ability	94
Table 3.1: Definitions of Dependent and Independent Variables Used in the Analysis..	122
Table 3.2: Summary Statistics	125
Table 3.3: Four Ordered Probit Models on the Teachers' Perception that They Would Not Become a Teacher Again	128
Table 3.4: Effects on the Teachers' Perception on Whether They Would Not Become a Teacher Again	130
Table 3.5: Four Ordered Probit Models of Effects on the Teachers' Perception that Teaching Hard is Not a Waste of Time	132
Table 3.6: Effects on the Teachers' Perception that Teaching Hard is Not a Waste of Time	133

Table 3.7: Four Ordered Probit Models on the Effect of Accountability on Teachers' Dissatisfaction	136
Table 3.8: Effects on Teachers' Job Dissatisfaction	137

CHAPTER I

DO CHARTER SCHOOLS SPUR IMPROVED EFFICIENCY IN TRADITIONAL PUBLIC SCHOOLS?

AN EVALUATION OF MICHIGAN SCHOOLS USING DATA ENVELOPEMENT ANALYSIS

1. Introduction

Charter schools are the most rapidly expanding form of school choice. The basic charter school concept is encompassed in the idea of “autonomy in exchange for accountability.” Charter schools are nonsectarian public schools of choice that operate with freedom from some of the regulations that apply to traditional public schools (U.S. Charter School Office). However, the degree of freedom from state or district regulations for charter schools varies across states (Kane & Lauricella, 2001). For instance, the laws governing charter school laws in six states (Arizona, Colorado, Florida, Michigan, North Carolina, and Texas) are relatively permissive. These six states generally foster the development of charter schools that are genuinely independent of local school districts. These charters enter the educational market place as competitors for the students and revenues of traditional public schools. However, except for Arizona, all these states have a cap on the number of operating charter schools. Michigan is one of the states, which allows many charter schools.

Currently, there are four general perspectives on the charter school movement. The first is the laboratory perspective. This perspective posits that charter schools, freed

up from regulations, will experiment with new educational practices. If these new practices are useful or innovative, they can be adopted by regular public schools. Second is the competition perspective. Since charter schools will attract children and money from the district schools, regular public schools will face financial incentives to persuade families not to exit. Public schools will improve their instruction and programs to avoid losing their students. For example, according to the Center for Education Reform, Rocky Mount County Public Schools in North Carolina had long considered an International Baccalaureate program, but had never acted on it. However, when the Rocky Mount Charter School applied for a charter, the district was spurred into action. Third is the alternative system perspective. As the number of charter schools increase, charter schools may replace district schools as the primary purveyors of public education. This perspective is appealing to those who believe that regular public schools are unlikely to respond constructively to the presence of charter schools. Thus, in the long term, the charter school system may be an alternative to the whole public education system. Last is the useless movement perspective. This perspective maintains that the charter school movement will not bring any positive effect. Examples of successful charters are anecdotal, while public schools are already implementing the good practices that are adopted by charter schools (Rothstein, 1998).

In fact, these various perspectives are indicative of the intensity of the current arguments concerning the charter school movement. A charter school evaluation could be implemented based on any one of the above four perspectives. Among these perspectives, my research intends to examine the competitive effect of charter schools on school districts hosting charter schools, using Michigan K-12 data. First, this chapter will

introduce current research on the competitive effects of charter schools on public schools. And, evaluation methodology will be discussed and analysis using Michigan Data will be presented. Specifically, the results will be compared with previous research implemented by Hoxby (2003) and policy implication will be discussed.

2. Michigan Charter School Studies

So far, there has been little research on the competitive effect of charter schools¹. Based on interviews with over 270 principals in Michigan, Mintrom (2000) concludes that there is little evidence that public schools near charter schools have been systematically changing their practices because of competition from charter schools. Bettinger (1999) finds that Michigan charter schools did not improve student achievement as rapidly as other regular public schools using school-level data in difference-in-difference regression models. In addition, Bettinger failed to find evidence that the presence of charter schools produces improvements in student achievement in nearby regular public schools. However, using similar difference-in-difference regression techniques, Hoxby (2003) finds productivity increases in Michigan school districts where charter enrollment represents at least six percent of total local enrollment. These studies, then, produce different conclusions regarding the competition effect. The different results may be due to differences in study design. For instance, Hoxby uses host districts where the enrollment share of charter schools is more than six percent as a dummy variable to

¹ Belfield and Levin (2002) review the cross-sectional research evidence on the effects of competition on educational outcomes. However, their review does not include charter school cases. Usually competition

measure the competitive effect between 1993 and 2000. Bettinger, on the other hand, uses the Herfindahl index of school enrollment for schools within a five-mile radius of charter schools using 1995 and 1999 school data. In addition, one of the important differences between the Hoxby and Bettinger studies is that they used different dependent variables. Hoxby uses the change in test score over the percentage change in per pupil expenditure, while Bettinger uses test scores as the dependent variable. That is, Hoxby's model measures the impact of charter schools on the efficiency of regular public schools. Bettinger's model, meanwhile measures the effect of charter schools on the effectiveness of regular public schools.

This paper is designed to investigate the Michigan story more fully, in order to provide a stronger empirical foundation for the policy debate. Like Hoxby and Bettinger, I will rely on Michigan Educational Assessment Program (MEAP) data and measures of district resource use. However, I will use an entirely different empirical methodology, Data Envelopment Analysis (DEA) to measure and evaluate school efficiency. Since my method permits the examination of changes in school efficiency associated with the presence of nearby charter schools, the result can be compared with Hoxby's.

3. Methodology

Evaluation criteria can be divided into two broad categories: those that measure good practices, and those that measure good student performance. The former is input-based evaluation, and the latter is output-based evaluation. My research basically adopts

is measured using the enrollment rate at private schools or by the potential for Tiebout-style competition

an output-based quantitative evaluation, since the educational practices in charter schools or nearby public schools are not explicitly investigated. This study looks at the efficiency with which schools use their resources (inputs) to generate educational outputs measured by student achievement. I employ Data Envelopment Analysis, specifically a super-efficiency DEA model, to evaluate whether school districts that host charter schools have improved efficiency.

DEA is a methodology which has been used to evaluate organizations' relative efficiency using input and output information (Charnes, et al., 1978; Bessent & Bessent, 1980; Anderson, et al., 1998). The basic conceptual model for DEA is designed to derive an organization's efficiency from its output/input ratio. That is, DEA derives an efficiency index for each organization as the weighted sum of its outputs over the weighted sum of its inputs. It sets the most efficient organization's index as 1 (or 100%), and scales the index of all other organizations relative to the most efficient organizations. The one input and one output case is easy to understand. For instance, assume that we have three schools, each using one input (expenditure) and one output (math scale score). Then, we can easily obtain the efficiency index for these three schools by dividing math score by expenditure. Table 1.1 displays the efficiency indexes for this hypothetical one input and one output case. As we can see, School 1 and School 2 achieved the same math scaled score, but with different amounts of inputs. School 1 achieved a 2000 scale score using only \$500, so it is more efficient than School 2. Since School 1 is the most efficient, its efficiency index is set as 1, and School 2's index is 0.65.

among regular school districts.

Table 1.1: Example of One Input and One Output

	School 1	School 2	School 3
Output(Math scaled score)	2000	2000	1400
Input (expenditure)	\$ 500	\$ 800	\$ 700
Efficiency	4	2.5	2
Weighted Efficiency Index (Most efficient unit is 1)	1	0.625	0.5

The mathematical model for DEA in situations involving more than one input and output was proposed by Charnes, et. al.(1978). The model defines the efficiency of organization, or decision-making unit (DMU), o as h_o , which is obtained by solving the following linear programming model².

$$\max h_o = \frac{\sum_{r=1}^s u_r y_{ro}}{\sum_{i=1}^m v_i x_{io}} \quad (1)$$

subject to:

$$\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1$$

where, $j= 1,, n$, and $u_r, v_i \geq \varepsilon$; $r=1,,s$; $i= 1,,m$. Here r and i are the number of outputs and inputs. y_{rj} and x_{ij} are known outputs and inputs of j th DMU and the u_r and $v_i \geq \varepsilon$ are the variable weights to be determined by the solution of this problem. The weight

for r is u_r , and v_i is the weight given to output i . ε is a non-Archimedean infinitesimal.

This is a non-linear model. We can convert it into a linear programming model as follows:

$$\text{Max } \sum_{k=1}^s \mu_j y_{jo} \quad (2)$$

Subject to:

$$\sum_{i=1}^m w_i x_{ij} - \sum_{r=1}^s \mu_r y_{rj} \geq 0,$$

$$\sum_{i=1}^m w_i x_{io} = 1$$

Where, $w_i = t v_i$, and $\mu_r = t u_r$, and $t^{-1} = \sum_{i=1}^m v_i x_{io}$, $t > 0$. Conversely, model

(2) can be transformed into the model (1). Although the weighting variables in (2) have been transformed, we can use (2) to solve (1), since the models are equivalent³.

This basic DEA model⁴ has been used to analyze school or school district efficiency by some scholars (Bessent & Bessent, 1980; Bessent, et al., 1982; Anderson, et al., 1998). Bessent & Bessent (1980) might be the first scholars who applied DEA to analyze the efficiency of the public schools. They examined efficiency ratings for 55 elementary schools in an urban school district. They selected input and output variables according to the following criteria: there was a conceptual basis for the relationship

² For basic information about linear programming model that is used in operations research (OR), please see Hiller, F.S. and Lieberman, G. J. (2001), Introduction to Operations Research, 7th edition. McGraw Hill.

³ Charnes, et al. (1978) illustrates how linear programming model minimizing input can solve (1). For how (2) can be drawn to solve (1) see Charnes, et al. (1978, p.432).

between input and outputs, there was an empirically inferred relationship of measured inputs and outputs, the relationship was such that increases in inputs were associated with increases in outputs (for instance, the percent of students not from low-income families was used rather than the percent of students from low-income families.), the measurement had no zero elements. If a given element was zero, a small value (0.01) was added. Based on these criteria, Bessent & Bessent, (1980) used two outputs and thirteen inputs to estimate efficiency indexes for each of 55 public schools⁵.

Recent application of DEA in education was done by Anderson, et al., (1998). They applied DEA to measure the efficiency of Chicago public schools using 1989, 1991, and 1993 data. The input variables used in the analysis were: student attendance, stability (the percentage of students remaining enrolled for the entire school year), percentage of students not classified as poverty level, percentage of students speaking English as their first language, teacher/student ratio, and total per-pupil direct educational expenditures. Output variables are grade equivalent scores in reading, mathematics, and vocabulary. Anderson, et al.,(1998) also obtain effectiveness ratings based on residual gain scores between seventh and eighth grades. They compare the effectiveness ratings and efficiency index and found that schools that were both effective and efficient in various years had more stable student populations, greater attendance, fewer students with Limited English Proficiency, more non-poverty students, and lower student expenditures. However, these results are not completely consistent across all three years. They

⁴ This basic DEA was proposed by Charnes, Cooper, and Rhodes (1978), so sometimes it is called the CCR model, or CCR ratio definition (Banker, Charnes, and Cooper, 1984).

⁵ The two outputs were median percentile test scores in reading and math for each school. The thirteen inputs were: previous year's median percentile reading and math test scores, percent of Anglo-American students, percent of students not from low-income families, percent in average daily attendance, total per pupil expenditure for instruction, number of professional staff per 100 pupils and so on

suggested that the reason might be multicollinearity among the independent variables in their regression. DEA methods also have been utilized to evaluate the efficiency of higher education departments (Johnes and Johnes, 1995).

Thus far, however, this established technique has not been utilized to evaluate the competitive effects of charter schools. Despite the intense policy debate regarding charter schools' competitive impact on traditional schools, researchers have yet to employ the best empirical methods to produce evidence that could inform the debate. DEA is well-suited to this task. The fact that the model is non-parametric is an advantage. We can check whether individual school districts with nearby charter schools improve their efficiency. In addition, the DEA model's ability to deal with multiple outcomes simultaneously is an advantage.

One potential drawback of past applications of the basic DEA model, however, is that it may assign an efficiency index of 100 to many organizations. For instance, among the 55 schools analyzed by Bessent & Bessent (1980), 30 received efficiency indexes of 100. A quarter of the schools evaluated with the basic DEA model in Anderson et al.(1998) were assigned efficiency indexes of 100. To adjust for this, Andersen and Petersen (1993) propose a super-efficiency DEA model, which provides a comparative ranking among the Decision Making Units (DMUs) with efficiency indexes of 100. All other assumptions of the super-efficiency DEA model remain the same as the basic DEA model, except for the fact that the DMU under evaluation is excluded from the constraints. Since the super-efficiency model provides relative efficiency rankings of all units, we can see whether charter-hosting school districts increased efficiency index compared to the non-charter hosting public school districts. These analyses will be tested

by using discrete categories indicating whether the increase of efficiency index is associated with charter-hosting status.

Also, using the efficiency index as dependent variable, we can run regressions to obtain difference-in-difference estimator to measure the effect of charter school enrollment on charter hosting districts. The regression model is

$$\Delta \text{super-efficiency index}_i = \delta_0 + \beta_1 \text{6\% CS Share}_i + \Delta u_i, \quad (3)$$

where, 6% CS share is a dummy variable which means that if charter school enrollment in the district i is more than 6% of the total enrollment of the district.

Unobserved independent variables that possibly could affect the efficiency are deleted in this model by differencing. Also, log-transformation of the dependent variable, change in log (super-efficiency index) will be used. This model, (3) assumes that 6%CS share is not correlated with the Δu_i and other general assumptions of Ordinary Least Square (OLS) regression are met.

Actually, (3) is identical with a First Differenced (FD) equation. Instead of using a dummy independent variable, 6%CS share cutoff dummy variable, we can include a continuous independent variable, change of percentage of charter school share in the district enrollment, which could provide more information compared to the discrete dummy independent variable. That is,

$$\Delta \text{super-efficiency index}_i = \delta_0 + \beta_1 \Delta \text{share of charter school}_i + \Delta u_i, \quad (4)$$

where Δ denotes the change from $t = 1$ to $t = 2$. This model also eliminates the unobserved time-invariant variables which could have effects on the efficiency index. The intercept indicates the overall change of efficiency index for all districts. Model (4) also must satisfy assumptions of OLS regression. For instance, β_1 of (3) and (4) could be biased if the charter school location is determined by some other factors such as school district efficiency and so on. In Michigan, many charter schools are located in urban counties and Detroit area (Arsen, et al., 1999). In this first-differenced equation, the correlation between independent variable and unobservable time-constant variable is allowed, thus the locality of charter schools location which is time-invariant does not matter. However, if the change of efficiency is correlated with some factors which are time-variant, β_1 would be biased. This issue is relevant to the possibility that the changes in the composition of students who remain in district schools could bring peer effect and could affect district efficiency. This would constitute a misspecification, or omitted variable problem. So, observable student characteristics will be used as input variables to yield an efficiency index which controls for composition of students.

4. Data

Since 1994, the number of charter schools in Michigan has grown (Arsen, et al., 1999). The state's charter schools are distributed unevenly among school districts. As of 2002, about 82, or 15 percent, of the state's 555 public school districts are hosting charter schools. This situation provides a good quasi-experiment condition for an evaluation of charter schools' effect on school districts over years. If only a few school districts hosted

charter schools or all school districts hosted charter schools, it may be difficult to isolate their competitive effect.

The input variables for the DEA analysis are: average teacher salary, current operating expenditures per pupil, teacher/student ratio, percent of students who are white, percent of students who are non-special education, and percent of students who do not receive free/reduced lunch⁶. The output variables are fourth and seventh grade math and reading scores (percent of students who received satisfactory scores), and graduation rates. These data were obtained from the Michigan Department of Education's K-12 database⁷. The data were gathered for two school years, 1994-1995 and 1999-2000, in order to measure the "change" in efficiency over time⁸. I assessed the efficiency of a DMU over time by treating it as a different unit in one time analysis⁹. Expenditure and teacher salary data were converted to real dollars using the CPI. Districts hosting charter schools between 1995 and 1999 are regarded as charter hosting districts. Districts which hosted charter schools for the first time in 1999-2000 were not regarded as charter hosting districts, since only one year would not be enough time to improve efficiency. Since 7th grade test scores and graduation rates were among the outputs, 31 elementary school districts were eliminated from the sample. This left 523 K-12 districts in Michigan that existed in both 1995 and 2000. Seventy-three of these school districts hosted charter schools as of 2000, and more than half of these charter schools were located in the metropolitan Detroit area (Please see Appendix).

⁶ Since the DEA input variables must be positively utilized inputs, the included variables, for instance, are students who are *not* receiving free or reduced lunch, and so on.

⁷ The website is "<http://www.state.mi.us/mde/cfdata/k12db/welcome.cfm>"

⁸ For the 1999-2000 school year, the 1999 spring graduation rate was used, since the 2000 graduation rate was not available)

5. Results of Analysis

5.1 Basic Results and Chi-Square Tests

Changes in school district performance will be reflected in changes in their efficiency indexes. If a school district's index increased, the district can be regarded as improving its efficiency compared to other districts. On the contrary, if a school district's index was decreased, then it did not show any improvement in its organization efficiency compared to other districts. Following is the result from the super-efficiency DEA analysis.

Table 1.2: Change of District Efficiency of Michigan Public School Districts, 1995 to 2000

	Number of Districts	% of Districts
Increase in Efficiency Index	334	63.9%
Decrease in Efficiency Index	188	35.9%
No Change in Efficiency Index	1	0.2%
Total	523	100%

Table 1.2 displays the directional changes in district efficiency for all Michigan school districts regardless of their charter-hosting status. Efficiency increased in 63.9 percent of public school districts, and decreased in 35.9 percent. That is, the majority of Michigan public school districts have improved their efficiency between 1995 and 2000¹⁰.

⁹ Using longitudinal data and treating all the data over year as different units in the data set is called "window analysis"

¹⁰ This finding appears inconsistent with the argument that American public schools suffered from deteriorating efficiency in recent years, a view that may need to be examined further. We need to control for the change of student characteristics and so on. For example, Hanushek (1994) looks at only one input

Now, we turn our attention to the competitive effect of charter schools. The competitive effect hypothesis suggests that the presence of charter schools will push the districts in which they are located to increase their efficiency. Although five years from 1995 to 2000 may not be a sufficiently long for the competition effect to fully work out, we would not expect the presence of charter schools to trigger deterioration in efficiency according to the competition perspective. Table 1.3 shows that about 64.4 percent of the 73 charter-hosting public school districts increased their efficiency index.

Table 1.3: Change in District Efficiency by Charter-Hosting Status

	Non-charter Hosting		Charter Hosting	
	Number of Districts	% of Districts	Number of Districts	% of Districts
Increase in Efficiency Index	287	63.8%	47	64.4%
Decrease in Efficiency Index	162	36.0%	26	35.6%
No Change in Efficiency Index	1	0.2%	0	0%
Total	450	100%	73	100%

School districts hosting charter schools are slightly more likely to increase their efficiency, than districts without charter schools (64.4 percent versus 63.8 percent). The chi-square test was used to assess whether this difference is statistically significant (I disregarded the 'no- change in efficiency index' row). The chi-square test shows that this is not statistically significant, $\chi^2=0.006$, P-value > 0.99. Thus, we conclude that there is no significant association between the change of district efficiency and the presence of

and one output, student achievement test score and per-pupil expenditure, to measure efficiency change over recent decades. However, this is not an appropriate efficiency measure. We would be better off to use

charter schools.

One interesting investigation would be to check whether the efficiency change in charter hosting districts is related to their location, such as urbanicity. In the U.S. Census locale code for Michigan, schools districts has eight categories: Detroit, mid-size city in urban county, suburban area of Detroit metro area, suburbs of other urban counties, large town in rural county, small town in rural county, rural area, rural area in urban county. Among these, Detroit and mid-size cities such as Flint and Grand Rapids are urban centers in Michigan. The suburban area of Detroit metro area is likely to be a low income area (Lee and Reimann, 2003). So, the efficiency change of the charter-hosting districts was sorted by location as Table 1.4.

Table 1.4: Location of Charter-Hosting Districts by Efficiency Change

	Detroit/Midsize city in urban, Suburban in Detroit	Other Areas	Total
Increase	26	21	47
Decrease	9	17	26
Total	35	38	73

Table 1.4 does not provide any evidence that charter hosting districts located in urban or low income suburban areas did not improve efficiency. Rather, it appears that more charter-hosting districts in urban or low income locations show more increased efficiency than other areas. Since DEA analysis uses some inputs related to socio-economic status (for instance, percentage of student without free/reduced lunch program), this result is not counterintuitive. To test whether efficiency change is associated with

more inputs such as the characteristics of student population and other outputs simultaneously to measure

the location of charter-hosting districts, the chi-square test was used. The test result, $\chi^2=2.87$, P-value = 0.218, shows that no association exists.

As we can see in the above tables, competition effect does not occur evenly. Some charter- hosting school districts increased efficiency, while others did not. According to the competition hypothesis regarding the charter schools, charter hosting districts ought to respond to charter school competition by increasing or at least maintaining their efficiency. However, many charter-hosting districts did not. Competition does not appear to be a panacea.

The DEA analysis that has been illustrated so far provides evidence against Hoxby's analysis. However, Hoxby (2003) shows that on average, school districts with more than 6 percent of local enrollment in charter schools have improved their efficiency over time. So, I checked the change of efficiency index for these districts. Hoxby (2003, Table XII) provides the list of school districts where at least six percent of local students enrolled in charter schools. They are total 39 school districts. Among them, 27 school districts increased their efficiency, while twelve school districts did not increase their efficiency.

Table 1.5: Change of Efficiency by Districts with At Least 6% Charter Enrollment

	Districts with At least 6% Charter Enrollment		Others	
	Number of District	% of District	Number of Districts	% of Districts
Increase in Efficiency Index	27	69.2%	307	63.4%
Decrease in Efficiency Index	12	30.8%	176	36.4%
No Change in Efficiency Index	0	0%	1	0.2%
Total	39	100%	484	100%

the efficiency change over long periods. This can be done with the DEA method.

Table 1.5 shows that districts with at least 6 percent of local enrollment in charter schools were a slightly more likely to increase their efficiency index compared to others. However, the difference is minor and we fail to find any evidence supporting for Hoxby's results ($\chi^2=0.502$, P-value = 0.84).

5.2 Difference-in-Difference Estimate and Robust Regression Analysis

So far, chi-square test was used to see the association between charter-hosting status and efficiency change. This chi-square test loses some information since it is based on discrete categorical analysis. And to enable more direct comparison with the Hoxby's analysis, I run the regression equations, (3) and (4). Following table reports the results of two regression equations, using change of efficiency and change of natural log of efficiency index as dependent variables.

Table 1.6: Difference-in-Difference Estimates of District Efficiency Change

Independent Variable	Dependent Variable	
	Δ super-efficiency index	Δ log of super-efficiency index
	Estimates	Estimates
6%CS share	0.110* (0.026)	0.064** (0.02)
Constant	0.023** (0.007)	0.028** (0.005)
N	523	523
R ²	0.03	0.019

Note: * means the coefficient is significant at the 0.05 level. ** means that the coefficient is significant at the 0.01 level. The quantities in parentheses below the estimates are the robust standard errors. 6% Ch share is an indicator variable for the districts which have at least 6 percent of enrollment in charter schools.

Table 1.6 displays that overall, all non-charter hosting school districts improved efficiency index by 0.027 significantly on average. The dummy independent variable, 6%CS share, appears to be significant when the dependent variable is transformed using natural log. This result is consistent with Hoxby's result in a sense. However, since the use of a dummy variable to represent the density of charter school competition loses some information, we need to check the equation (4) to see whether a continuous measure of charter school enrollment is associated with efficiency change in public school districts. A scatter plot was examined to determine whether there is non linear relationship between the change of efficiency and the change of charter school enrollment share and also to determine if outliers exist (See Appendix 2).

The scatter plot shows that there are two outliers even using a log-transformation of the dependent variable. Thus, we also need to run a robust regression, which deals with outliers using Cook's distance, or median regression to examine whether the robust regression or least absolute deviations (LAD) estimator from median regression will generate different results. The robust regression calculates Cook's Distance and eliminates any observation for which Cook's Distance is bigger than 1. After eliminating outliers, the robust regression weights each case¹¹ and obtain an estimate which is robust to outliers. Median regression produces median expected value, so it is also robust to the outliers.

Table 1.7 displays the results of estimating model and using two First Differenced (FD) regression equations which do not address the outlier concern and two robust FD regression equations which do.

¹¹ For more detail about robust regression, please see STATA technical manual, Reference [R].

Table 1.7: FD and Robust FD regression. Effects of Change in Charter School Share of Enrollment on Change in District Efficiency.

	<u>FD regression</u>		<u>Robust FD regression</u>	
	Dependent variable: Δ super-efficiency index	Dependent variable: Δ log of super-efficiency index	Dependent variable: Δ super-efficiency index	Dependent variable: Δ log of super-efficiency index
	Estimate	Estimate	Estimate	Estimate
Δ CSshare	1.045** (0.137)	0.589** (0.105)	0.108 (0.081)	0.121 (0.088)
Constant	0.018* (0.007)	0.024** (0.005)	0.023** (0.003)	0.025** (0.004)
N	523	523	521	521
R ²	0.099	0.057		

Note: * means the coefficient is significant at the 0.05 level. ** means that the coefficient is significant at the 0.01 level. The quantities in parentheses below the estimates are the robust standard errors. Δ Ch share is continuous explanatory variable, change of enrollment in charter schools for the school district.

Table 1.7 exhibits that robust FD regression significantly reduces the size of the independent variable, change in charter school share of enrollment, than usual FD regression. Standard error of the estimates decreases a little bit in the robust equation. Two observations are eliminated from the robust regression and the effect of explanatory variable becomes statistically insignificant.

Table 1.8 displays the results of equation (4) using median regression. The coefficient size of median regression and robust regression appear to be similar. The median regression shows that change in charter school share of enrollment does not appear to be significant.

Table 1.8: Median FD regression. Effects of Change in Charter School Share of Enrollment on Change in District Efficiency.

	Median Regression	
	Dependent variable: Δ super-efficiency index	Dependent variable: Δ log of super-efficiency index
	LAD Estimate	LAD Estimate
Δ CSshare	0.106 (0.079)	0.127 (0.088)
Constant	0.022** (0.004)	0.026** (0.005)
N	523	523
Pseudo-R ²	0.0028	0.003

Thus, we conclude that the charter school enrollment share does not increase districts' efficiency significantly. In contrast to Hoxby's findings, these results suggest that merely increasing the number of charter school will not automatically generate an increase of efficiency of charter-hosting district.

Table 1.9 reports the robust regression and median regression for the difference-in-difference equation and shows that indicator variable, 6% charter school enrollment share, does not affect efficiency change significantly. When we log-transformed the dependent variable, one observation is eliminated from the robust regression model. Median increase in log of efficiency for non-charter hosting school districts is 0.026, which is statistically significant. Both coefficient and its standard error are decreased in this robust regression equation and median regression. The coefficient of independent variable, 6%CS share, in median and robust regression becomes insignificant.

Thus, we conclude that the indicator variable, 6% charter school share, does not have any effect on the change of school district efficiency.

Table 1.9: Robust and Median Regression. Difference-in-Difference Estimates

Independent Variable	<u>Robust Regression</u>		<u>Median regression</u>	
	Dependent variable: Δ super-efficiency index	Dependent variable: Δ log of super-efficiency index	Dependent variable: Δ super-efficiency index	Dependent variable: Δ log of super-efficiency index
	Estimates	Estimates	LAD Estimates	LAD Estimates
6%CS share	0.012 (0.013)	0.014 (0.014)	0.0085 (0.014)	0.0093 (0.016)
Constant	0.023** (0.0036)	0.025** (0.0039)	0.023** (0.004)	0.026** (0.004)
N	522	523	523	523

Note: * means the coefficient is significant at the 0.05 level. ** means that the coefficient is significant at the 0.01 level. The quantities in parentheses below the estimates are the robust standard errors. 6% Ch share is an indicator variable for the districts which have at least 6 percent of enrollment in charter schools.

6. Discussion and Concluding Remarks

The analysis of charter schools' competitive effect on school districts will enhance our understanding of a key dimension of how charter schools are (or are not) changing the educational system. We should see whether the charter schools spur improved organizational efficiency in hosting public schools and identify patterns in efficiency changes across districts. DEA can be utilized in various ways to see the competitive effect of charter schools on the efficiency of regular public schools.

After we obtain efficiency index using DEA, we used categorical table to see the overall effect of charter schools. A Chi-square test shows that charter-hosting status is not significantly associated with district efficiency change. Despite some limitation of the chi-square test due to its reliance on discrete categorical analysis, the test indicates that

charter-hosting districts do not respond to the competition from charter schools in a monolithic way. Some districts did not improve efficiency, while others improved. Although more charter-hosting districts appear to increase their relative efficiency compared to other non-charter-hosting districts, this is not a significant difference. We will need to wait and see whether the number of charter hosting districts that improved relative efficiency will be maintained or become significant.

To more direct comparison with Hoxby's analysis, the difference-in-difference estimation and FD regression analysis were implemented. The regression analysis illustrates that change of the enrollment share of charter schools does not lead to the improvement of efficiency of the charter hosting districts significantly. So, this implies that merely increasing the number of charter schools will not necessarily produce significant improvement in district efficiency. However, the regression analysis has limitations since it does not control for ceiling effect on the test scores and possible unobservable time variant factors are correlated with change in charter school share. This limitation also applied to the Hoxby's (2003) study.

In contrast to the simple market theory that presumes charter schools will yield a more competitive environment in the public school system will not work. Districts may respond to the competition in various ways. For instance, regular public schools may want to cooperate with charter schools if charter schools take out disadvantaged students from them. In North Carolina, charter schools have taken low performing students out of the nearby public schools, so nearby regular public schools become very favorable to charter schools and do not feel any competition. According to Kathryn Meyer, Chairperson of Durham School Board, this situation is occurring to some extent in

Durham, with the charter schools attracting low-performing students¹². In this case regular public schools will be comforted by the existence of charter schools because of this kind of “take-out-of” effect. If all schools in Durham turn into charter schools, a similar equilibrium could occur. Currently, charter schools serve minority or white students disproportionately in North Carolina. This is not inconsistent with the theory of market system. Separated education markets (that is, charter schools) for Black, Asian, White, or Hispanic students could emerge.

Similarly, in Michigan, not only competitive response to the charter schools exists, but also cooperative or collusive responses exist (Arsen, et al. 2002). For instance, some intermediate school districts that saw the trend of responding to charter schools are moving in the cooperative direction as the administrators are adapting to the new policy environment from the competition. According to another mid-Michigan ISD superintendent, the knee-jerk reaction was competitive, but when districts realized there would not be a mass exodus they actually started to cooperate more than they did prior to choice (Arsen, et al. 2002).

Finally, it must be noted that this quantitative study cannot fully illuminate how the inside of public schools is changing. For instance, one can question whether observed changes in efficiency relate instead to the changes in instructional practice or other institutional characteristics that were not captured in the statistical analysis. An efficiency improvement could come from teachers teaching to achievement tests, resulting from pressure to improve achievement scores on standardized tests. In addition,

¹² This story is from my master’s thesis. Lee, K (2000) “What lessons learned by charter schools should the resource center be disseminating to other regular public schools?” Sanford Institute of Public Policy, Duke University.

it is reasonable to ask why some charter schools did not improve their efficiency. It is hard to answer to this question from the data analysis. Thus, quantitative analysis is limited in illuminating real changes inside public schools. We may need to investigate how schools change their practice using qualitative evaluation methods.

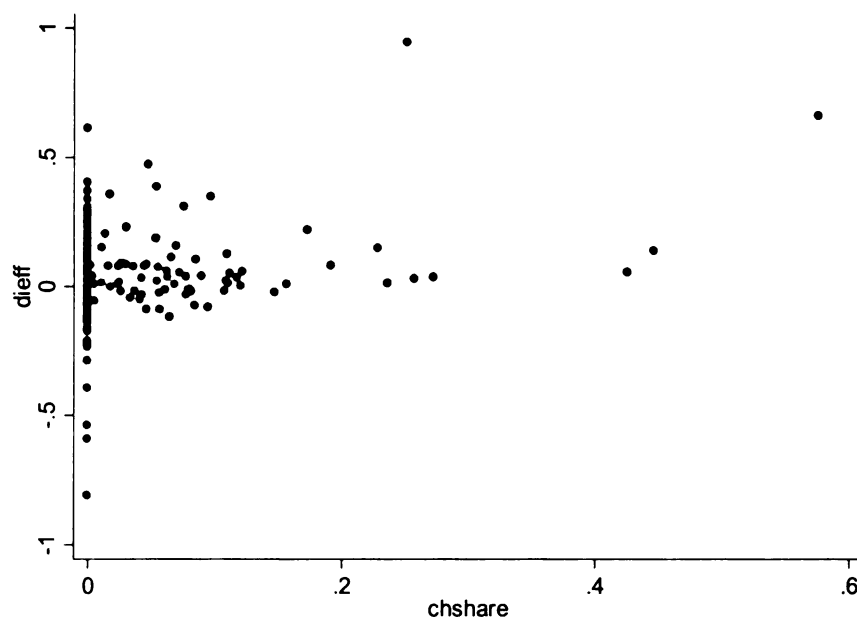
APPENDIX

1. Table: Number of Charter School by Locale Code in Michigan (as of 2000)

Locale Code	Number of Charter Schools	% of Total Sum
Detroit	45	27.4
Mid-size City, urban county	48	29.3
Suburban area of Detroit metro area	27	16.5
Suburbs of other urban counties	16	9.8
Large Town in rural county	2	1.2
Small Town in rural county	11	6.7
Rural area	5	3.0
Rural area in urban county	10	6.1
Total	164	100

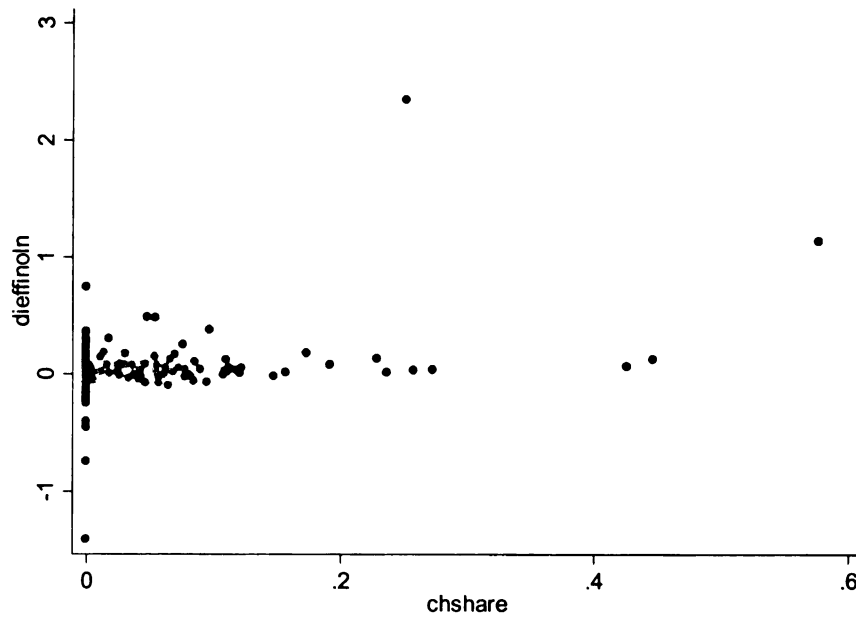
(Source: Michigan Department of Education K-12 Database)

2. Figure: Scatter Plot. Change in Log of Efficiency by Change of Charter School Enrollment Share in District



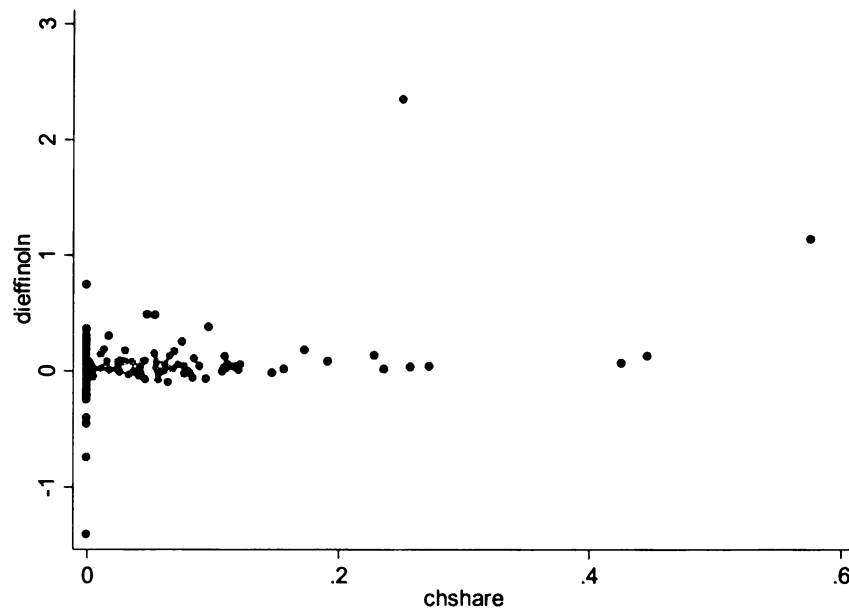
(Note: dieff = change in log of efficiency. Chshare= charter school enrollment)

3. Figure: Scatter Plot. Change in Efficiency by Change of Charter School Enrollment Share in District.



(Note: dieffinln = change in efficiency. Chshare= charter school enrollment.)

3. Figure: Scatter Plot. Change in Efficiency by Change of Charter School Enrollment Share in District.



(Note: dieffinln = change in efficiency. Chshare= charter school enrollment.)

REFERENCES

- Andersen, P. and Petersen, N.C. (1993). A procedure for ranking efficient units in Data Envelopment Analysis, *Management Science*, 39, p.1261-1264.
- Anderson, L., Walberg, H.J. & Weinstein, T. (1998). Efficiency and effectiveness analysis of Chicago public elementary schools: 1989, 1991, 1993. *Educational Administration Quarterly* Vol. 34, No. 4 (Oct) 484-504.
- Arsen, D., Plank, D, and Sykes, G. (1999), *School Choice Policies in Michigan: The Rules Matter*. The Education Policy Center at Michigan State University, 1999. Available at www.epc.msu.edu.
- Arsen, D., Plank, D., and Sykes, G. (2002), School Choice Policies: How Have They Affected Michigan's Education System. The Education Policy Center at Michigan State University, 2002, Working Paper #10. Available at www.epc.msu.edu
- Benfield, C.R. & Levin, H. M. (2002), "The Effects of Competition on Educational Outcomes: A Review of US Evidence", Occasional Paper #35, National Center for the Study of Privatization in Education, Columbia University. Available at www.ncspe.org
- Bessent, A., and Bessent, E.W. (1980). Determining the comparative efficiency of schools through data envelopment analysis. *Educational Administration Quarterly* 16:57-75
- Bessent, A., Bessent, E.W., and Reagan, K.B.(1982). An application of mathematical programming to assess productivity in the Houston Independent School District. *Management Science*, Vol. 28, Issue 12 (Dec.) 1355-1367.
- Bettinger, E. (1999). The effect of charter schools on charter students and public schools, Occasional Paper No.4. National Center for the Study of Privatization in Education. Teachers College, Columbia University.
- Center for Educational Reform. <http://edreform.com>
- Charnes, A., Cooper, W., and Rhodes, E. (1978), Measuring the efficiency of decision making units. *European Journal of Operation Research* 2, 6 (Nov.): 429-444.
- Charnes, A., Cooper, W., and Rhodes, E. (1979), Short communication: measuring the efficiency of decision making units, *European Journal of Operation Research*, p.339
- Chubb, J. and Moe, T., (1990), *Politics, Markets, and America's Schools*. Brookings.
- Hanushek, Eric (2002), Publicly Provided Education, The Handbook of Public Economics, Amsterdam: North-Holland

Hassel, B.C. (1999) "Charter Schools: Politics and Practice in Four States." In Peterson, P. E. & Hassel, B.C. (Coed.), *Learning from School Choice*, Brookings.

Hoxby, C.M. (2003), School Choice and School Productivity. In Hoxby, C.M. (ed.), *The Economics of School Choice*, University of Chicago Press

Johnes, J & Johnes, G. (1995), Research funding and performance in U.K. university departments of economics: A frontier analysis. *Economics of Education Review*, Vol. 14, No. 3, pp. 265-284.

Kane, P.R. & Lauricella, C.J. (2001) "Assessing the Growth and Potential of Charter Schools" In Levin, M.H. (ed.), *Privatizing Education*, Westview.

Lee, K. (2000). What are the lessons learned by charter schools. Master's Memo. Terry Sanford Institute of Public Policy. Duke University.

Lee, K and Reimann, C. *Who's Attending Michigan's Priority Schools?* The Education Policy Center at Michigan State University, Data Brief #12, July 2003. Available at www.epc.msu.edu.

Mintrom, M. (2000), Leveraging Local Innovation: the case of Michigan's charter schools. Working Paper 6, The Education Policy Center. Michigan State University

Rothstein, R. (1998), Charter Conundrum. *The American Prospective*.
<http://www.prospect.org/print/V9/39/rothstein-r.html>

The Michigan Department of Education. <http://www.michigan.gov/mde>

The US Charter School Office
www.uscharterschools.org

Wooldridge, J. (1999). *Introductory Econometrics: A Modern Approach*. South-Western.

Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

CHAPTER II

ACCOUNTABILITY POLICY OR OPPORTUNITY TO LEARN? THE IMPACT SIZE OF ACCOUNTABILITY AND PROFESSIONAL DEVELOPMENT ON INSTRUCTIONAL PRACTICE -EVIDENCE FROM 2000 NCES SASS DATABASE

1. Introduction

Even though the body of research on the effect of accountability policy is growing, there are relatively few systematic studies to test the underlying theories of how this policy works. The present study will examine two competing theories regarding the impact of standards-based accountability policy¹³ on teachers' instructional practice.

The first theory is the principal-agent model derived from the behavioral predictions or "theory of action" put forth by accountability policy advocates. According to this model, a properly designed state-level accountability system will induce teachers (agents in the principal-agent model) to use state curriculum guidance and students' test results to modify their instructional practice in order to improve student performance. A properly designed accountability system involves a well-designed assessment system and a sufficient incentive mechanism attached to the outcomes. For instance, a state

¹³ Actually, standards-based reform and accountability policy are not synonym. Standards-based reform put more focus on standard and aligned assessment. Accountability policy shares the same components with standards-based reform, however additionally accountability policy includes and emphasizes an incentive mechanism, monetary rewards or sanction to schools/teachers for the outcome, student achievement test. It is hard to disentangle accountability policy from standards-based reform or vice versa. For instance, the book, *Holding Schools Accountable* (Ladd, ed, 1996) discusses accountability policy. Cohen discusses the effect of standard based reform in California in the book, while others discuss accountability system. I use accountability policy as broad concept which covers standards-based reform and accountability, since accountability policy covers standards-based reform idea although it entails incentive mechanisms.

assessment system such as value-added assessment has been proposed since it can provide relatively good information on student achievement growth.

Enough monetary reward for the attainment of test score goals or sanction for the failure to meet them are major components of the accountability policy since the theory says that an incentive system will induce the teachers and schools to make an effort to increase students' test scores. Provided with value-added test results and incentives attached to them, teachers (agents) working for the state (principal) will focus on student achievement and use test results and the state curriculum to adjust their instructional practice. Thus it is reasonable that some scholars think that accountability is based on the business incentive model emphasizing observable outcomes because incentives are the key mechanism of the accountability policy.

The second theory is derived from institutional research on school operations and predicts that idiosyncratic features of teachers and schools will determine their responses to a standards-based accountability system. For instance, David Cohen (1996) suggests that schools and teachers impose fundamental changes on the standards-based reform. Their responses vary depending on the local school context and teacher's individual characteristics, undermining the theory of action assumed by standards-based accountability proponents. According to this theory, learning policy for teachers and school administrators to enhance their capacity such as content and pedagogical knowledge would be critical for the successful implementation of accountability policy. Thus, professional development (PD) will have more influence on teaching practice than external accountability policy (Cohen & Hill, 2001).

According to this second perspective, accountability is not a sufficient education

policy. Test results and state curriculum guidance can speak to some aspects of the pedagogy. However, the pedagogical aspects of accountability policy are not self-implementing.

In addition, internal accountability or organizational capacity may determine the successful implementation of external accountability. However, an internal accountability system can be developed without a bureaucratic state-level accountability system (Newmann, et al., 1997). By this theory, given the influence of local context on the success of policy implementation, education policy must be designed in ways to enhance the capacity of enactors or organizations, and an accountability system does not meet this condition necessarily. Different scholars have slightly different interpretations of this second theory. Some scholars put more attention on the enactor's knowledge or belief (Cohen and Barnes, 1993; Cohen and Hill 2001), while others emphasize organizational norms, or internal accountability in their research (Abelmann et al, 1997; Newmann et al., 1997), although they are not mutually exclusive. This paper will mainly examine the Cohen and Barnes's conception, namely the pedagogy of policy, or the opportunity to learn for enactors (teachers) will be a more important factor affecting the change of teacher behavior in response to state standards.

This paper conducts an empirical study to examine the relative merits of these two theories. In other words, the two theories, the principal agent model and the power of local actors as highlighted here will be evaluated empirically. The study utilizes National Center for Educational Statistics (NCES) 2000 Schools and Staffing Surveys (SASS) in order to illuminate a major aspect of the efficacy of standards-based accountability. SASS data provide some information on the effect of state accountability

policy on the teaching practice through teacher and school level variables.

School-level organizational factors and individual teacher characteristics also could have an effect on teachers' behavior so these variables will be controlled in the analysis. This paper tries to answer the following questions:

- Do teachers in states with strong standards-based accountability policies show more consistent practice in their use of test data and state curriculum guidance?
- Or, are opportunities to learn through PD programs more important in determining their use of state standards or test results?

2. The Principal-Agent Model and Theory of Action

Accountability is now the first most important agenda in the education policy. President Bush's ambitious educational reform plan, No Child Left Behind, states that all states must implement annual testing in math and reading. All students need to be at the proficiency level until 2014.

Elmore and his colleagues (1996) identify the main components of accountability policy: 1) setting a target for student achievement at the school level and creating methods for measuring school performance (Standards and Fair Measurement System), 2) implementing statewide assessment to measure student performance at the school or district level (Statewide Test), and 3) rewarding, assisting, or punishing high or under performing schools (Incentive System)

Beyond these basic components, Elmore, et al. note that the success of the

accountability system will depend on the extent to which the following conditions are satisfied:

- School administrators, teachers, parents, and students understand the accountability system and know what to do to improve performance.
- The state has clear performance goals and systems that reward both improvements in student performance and the attainment of an absolute standard.
- The state has technical expertise and capacity in assessment and evaluation and assists schools to improve their performance.
- There is a stable political environment for the accountability system.

The assessment system is critical to the theory of action underlying accountability policy. Elmore and Rothman state that the theory of action of standards-based reform envisions that teachers armed with data on how students perform against standards will make the instructional changes needed to improve performance. Smith and O'Day (1990) note that clearly specified curriculum guidelines and high quality statewide tests are important instruments for systemic reform.

However, if the assessment system does not provide adequate information on the progress of student achievement, teachers cannot utilize these data. For instance, if the state provides only one snapshot test scores of students' achievement level, not growth over time, then teachers cannot know whether their instructional methods improved student achievement. An Education Commission of the States (2000) report says, "school personnel can learn a great deal from available student learning data, especially

when the data are *followed over time*, disaggregated and augmented with additional information.” So, nowadays, some scholars have argued that a value-added assessment system fits the purpose of the accountability system; that is, they can hold schools accountable for their contribution to improved student achievement, and therefore can make the theory of action work (Hanushek, 1994; Hanushek & Raymond, 2001; Clotfelter & Ladd 1996). Value-added assessment is even referred to as a revolution in accountability policy (J.E. Stone). Only when the state provides information about how much a teacher and school contributed to student achievement can they utilize the test results to improve performance¹⁴ (ECS, 2000; Sanders, 2000).

From this perspective, a well-designed value-added accountability system will change schools and teachers and produce higher quality public education. Thus, how well teachers know the standard and utilize student achievement test scores will be critical for the success of accountability. For this theory to hold at the school level, teachers must possess sufficient knowledge about both the curriculum framework and how to utilize test results to guide changes in their teaching.

However, accountability policy itself does not focus on such knowledge of teachers. The main conceptual framework of accountability policy, as noted above is based on the principal-agent model. The model attaches monetary reward and sanctions to visible student test outcomes. Such an incentive will lead teachers and schools (agents) to dedicate their energy to increase student achievement that is the main goal of the principal.

Therefore, a value-added assessment system plus monetary rewards or sanctions

¹⁴ Dr. William Sanders is the leading proponent of this argument. Recently, influenced by the Tennessee

attached to the student test scores will improve public education, since according to the principal-agent model, they provide sufficient conditions for the agents, teachers and school to improve test scores. A properly designed accountability system with value-added assessment and sufficient incentive system will drive teachers and schools to make more effort to increase student achievement which is the principal's interest.

3. Pedagogy of Policy and Capacity of Teachers and Schools

Some research demonstrates that teachers' opportunity to learn, or school circumstances influence the success of standards-based reform, or accountability policy (Cohen & Hill, 2000; Abelmanne et al., 1999; Spillane, 2000; Cohen 1996; Cohen 1990).

One example regarding the important role of teachers' knowledge in the success of standards-based reform comes from the story of Mrs. Oublier (Cohen 1990). The story shows that although California implemented an ambitious effort to revise mathematics teaching, a teacher's teaching practice did not change. Teachers' traditional practice and habits as well as school norms have a great effect on policy implementation. A California survey shows that more teachers are using new curriculum materials that are associated with the standards-based reform. However, the results are not uniform. Many teachers use the language of instructional and curriculum guidance in remarkably different ways (Cohen 1996). Recent Study by Cohen and Hill (2001) provide evidence that when teachers are provided with more opportunity to learn, such as, through profession development programs which enhances their understanding of the curriculum

guidance, they will ultimately alter their teaching practice.

Some scholars have examined the local (or organizational) context as the main determinant of variation in the success of accountability policy implementation. For instance, Abelman and Elmore (1999) argue that the internal school accountability system will be critical in the success of an external state accountability system. Using Abelman & Elmore's framework, Debray, et al., (2001) found that state policies interacted with existing school structures and norms to produce divergent responses across school types. They argue that the variation among schools within a state in response to the policies far exceeds the variation due to interstate differences in accountability policies. Further, they show that when a strong internal accountability system does not exist, individual teachers' knowledge or beliefs will play a key role in instructional practices. They comment, "Without a coherent school or departmental response to the policy, there was a range of teacher responses based upon their personal senses of responsibility and efficacy."

In summary, teachers and schools will change the standards-based accountability system and more attention in policy decision must be given to lower level enactors and the local organizational context. By this perspective, rather than external incentives or regulatory accountability policy, PD programs for teachers will induce the desired changes in teacher behavior. This paper will examine this notion of pedagogy of policy by evaluating whether opportunities to learn are critical for changes in enactor's practice.

Arizona, Florida, are creating or considering a kind of value-added assessment system (Kanstoroom, 2000).

Table 2.1: Underlying Theories of Accountability and Professional Development

1. Theory of Action of the Accountability Policy Framed by Principal Agent Model

Standards/Test Scores		Principal (State Government)'s Goal
Agent (Teachers)	→ Instructional Change	→ Increase of Students' Test Scores
	Monetary Rewards/Sanction	

2. Policy Implementation Generated by Opportunity to Learn

Teacher	→ Instructional Change	→ Increase of Student Learning
	Opportunity to Learn	

4. Current Literature on the Effect of Accountability Policy

4. 1 The Effect of Accountability Policy on Student Academic Achievement

There have been some debates on the effect of accountability policy on student achievement gain. Specifically, the debate has focused on the achievement gains of students in Texas and North Carolina, states with strong accountability policies. After finding the largest student achievement gain in the National Assessment of Educational Progress (NAEP) controlling the state level socio-economic variables and expenditure variable, Grissmer et al., (2000) attributes the gains to the aligned standard, assessment and accountability system. They find that there is large variation in NAEP gains across states and the variation is not explained by some major resource variables such as per pupil expenditure, student-teacher ratio, teacher resources and so on.

Case studies in North Carolina and Texas which attain the largest gain in NAEP

between 1990 and 1996 shows that these two states have strong accountability systems compared to other states. Grissmer, et al., (2000), thus, argues that the strong accountability policies in North Carolina and Texas are the main source for their achievement gain.

As soon as the Rand study by Grissmer, et al., (2000) was published, Klein et al., (2000), the other Rand researchers provided different story of the achievement gain in Texas. Klein, et al., (2000) shows that except for 4th grade math achievement, the gains in Texas were not significantly different from national trends. Klein, et al., (2000) also show that the NAEP achievement gap between white and students of color increased in Texas.

Actually, their studies' methodologies are somewhat different. Klein, et al. (2000) uses descriptive methods and correlation test, while Grissmer et al. (2000) use a more advanced regression model. So, is the Grissmer et al.,(2000) study more reliable? We cannot answer this question, since the key source of the limitation of both studies does not come from methodology, but the unit of analysis and their study design is different. Both of them used state-level aggregated variables. Grissmer, et al. (2000) argue that if state-level aggregate data can provide sufficient control variables, using state-level data will yield more valid result than using student-level data which do not have control variables such as family background and school characteristics in which students are nested. This argument is somewhat arbitrary, because using state-level data would produce unreliable results.¹⁵ Aggregated control variables cannot eliminate the

¹⁵ Recently, whether high stake accountability policy represented by graduation test will increase student achievement was debated (Amrein and Berliner, 2002a ; 2002b; 2003; Braun, 2004; Rosenshine, 2003). Their analysis focus on the high-stake graduation exam on the student achievement, using NAEP, SAT, and other achievement test. They provide different conclusions again. The similarity between the debates by

possibility of aggregation bias (Bryk and Raudenbush, 1992). In addition, both Grissmer, et al., (2000) and Klein, et al., (2000) only mention math score gains, not reading scores. Grissmer, et al., (2000) say that the reading data are not sufficient for their analysis, so they did not analyze them. The Grissmer, et al., (2000) study, despite of its better methodology, did not aim to illuminate the achievement gap pattern between student races. Achievement gap by student background can be analyzed well with descriptive and correlation test methods used by Klein, et al.

The Grissmer, et al., (2000) analysis also was challenged by Darling-Hammond (2000). She also analyzed NAEP data. However, she utilized teachers' characteristic variables as well as other state-level SES variables from the SASS database. Her analysis of state policy was implemented by looking at teacher policy, accountability policy, and other factors. She attributed the achievement gain of Texas and North Carolina to the states' teacher policies such as certification policy, rigorous PD, teacher education reform, increasing teacher salary, and so on. Darling-Hammond mentions that North Carolina Accountability policy (so called, ABC, Accountability, Basics, and Control) was fully implemented in 1997. Thus, it is hard to attribute the NAEP test gains to accountability policy since Grissmer's data analysis was based on the NAEP scores between 1990 and 1996. Darling-Hammond (2000) notes that further analysis adding information about parent education level, curriculum and testing approaches using the NAEP background surveys¹⁶ would shed greater light on school factors that matter.

Other recent analyses of the effect of accountability on student achievement using

Grissmer et al., and Klein, et al., and that by Amrein and Berliner, and Braun is that they used state-level aggregated data. Until student level database with student characteristic variables is available, we would only receive limited information from the state-level aggregated analysis, which will cause another debate.

district or city samples were done by Ladd (1999) and Roderick, et al. (2002).

Ladd (1999) looked at large Texas cities and measured the gains in student performance in Dallas which has stronger accountability relative to other cities. She found that Hispanic and white seventh graders achieved more gains, but not Black students. For third grade outcomes, the Dallas accountability program did not seem to work. However, her analysis was limited to only in some cities in Texas and did not produce general conclusion for the effect of the state-level accountability policy.

Roderick, et al., (2002) analyzed Iowa Test of Basic Skills (ITBS) results for Chicago public schools and found that introduction of a high-stake accountability policy boosted achievement in promotional gate grades. However, they mention that their analysis did not examine whether the achievement gains of the ITBS can be generalized to other achievement measures. In addition, since their analysis uses only Chicago Public School data, they cannot address the extent to which their findings can be generalized to other school districts and other context.

Recently Hanushek and Raymond (2003) analyzed the effect of accountability systems on student achievement using 1992 and 1996 NAEP data. They used a random effect model, which controls for time-invariant state-level variables. Hanushek and Raymond (2003) assume that other state policies are invariant over time and use dummy variables to represent states with strong accountability systems. They found that accountability policy has positive effects on achievement gain in math. However, reading scores were not significantly affected by the accountability system. They analyzed the effects by race, and found that Hispanic and black students did not improve

¹⁶ This NAEP survey data was used by Swanson and Stevenson (2002) to illuminate the effect of standards

NAEP achievement in math and reading as much as White students. They mention that whites gain more than blacks after accountability is introduced, so the achievement gap between white and black students widens with the introduction of accountability policy. Some of their results are consistent with Klein, et al., (2000): no effect of accountability on reading, and achievement gap between white and student of color widens. This could be because both Klein, et al. (2000) and Hanushek and Raymond (2003) used NAEP data during a similar period, early and mid 1990s.

Carnoy and Loeb (2002) also implemented a similar analysis with Hanushek and Raymond (2002), however they used different NAEP data periods, 1996-2000. Carnoy and Loeb (2002) developed a zero-to-five index of the strength of states' accountability systems and analyzed whether their accountability index is associated with student gains on the NAEP mathematics tests. They found that stronger accountability policy has a positive effect on the achievement gain in 8th grade NAEP math examination between 1996 and 2000. However, they did not find a significant relationship between 9th grade retention and accountability index. They mention that the 8th grade achievement gain in math must result in better performance in 9th grade. For instance, if 8th grade did a good job then 9th grade must show a higher graduation rate. However, the relationship between the 8th grade math gain and 9th grade retention rates is not so significant.

So far, different results about the effect of accountability policy on the student achievement based on the NAEP tests are prevailing. However, we can summarize the findings as following:

- Most analysis says that accountability policy does not have an effect on NAEP reading test scores. Usually the accountability policy has some positive effect on 8th grade math scores, but not 4th grade math.
- Whether accountability policy will close the achievement gap between white and students of color is not likely, and no strong evidence that the accountability policy will reduce the gap exist. Ladd (1999) shows that black students did not achieve significant gain compared to white and Hispanic. Hanushek and Raymond (2003) show that accountability may widen the achievement gap. Only Carnoy, et al., (2002) shows that strong accountability index is associated with math gain and the gain is larger for black students. However, they mention that the achievement gain is likely due to teaching to the test and test-taking skills.
- State-level aggregate data analysis produces another debate. Using different methods and purposes, researchers draws different conclusions, making the effect of accountability system more uncertain. Murnane and Levy (2001) also mentions that the different interpretations of the NAEP data underline the difficulty in evaluating the consequences for children of color of a particular standards-based reform effort. Without disaggregated data with more variables, future research will only confirm the difficulty (or even impossibility) of evaluating the effect of accountability systems on student achievement.

4. 2 Effect of Accountability Policy on Classroom Instruction

While some economists have focused on the effect of accountability on student test scores, education researchers have given more attention to the effect of accountability on teaching practice. This research is important since the underlying theory of action of an accountability system is that teachers provided with student test data and curriculum guidance will change their instructional practice. The monetary reward or sanction will make teachers change their behavior toward using state curriculum guidance and test results. Then, what if the strong accountability system improves student test scores without making teachers understand or utilize state curriculum guidance that is usually based on constructivism. We can infer that teachers may just teach to the test so only test-taking skills of students are improved. The improvement of student test scores may not be caused by the external accountability system, but by some other factors researchers have not found. Thus, in this regard, the research on the impact of accountability policy on teaching practice is important.

Cohen and Hill (2001) provide a profound analysis of the effect of standards-based reform in California. One of the main findings of their study is that changing teacher's classroom instruction by using high-stakes tests failed either to align the tests with the student curriculum or offer teachers substantial opportunities to learn about reform idea. Instructional improvement works best, when teachers are provided with opportunities to learn about specific academic content, the state's curriculum and its assessment for students.

Specifically, Cohen and Hill (2001) analyzed teacher survey data and ran the

regression to see what variables are associated with teachers' ideas and practices.

Independent variables included in their model are: measure of teachers' time in student curriculum workshops, their familiarity with reform, their attitude toward reform, and toward the California Learning Assessment System (CLAS), norms of collaboration, administrative support, and so on. They looked at whether these independent variables have an effect on teachers' ideas and practices related to the reform frameworks.

Interestingly, professional or organizational conditions such as administrative support for the state reforms, professional norms of collaboration were not significantly correlated with teacher's ideas and practices toward reform. Collegiality around mathematics had no effect. Professional norms of collaboration were associated with conventional ideas, which could imply that professional communities could be conservative as well as progressive, although the association is not statistically significant.

On the other hand, time in curriculum workshops and teacher attitudes toward reform had significant positive effects on teaching practices. Thus, teachers' opportunities to learn and attitude or familiarity with reform had positive effects on the teachers' usage of the reform framework.

Whether teachers can join the PD opportunities is another issue. Cohen and Hill (2001) investigate what factors will influence teachers' choice of mathematics PD. Teachers working at schools serving more low-SES students are less likely to have effective opportunities to learn. This could be evidence that students with low-SES will not benefit from accountability policy.

Swanson and Stevenson's (2002) analysis of NAEP teacher survey data shows some similar results to Cohen and Hill's (2001). They used hierarchical linear modeling

to examine the extent to which coherent standards-based reform has actually produced systematic policy activism across states, and whether the activism has an effect on teachers' use of instructional practices. Their analysis shows that teachers' knowledge from PD (instruction) or about NCTM standards and their attitudes about standards-based practices have significantly positive effects on standards-based instructional practice. State's standards-based policy activism has a positive effect only when the above variables are excluded.

Why does the effect of standards-based policy disappear when variables representing teachers' knowledge and attitudes are added to the model? Swanson and Stevenson (2002) says that state policies may support the adoption of standards-based instructional practices through the promotion of local level change, which might increase teacher knowledge about the reform, and shaping teachers' positive attitudes toward the reform. This interpretation is possible. However, this result could also imply that without enhancing teacher knowledge and attitude, the standards-based accountability policy itself does not work as an independent policy in changing teacher's instructional practice. Swanson and Stevenson (2002) developed a solid analysis using a national sample, however, their linear model did not include school-level variables which reflect some organizational features. If they had incorporated such variables in their analysis, it would have been more useful.

While Cohen and Hill (2001) and Swanson and Stevenson (2002) analyze the effect of standards-based accountability using a state and national sample, Barnes (2002) provides a specific story about how a high-poverty school experiences the standards-based reform from her qualitative analysis. Barnes (2002) shows that the outside

accountability policy will bring some conflicts inside of schools. Whether the conflicts will become productive or counterproductive depends on the schools' (teachers') capacity or resources which are personal, social, or professional. When capacities such as opportunities to learn, experience or knowledge, or professional norm are lacking or limited, the conflict brought by outside reform will be counterproductive. The problem is that high-poverty schools do not have such capacities. The schools that need to receive benefit from state policy are high-poverty schools. However, without additional provision to increase capacities and resources for the disadvantaged schools, external accountability policy will place more burdens or difficulties on schools serving disadvantaged children.

Newmann, King, and Rigdon (1997) provide some critical analysis of the effect or utility of external accountability policy. They selected twenty four schools which are equally divided among elementary, middle, and high schools and reflects a broad spectrum of locations, size, and student body. The schools come from different states. They analyzed whether the schools have strong external accountability system, strong internal accountability systems and internal capacity. Newmann, et al. (1997) found that schools with strong external accountability tended to be low in organizational capacity. Some schools generated strong internal accountability system internally by the school community without prescriptive mandates from a district or state. And outside state accountability system did not necessarily boost organizational capacity and sometimes, it caused contention between teachers in a school. They shows that when highly specific prescriptive standards connected to high-stakes consequences are mandated by state authorities, school staff can lose the ownership or commitment and authorities that they

need in order to work collaboratively to achieve a clear purpose for student learning.

However, they state that their findings are from only a small sample of schools, and they did not select schools from states with vigorous accountability systems. So, in the future, more extensive analysis using large samples would be necessary to support their findings more fully.

Debray, et al. (2001) also provide similar result to the Newmann, et al.(197) study, even though they also investigated only four schools. Debray, et al. (2001) found that the variation between the variation in internal accountability in response to the state accountability policies far exceeds the variations in state accountability policies. They show that when a strong internal accountability system does not exist, individual teachers' knowledge or belief will play key role in the instructional practices. Their argument shows that teachers' individual knowledge and personal sense that is inconsistent with reform must be controlled by the internal norm. However, they did not show how a school can construct a strong internal accountability system and what is the role of individual teachers in the process of building the strong internal accountability system.

It may be worth looking at whether school staff and teachers regard test results useful for improving their instruction. Supovitz and Klein (2003) provide research on whether teachers and school staff regard the student performance data useful, although their main purpose is to provide a framework on how to use student performance data to guide improvement in instruction. Supovitz and Klein (2003) say that there are three general sources of student performance data: external assessment data, individual teacher assessment data, and school-wide assessment data. Among these performance data,

accountability systems usually imply that schools and teacher will benefit from the external assessment data. In their survey, teachers and administrators feel that the state test results are moderately or minimally useful, providing limited information. External test scores do not provide adequate details to guide teaching and learning. Only around half of the leaders responded that state tests provide adequate information. And less than half of school leaders responded that the state tests were not timely enough to inform classroom instruction. Surveys of perceptions regarding the utility of different data sources also illuminate that overall school leaders feel that internal data have greater value for providing instructional guidance. Student portfolios, open-ended assessments, and running records that are developed inside of the school are considered more useful than external state or district assessment.

Ladd and Zelli's (2002) survey also shows that the statewide standardized test results were not seen as a good measure of a students' mastery of the curriculum by principals in North Carolina. Less than half of principals agreed that end of grade test can measure students' mastery of the material taught in schools well.

In summary, research about the effect of accountability policy on teachers' instruction provides the following tentative lessons.

- Teachers' knowledge of standards and their attitudes toward reform appear to be key determinants of the success of outside accountability policy.
- Internal accountability can be created by a school community, not by external state accountability system. Variation in internal school accountability exceeds the variation of state accountability policy. The problem is that current state accountability policy brings a conflict which could be

counterproductive to the enhancement of internal school accountability of the high-poverty schools.

- Statewide test results are somewhat useful for improving instruction, however, teachers think that internal assessment is more useful. In addition, statewide tests so far have not convinced teachers and school leaders that the statewide standardized test measure student learning adequately.

Since these lessons come from the limited research currently available, they are subject to challenged. As more systemic and comprehensive research on the effect of accountability on instructional practice grows, we should be able to confirm the above lessons or find different lessons.

5. Research Design

Recently Carnoy and Loeb (2002) developed a zero-to-five index of accountability measure. Five means that the state builds the strongest accountability system, while zero means the state did not construct any relevant policy instrument for the accountability policy. The higher the index is, the stronger the accountability is in the state. An accountability index was created by examining testing grades, repercussion for schools, strength of repercussion, high school exit exam, and so on (Carnoy and Loeb, 2002). Appendix provides the accountability index table created by Carnoy and Loeb (2002, p.324-326).

The accountability index in Appendix shows that North Carolina, Texas, and

California are implementing strong accountability policy as the literature review shows in the previous section. Kentucky and Florida, which are also known for the implementation of strong accountability policies, have higher index scores. Thus, Carnoy and Loeb's index appears to be an appropriate measure of the strength of a state's accountability policy. Using this index, we categorize states into two groups; one group is composed of states having strong accountability indexes of four or five, and another is composed of other states which does not have strong accountability. Alabama, California, Florida, Kentucky, Maryland, New Jersey, New Mexico, New York, North Carolina, and Texas are included in the strong accountability group. Table 2.2 exhibits that the average index value is 4.6 for these ten states and 1.5 for the other states. Thus the difference of average accountability index of these two groups is large and it implies that teachers in the strong accountability states confront much more external accountability pressure than those in the states with weak or no accountability policy.

As of 1999-2000, all ten of these states have strong repercussion for schools. These include ratings/intervention (Alabama, Texas), monetary awards/ intervention (California, Kentucky, Maryland, New Mexico, North Carolina), ratings/subject to vouchers (Florida), or audits/state takeover/freeze on pupil registration (New Jersey, New York). States with weak or no accountability policy do not have strong repercussion for schools except for West Virginia (see the column of strength of repercussion in Appendix 1 and do not provide any monetary reward except for Pennsylvania. West Virginia only has intervention for schools and has marginally strong accountability index, 3.5. Pennsylvania provides money only for high school improvement and does not have any strong accountability policy for K-8 schools.

Table 2.2: Two Groups by the Intensity of Accountability in 1999-2000

<u>States with Weak Accountability</u>		<u>States with Strong Accountability</u>	
State	Index	State	Index
Alaska	1	Alabama	4
Arizona	2	California	4
Arkansas	1	Florida	5
Colorado	1	Kentucky	4
Connecticut	1	Maryland	4
Delaware	1	New Jersey	5
Georgia	2	New Mexico	5
Hawaii	1	New York	5
Idaho	1	North Carolina	5
Illinois	2.5	Texas	5
Indiana	3	Average Index Score	4.6
Iowa	0		
Kansas	1		
Louisiana	3		
Maine	1		
Massachusetts	2		
Michigan	1		
Minnesota	2		
Mississippi	3		
Missouri	1.5		
Montana	1		
Nebraska	0		
Nevada	1.5		
New Hampshire	1		
North Dakota	1		
Ohio	3		
Oklahoma	1		
Oregon	2.5		
Pennsylvania	1		
Rhode Island	1		
South Carolina	3		
South Dakota	1		
Tennessee	1.5		
Utah	1		
Vermont	1		
Virginia	2		
Washington	1		
West Virginia	3.5		
Wisconsin	2		
Wyoming	1		
Average of Index Score	1.5		

(Source: Carnoy and Loeb, 2002)

High school exit exams have been implemented in 21 states. Among the 10 states with strong accountability policy, California and Kentucky did not have high school exit exams as of 1999-2000. Thirteen states without strong accountability systems have high school exit exams. So, accountability index is mainly based on the strength of repercussion. Since we can infer that high school exit exams will mainly have an effect on teaching practice of high school teachers, we want to examine the effect of these exams on teachers' practices. However, this will need to reduce the sample only to the high school teachers. Actually, this is related to the issue of the impact of high-stake testing on high school students and somewhat deviated from the issue of accountability policy such as implementing annual test for 3 through 8 grade and rewarding and sanctioning schools based on the achievement test¹⁷.

A dummy variable for teachers from these ten states with strong accountability policies would exhibit whether teachers under the strong pressure of accountability policy change their behavior more than teachers from states which do not implement strong accountability policies.

The other variables we are interested in are PD program participation variables. We will examine the effect size of teachers' PD program participation on their instructional practice. That is, we will assume that teachers who participated in PD programs will lead them to change their instructional practice than those who did not, and compare the PD participation variable with accountability policy dummy variable.

As mentioned earlier, this paper does not measure the impact of accountability policy on the student test scores. Rather the underlying two theories explained in the

¹⁷ We can create a dummy variable for the states with high school exam to see the effect of high-stake

previous section will be investigated. The theory of principal-agent model says that teachers under strong accountability policy would use students' test scores and state standards more frequently and rigorously. The test scores published by the state and the rewards and sanctions attached to the test scores will drive teachers to use state curriculum and test information to change their instructional practice.

However, the second theory, pedagogy of policy, suggests an alternative hypothesis. It predicts that teachers' opportunities to learn represented as PD program participation will be key determination of teachers' use of test results and state curriculum guidance more than the external accountability policy. Teachers in other states which have relatively weak accountability system will utilize the test results if they have sufficient knowledge on how to use test results and curriculum. Teachers, despite of the external strong accountability policy, will not use children assessment results or state standards if they do not have knowledge on how to use test results and state's curriculum guidance.

Thus, we have a following question to test:

Between strong external accountability policy and opportunities to learn, which will have more effect on the teacher's use of state standards and state-wide test information to change or improve their instructional practice?

6. Data¹⁸

testing on teaching practice of high school teachers. However this will be outside the topic of this chapter, although it may be worth investigating.

The National Center for Educational Statistics has conducted a national teacher and school staff survey, the Schools and Staffing Survey (SASS). SASS has been implemented in school years 1987-1988, 1990-1991, 1993-1994, and 1999-2000. SASS uses stratified random sampling to represent the national population. SASS surveys teachers, principals, administrators, district administrators and includes public, charter, private schools.

We utilize 1999-2000 SASS public school teacher survey data and public school administrator survey data which provide many useful variables at the teacher and school level. Importantly, the database provides critical information about whether teachers have participated in PD programs and to what extent they use state or district standards and test results.

Data from the private school survey, charter school survey, and Indian affairs schools were excluded. The basic unit of analysis is the teacher level and school level variables are incorporated into the analysis. Also, the survey provides information about whether teachers work as full-time or part-time. Since accountability policy would affect public school full-time teachers, only full-time public school teachers are included for the analysis. This reduces the sample size from 42,086 to around 38,375. Appendix provides sample number of full-time teacher by state.

The sample size is subject to change by including some variables because some teachers did not respond to the survey items. The STATA software provides information about the represented population number by the sample size, so we can see how much

¹⁸ I appreciate my colleagues, Debbi Harris and Marisa Burian-Fitzgerald for helping me to manage the SASS database and obtain some variables.

population is represented by the analysis¹⁹. There is a trade-off between inclusion of more variables and the representation of the reduced sample size. Including more variables will control for other factors and provide more reliable estimate of the effects of PD, accountability policy, and other independent variables we are interested in. However, this could reduce sample size and the population represented in the analysis. This will be discussed when results of various models are presented.

7. Methods

7.1 Regression Model and Independent Variables

The basic method to measure the effect of PD programs and accountability policy is regression technique. Full regression model would be:

Teacher's use of state standard or test

$$= B_0 + X_1B_1 + X_2B_2 + X_3B_3 + X_4B_4 + X_5B_5 + u, \quad (1)$$

where X_1 is the vector of teacher characteristics such gender, race, college major, and so on. X_2 is the vector of teacher perceptions. X_3 is the vector of school characteristics and X_4 is the vector of teacher PD relevant variables, X_5 is a dummy variable indicating

¹⁹ STATA software easily can deal with survey data, incorporating weighting variables such as teacher final weight and strata or clustering, so it produces reliable estimation on the effect of independent variables and robust standard error for the clustering survey data.

whether teachers are from the states with strong accountability policy.

An ordered probit regression model and an ordered logit regression model will be employed since the dependent variable is an ordered response of teachers on the survey question (please refer to Appendix). For instance, to answer the survey question, to what extent the teacher uses state/district standards to guide his/her instructional practice, which is used as the dependent variable, the teacher needs to choose one to five scale of answer where five means to a great extent and one means not at all (see Appendix). So, five means that the state or district standards guide teachers' instructional practice to a great extent. It is hard to say that the scale exactly has the numeric mean. The difference between scale four and scale two does not necessarily mean that it is twice as influential as the difference between scale one and two. We can only know that five means more influence of state or district standards than four, and four means more influence than three, in other words, the response scale has ordinal meaning. However, linear regression result also will be provided to check whether ordinary least square linear regression produces significantly different results compared to the ordered probit model and ordered logit model. If so, we better use ordered probit or ordered logit model. Otherwise, looking at the results of linear regression model for the convenient interpretation of coefficient size will be fine.

The ordered probit or ordered logit model illuminates the effect of accountability policy and PD on the ordered response, controlling for other school-level and teacher characteristics variables like the usual linear regression model. The signs and statistical significance of coefficients can be interpreted the same as the linear regression results. For instance, if the positive coefficient of the variable that indicates participation in PD

means that teachers who join PD are more likely to use state or district standards to guide their instructional practices. The specific magnitude of the coefficient only can be obtained by a complicated calculation. Table 2.3 provides the definitions of variables used in the analysis.

Table 2.3: Definitions of Dependent and Independent Variables Used in the Analysis

Independent Variables:

Teacher (Basic) Characteristics Variables

Male: Dummy variable which takes on the value 1 if the teacher is male and 0 if the teacher is female.

Minority: Dummy variable which takes on the value 1 if the teacher is minority and 0 if the teacher is white.

Age: Continuous variable indicating the age of teacher.

TotExp: Continuous variable. Total teaching experience measured by year.

Sqtotexp: Continuous variable.

Square value of total teaching experience measured by year.

Salary: Continuous variable. Teacher Annual Salary

Unionmem: Dummy variable which takes on value 1 if the teacher is union member, otherwise 0

Remaassi: Dummy variable which takes on value 1 if the teacher's main teaching assignment field is reading or math.

HRSMATH: Teaching hours in math per week during most recent full week of teaching.

HRSEng: Teaching hours in reading/English per week during most recent full week of teaching.

Teacher Knowledge or Ability Variables

MA: Dummy variable which takes on value 1 if the teacher has a master degree, otherwise 0.

Mathalba: Dummy variable which takes on value 1 if the teacher's college major is math or math education.

English: Dummy variable which takes on value 1 if the teacher's college major is English/language arts or English literature or composition.

Verycomp: Selectivity of undergraduate institution. Dummy variable which takes on value 1 if the teacher's undergraduate institution is very competitive, highly competitive or the most competitive, 0 if the teacher's undergraduate institution is competitive or less competitive, non competitive or special. This selectivity of undergraduate institution is from the ratings of Barron's 2001 Profiles of American Colleges. This

variable can be a proxy for the teacher's innate ability.

Certrec: Dummy variable which takes on value 1 if the teacher obtained teaching certification which is regular, advanced, provisional or probational in her/his main teaching assignment, 0 if the teacher reports that temporary, emergency or no certification.

Teacher Professional Development

PDindepth: Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on in-depth study of the content in his or her main teaching assignment field in the past 12 months. 0 means the teacher did not participate.

PDstandards: Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on content and performance standards in his or her main teaching assignment field in the past 12 months. Otherwise 0.

PDmethodte: Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on methods of teaching in the past 12 months. Otherwise 0.

PDassessme: Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on student assessment, such as methods of testing, evaluation, performance assessment, etc in the past 12 months. Otherwise 0.

PDdiscipline: Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on student discipline and management in the classroom in the past 12 months. Otherwise 0.

Teacher Perception Variables ²⁰

ZInfluence: Continuous (scaled) variable. Higher score indicates higher perception of influence in school policy such as setting performance standards for students, establishing curriculum, evaluating teachers, hiring new full-time teachers, setting discipline policy, deciding the usage of school budget, and determining the contents of in-service professional development program.

ZControl: Continuous (scaled) variable. Higher score indicates that the teachers perceive that they have much control over following areas such as selecting textbooks and other instructional materials, selecting content, topics, and skills to be taught, selecting teaching techniques, evaluating and grading student, disciplining students, and determining the amount of homework to be assigned.

²⁰ Please see the appendix A in the working paper, Debbi Harris (2002), Lowering the bar or moving the target: A wage decomposition of Michigan's charter and traditional public school teacher, for more information about these scaled variables. The paper is available at www.epc.msu.edu. Also please refer to Wolfe, E. W., Ray, L. M., & Harris, D. C. (in press). A Rasch analysis of three measures of teacher perception. *Educational and Psychological Measurement*.

ZStudent: Continuous (scaled) variable. Higher score means that teachers perceive no serious student problem and low score means that teachers perceive serious student problem. Examples of student problems are: student tardiness, absenteeism, robbery of theft, pregnancy, alcohol, and so on.

ZClimate: Continuous (scaled) variable. Higher score means that teachers perceive a worse school climate and lower score means that teachers perceive a better school climate.

School Variables

PerFRLkw: Continuous variable. Percentage of student receiving free or reduced lunch.

NewminPER: Continuous variable. Percentage of student of color

Totalenroll: Continuous variable. School size. Total enrollment of student.

Stutearatio: Continuous variable. Student teacher ratio.

Suburban: Dummy variable that takes value 1 if the school is located at suburban area

Accountability Policy Variable

Strongacc: Dummy variable which takes the value 1 if the teacher is from the states with strong accountability policy (Alabama, North Carolina, Texas, California, Florida, New Jersey, New York, New Mexico, Kentucky, Maryland), 0 otherwise.

Dependent Variables²¹

Useofstandard : Scale is one to five. One means that the teacher does not use state/district standards to guide his/her instructional practice at all. Five means that the teacher uses them to a great extent. Higher scale means more use of state/district standards to guide his/her instructional practice.

Useforarea: Scale is one to five. One means that the teacher does not use state or local achievement test information to assess areas where he/she needs to strengthen his/her content knowledge or teaching practice at all. Five means teacher uses them to a great extent. Higher scale means more use of test result to assess areas where the teacher need to strengthen content knowledge and teaching practice.

UseTcurri: Scale is one to five. One means that the teacher does not use state or local achievement test information to adjust his/her curriculum in areas where his/her students encountered problems at all. Five means that teacher uses them to a great extent. Higher scale means more use of test result to adjust curriculum in areas where students encountered problems.

Usegrouping: Scale is one to five. One means that the teacher does not use state or local achievement test information to group students into different instructional groups by achievement or ability at all. Five means that

²¹ Appendix provides related survey questionnaires on these dependent variables.

teacher uses them in such a way to a great extent. Higher scale means more use of test information to group students by achievement or ability.

These independent variables capture most teacher and school characteristics that education researcher have been interested in. Previous empirical research did not find a good proxy for teacher's innate ability, however, in this analysis, selectivity of teachers' college (Verycomp) variable was included as a proxy variable for teachers' innate ability. Recently this variable was also used by Clotfelter, et al. (2003) and Lankford, et al (2002) as a proxy variable to capture teachers' innate abilities.

Accountability policy heavily focuses on two areas, math and reading. The No Child Left Behind Act also specifies that all students' math and reading scores must be at the proficiency level by 2013-2014 school year. Thus, accountability policy should have more effects on the teachers whose main teaching areas are math or reading.

Independent variables such as HRSMath, HRSReading, and Remaassi are created to capture whether math and reading teachers are more likely to adopt their instruction following the state/district standards. Relative teacher knowledge variables, whether undergraduate major was math, math education and English/Art and reading were included to see whether teachers' math or reading college major could have a positive (or negative) effect on teachers' use of state/district standards and student test scores.

In addition to total experience variable, square of total experience variable is created, since the relationship between experience and teachers' use of state or district standards could have a parabolic shape. Age and experience, and possibly salary could be correlated with each other, then their coefficients might not be significant. However,

the multicollinearity would not be a problem since we are not interested in looking at the effects of these variables. Or, dropping one of these variables could solve this problem.

Teacher perception variables, especially Zinfluence and Zclimate, provide some information on the internal accountability of schools, or they may capture dimensions of organizational culture. The variables indicate whether teachers did participate in the PD program. Teacher PD variables also provide information on the content of PD programs teachers joined.

7.2 Self-Selection Problem and Quality of Professional Development Program

There are two emerging issues with these PD participation variables. First is that the programs are not likely to be mandatory, so there could be a self-selection problem. The participation rate on each PD program varies between states (see Appendix). Thus it is likely that teachers are voluntarily selecting the PD programs based on their own preferences. This will bring the self-selection problem in the analysis. If some teachers' characteristics and school-level factors which are not controlled in the equation are correlated with the teachers' participation in PD programs, then the estimator will be biased. We may expect that the self-selection would cause overestimation of the true effect of the PD participation.

There are some ways to solve this self-selection problem. First of all, controlling for sufficient variables which could be correlated with the decision of teachers to participate in the PD programs would alleviate self-selection problem. Actually, this is what we expect: most factors that are possibly correlated with

participation are controlled in the full model (1). For instance, teachers who are working with more disadvantaged students are more likely to attend to the PD programs.

Including school level variables for students' socio-economic status will alleviate the self-selection caused by such school factors. Alternatively, teachers may be more likely to participate in PD if they wish to overcome limitations in their knowledge or ability regarding teaching and learning. In this case, controlling for teacher knowledge variables or ability with such variables as college degree, major, or selectivity of college would address the self-selection problem. Since numerous control variables are incorporated in the full model, we suspect that the self-selection problem would not be a serious problem.

Other solutions might be to use panel data methods such as random effect model or to use instrumental variable (IV) for program participation. Since the SASS database is cross sectional data, it is not possible to use panel data methods in this case. Instead, using IV is worth trying. We will try to find and use an IV for program participation. This would assume that X4 (PD program participation variables) is correlated with u in the following equation (1):

Teacher's use of state standard or test

$$= B_0 + X_1B_1 + X_2B_2 + X_3B_3 + X_4B_4 + X_5B_5 + u \quad (1)$$

And if the valid instrumental variable for X4 is Z1, then the Z1 must be uncorrelated with u in the equation (1) and $\Theta_1 \neq 0$ in the reduced form equation (2), which means that Z1 has a partial correlation with X4.

$$X_4 = \delta_0 + Z_1\theta_1 + X_1\delta_1 + X_2\delta_2 + X_3\delta_3 + X_5\delta_5 + k \quad (2)$$

One possible candidate as an IV would be expenditure for PD program. We can imagine that the possibility of teachers' participation in the program will be increasing when more programs are provided. Since district must spend more money to support such programs, expenditures on PD programs could be correlated with teachers' participation. We can infer that the expenditure would not affect teachers' use of standards directly and is not correlated with u .

The problem with using PD expenditures as an IV is that states do not provide information on expenditures specifically for PD program support. Only expenditures of very general instructional support services are provided and it includes expenditures on the supports for speech therapists, guidance counselors, and school nurses as well. This may result in the weak partial correlation between IV and the program participation variable and could cause asymptotic bias on the estimator of IV. We will use the Michigan portion of the SASS sample to check whether expenditures for instructional support services can be a good IV for PD participation.

Another possible candidate for an IV is the number of PD programs offered to school or district administrators by district²². Districts which provide more PD opportunities to school or district administrators are also likely to provide more PD opportunities to teachers. And we can infer that the number of PD program may not have a direct effect on teachers' use of state or district standards. Also, some teachers'

²² Please see Appendix for the relevant SASS questionnaire.

preference which is hidden in u is not correlated with the number of PD programs for administrators. We will use the number of PD programs for staff and administrators as the IV and discuss the result of IV estimate when analyzing the effect of standards-related PD participation on teachers' use of state standards for their classroom instruction.

The second issue is the quality of PD programs. We have assumed that PD programs have good and useful contents to help teachers change their instructional practice. If the programs are ineffective and have poor quality, then the theory of opportunities to learn would not work and the comparison between two implementation theories will be invalid. Literature raises this concern that PD programs did not contribute to substantial learning for teachers (Hawley and Valli, 1999). Most workshops, conferences, and other PD programs are so wasteful that they did not lead teachers to change in practice significantly (as quoted in Hawley and Valli, 1999).

The SASS database fortunately provides information about the usefulness of PD programs as rated by teachers. Teachers are asked to rate the usefulness of PD activities that they participated in over last 12 months. Teachers who think that the program was very useful would mark on number 5. Teachers who regard that the program was not useful at all would mark on number 1. Thus, if average rate is 3 then it would mean that teachers are kind of neutral to the usefulness of the programs²³. Following Table 2.4 provides the average rate of the usefulness of PD programs evaluated by teachers.

²³ Please see Appendix for the questionnaires.

Table 2.4: Average Usefulness of Five Professional Development Programs Rated by Teachers.

Professional Development Activities	Estimate	Std. Err.	Observation	Population
Usefulness of PDindepth	3.956	0.0101	20,812	1,616,789
Usefulness of PDmethodte	3.726	0.0096	26,847	1,999,429
Usefulness of PDstandards	3.638	0.0097	26,588	2,002,183
Usefulness of PDdiscipline	3.585	0.0138	15,974	1,122,059
Usefulness of PDassessme	3.518	0.0106	23,125	1,746,671

*Note: the definition of each category on the professional development activities is provided in Table 2.3. Estimates and standard errors are weighted and robust since they are obtained after controlling for sampling weight and correlation within stratum.

The average rate on the usefulness of PD activities related to in-depth study on main teaching area received highest rate, 3.956, which implies that teachers think that the programs on in-depth study are (somewhat) useful. In the case of PD programs regarding standards, the average is 3.726, which lies between neutral, 3, and somewhat useful, 4. Thus, the quality of PD programs regarding standards is weakly modest. So, the effects of PD programs in the equation (1) will be conservative estimates, considering that the teachers' perception on the usefulness of the programs is weak or moderate. That is, if states and districts offered more effective and useful programs, then the effects of PD program would increase.

7.3 Dependent Variables and Summary Statistics

Analysis will be implemented using four dependent variables, use of standards, use of student test scores to assess areas where the teacher needs to strength his/her content knowledge or practice, use of student test scores to adjust curriculum, and use of test scores to group students by achievement or ability level.

One thing we need to mention is that among these dependent variables, the last one, the extent that teachers use test score to group students by achievement or ability level within classroom is a somewhat complicated issue (Good & Brophy, 2002). However, enough problems with within-class ability grouping have been reported to cause educators to question and avoid this practice. Generally recent trend is away from ability grouping toward whole-classroom instruction (As quoted in Good & Brophy, 2002, p. 274).

Also, the current reform recommends that the practice of grouping student by ability within classrooms should be ceased. For instance, standards-based reform or accountability curriculum guidance emphasize that “all” students must achieve high level of learning. In California, one of key reform positions is that grouping students by ability should be ended (Cohen and Hill 2001, p.68).

North Carolina Mathematics curriculum standards guidance says that “every” student is challenged to meet a higher standards and fluency in mathematics is an expectation for all students (NC Department of Public Schools, 2003). North Carolina mentions that ability grouping which is commonly practiced does little to reduce the achievement gap and the state certainly discourages schools to group students by ability. Thus, the dependent variable, whether teacher are more likely to use test result to group students by achievement or ability level, will mean that whether accountability policy enhance or discourage such practice. If teachers who are from strong accountability states are more likely to group students by achievement or ability, then this could be evidence that accountability policy yield unintended teaching practice.

Table 2.5: Summary Statistics for Variables

	Mean	Std. Err.	Observation	Pop. Size
Useofstandard*	4.127	0.008	38,375	2,727,067
Useforarea*	3.640	0.013	22,115	1,722,596
UseTcurri *	3.789	0.012	22,115	1,722,596
Usegrouping*	2.587	0.015	22,115	1,722,596
Age	42.236	0.090	38,375	2,727,067
Salary	39,928.240	99.506	38,375	2,727,067
Remaassi	0.177	0.003	38,375	2,727,067
HRSEng	10.444	0.080	14,328	1,410,737
HRSMATH	5.348	0.046	14,328	1,410,737
Mathalba	0.040	0.001	38,375	2,727,067
English	0.066	0.002	38,375	2,727,067
Male	0.255	0.003	38,375	2,727,067
Minority	0.160	0.003	38,375	2,727,067
Totexper	14.808	0.084	38,375	2,727,067
Sqtotexp	321.613	2.894	38,375	2,727,067
Unionmem	0.797	0.003	38,375	2,727,067
MA	0.459	0.004	37,994	2,709,439
Verycomp	0.269	0.004	38,375	2,727,067
Certrec	0.930	0.002	38,375	2,727,067
Zclimate	0.026	0.008	38,375	2,727,067
Zcontrol	-0.028	0.008	38,375	2,727,067
Zinfluen	-0.019	0.008	38,375	2,727,067
Zstudent	-0.032	0.008	38,375	2,727,067
Totalenroll	825.902	4.165	35,333	2,495,093
PerFRLkw	38.582	0.256	34,421	2,455,204
NewminPER	34.977	0.270	38,214	2,718,586
Stutearatio	15.830	0.031	35,333	2,495,093
Suburban	0.501	0.004	38,375	2,727,067
PDdiscipline	0.411	0.004	38,375	2,727,067
PDindepth	0.593	0.004	38,375	2,727,067
PDmethodte	0.733	0.004	38,375	2,727,067
PDassessme	0.640	0.004	38,375	2,727,067
PDstandards	0.734	0.004	38,375	2,727,067
Strongacc	0.351	0.003	38,375	2,727,067

(* indicates that the variable is a dependent variable. Other variables are independent variables)

Table 2.5 displays summary statistics for dependent and independent variables.

This summary statistics table provides estimated mean, standard error, the observation

number which excludes missing cases, and equivalent population size which are represented by the observation number. The weighting variable and strata variable enable us to obtain the estimated mean and standard error that approximately capture the population statistics. Also, population size is obtained from the sample observation.

According to the NCES SASS data guidance book, the total headcount of teachers in 1999-2000 is 2,984,781 (U.S. Department of Education, NCES, SASS, 1999-2000, 2002). Thus, the population size of all public school teachers is 2,984,781. However, the population size for variables in the summary statistics is usually 2,727,067, since part-time teachers are excluded in the analysis. Exceptions exist. Two variables, HRSEng and HRSmath have smaller observation numbers, so the represented population size is reduced to 1,410,737, which is almost half of 2,727,067. The same sample number of these two variables could mean that teachers who have taught math also have taught reading recently.

Average math teaching hour is around five and average reading teaching hour is ten. The sample size, and equivalent population size will be reduced, if we include these two variables in the model. The smaller sample resulted from the inclusion of these two variables would narrow the analysis to the teachers who responded that they have taught some hours in math or reading recently. This might control for the effect of accountability policy on the teachers who really teach math and reading recently. However, in order to estimate the effect of the accountability policy and PD on teaching practice for a larger population regardless of teaching hours, we will also present the models which exclude these two variables, HRSMath and HRSread. Instead, the dummy variable, Remaassi will control for whether the teacher's main teaching field is math or

reading.

Although we exclude these two variables, the sample size will be slightly reduced to 2,455,304, if we include the variable, percentage of student receiving free/reduced lunch program which has been used to control SES in most education research. This slight decrease in sample size will be a trade-off to control SES factor at the school level.

The summary statistics also shows that about 46 percent of full-time teachers earn masters degree in their main teaching area. And most teachers, 93 percent, obtained certification. Male teachers comprise only one quarter of full-time teachers. PD variables show that 73 percent of teachers have participated in PD program focusing on teaching methods and standards area. Participation in PD on discipline is less likely than other PD programs since only 41 percent of teachers have joined such PD programs. Twenty seven percent of teachers earned their bachelor's degree from undergraduate institutes are ranked as the most, highly, or very competitive. Thirty five percent of teachers are from the states with strong accountability policy.

Most analysis will be given to the effects on the first dependent variable, Useofstandard, that is, the effects of standards-related PD participation status and strong accountability on teachers' use of state or district standards. For the other three dependent variables which are related to teachers' use of students' test score, similar but short analysis will be implemented without examining IV estimation.

8. Results

8. 1 The Effect on the Teacher's Use of State/District Standards

8.1.1 Specification and Change of the Size of Coefficients

First we examine relatively simple models to see the effects on the teachers' use of state/district standard for their instruction. Table 2.6 shows four ordered probit models that have different specification.

Table 2.6: Four Ordered Probit Models: the Effect on the Use of State Standards

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>		<u>Model 4</u>	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Salary					-0.000004**	0.000001	-0.000004**	0.000001
Remaassi					0.143**	0.028	0.162**	0.029
Mathalba					-0.149**	0.04	-0.122**	0.042
English					-0.100**	0.036	-0.077*	0.038
Male					-0.344**	0.019	-0.315**	0.021
Minority					0.116**	0.027	0.083*	0.032
Totexper					0.007	0.004	0.010*	0.004
Sqtotexp					-0.00003	0.0001	-0.0001	0.0001
Unionmem					-0.055*	0.022	-0.057*	0.023
MA					0.033	0.021	0.026	0.022
Verycomp					-0.045*	0.021	-0.036	0.022
Certrec					0.093*	0.036	0.103**	0.037
Totalenroll							-0.00012**	0.00002
PerFRLkw							0.002**	0.001
NewminPER							0.001	0
Stutearatio							0.012**	0.003
Suburban							0.086**	0.021
PDdiscipline			0.039*	0.019	0.034	0.019	0.025	0.021
PDindepth			0.173**	0.02	0.156**	0.02	0.141**	0.021
PDmethodte			0.111**	0.021	0.100**	0.021	0.096**	0.022
PDassessme			0.163**	0.02	0.136**	0.02	0.127**	0.021
PDstandards	0.422**	0.02	0.279**	0.022	0.255**	0.022	0.249**	0.023
Strongacc	0.227**	0.02	0.203**	0.02	0.201**	0.021	0.186**	0.023

Note: • Dependent Variable: Useofstandard. • Pweight: tfnlwgt. Strata: state. Number of strata: 51, for model 1, observation number is 38,375, population size is 2,727,066. For model 2, observation number is 38,375, population size is 2,727,066. For model 3, observation number is 37,994, population size is 2,709,439. For model 4, observation number is 34,109, population size is 2,440,181. • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.

Model 1 only includes two variables, accountability variable and standards-related PD participation variable. Model 2 adds other four PD variables in the equation. Model 3 includes teacher level variables and Model 4 has school level variables. In Model 1, the coefficient of the dummy variable, PDstandards, which indicates whether teachers participated in PD program related to standards, is much larger, almost twice, than the coefficient of accountability dummy variable. However, as we include more independent variable into Model 2 through 4, the coefficient of the variable, PDstandards, decreases significantly from 0.422 to 0.249.

Decrease of the coefficient of the PD participation variable, PDstandards, implies that self-selection problem possibly exists and the initial estimate of the effect of standards-related PD program participation in Model 1 was overestimated. Controlling teacher characteristics and school characteristics variables reduce the estimate significantly and systematically, thus it alleviates the self-selection problem. However, the difference between the effects of accountability policy variable and PD participation variable, PDstandards, remains significant, in spite of the decrease of the coefficient of PDstandards. The Wald test shows that the effect of standards-related PD program participation is larger than that of accountability policy variable in Model 4 and it is statistically significant at the one-tail test ($H_0: B_{\text{Strongacc}} = B_{\text{PDstandards}}$, $H_1: B_{\text{strongacc}} < B_{\text{PDstandards}}$, $P\text{-value} = 0.027$).

In addition, we find that the effect size of accountability policy is also decreasing systematically as we accommodate more control variables. Below Table 2.7 exhibits how much the coefficient size of the accountability policy variable, Strongacc, will be changing by the each model.

Table 2.7: Change of Coefficient Size on Accountability in Ordered Probit Model

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>		<u>Model 4</u>	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Salary	-0.000005**	0.000001	-0.000006**	0.000001	-0.000006**	0.000001	-0.000005**	0.000001
Remaassi	0.139**	0.028	0.167**	0.028	0.191**	0.029	0.186**	0.029
Mathalba	-0.173**	0.040	-0.162**	0.040	-0.127**	0.042	-0.115**	0.042
English	-0.103**	0.036	-0.074*	0.036	-0.056	0.038	-0.056	0.039
Male	-0.400**	0.019	-0.358**	0.020	-0.326**	0.021	-0.289**	0.021
Minority	0.159**	0.027	0.170**	0.027	0.069*	0.032	0.058	0.032
Totexper	0.013**	0.004	0.017**	0.004	0.021**	0.004	0.015**	0.004
Sqtotexp	-0.0002	0.0001	-0.0003**	0.0001	-0.0004**	0.0001	-0.0002*	0.0001
Unionmem	-0.036	0.022	-0.020	0.022	-0.016	0.023	-0.035	0.023
MA	0.038	0.020	0.051*	0.021	0.041	0.022	0.036	0.022
Verycomp	-0.041	0.021	-0.027	0.021	-0.018	0.022	-0.024	0.022
Certrec	0.111**	0.037	0.107**	0.037	0.126**	0.038	0.106**	0.038
Zclimate			-0.202**	0.013	-0.196**	0.013	-0.183**	0.014
Zcontrol			-0.131**	0.010	-0.117**	0.011	-0.105*	0.011
Zinfluen			0.039**	0.012	0.045**	0.012	0.030*	0.012
Zstudent			0.014	0.012	0.039**	0.013	0.034*	0.013
Totalenroll					-0.00008**	0.00002	-0.00006**	0.00002
PerFRLkw					0.003**	0.001	0.002**	0.001
NewminPER					0.002**	0.000	0.0014**	0.0005
Stutearatio					0.012**	0.003	0.011**	0.003
Suburban					0.078**	0.021	0.072**	0.021
PDdiscipline							0.015	0.021
PDindepth							0.118**	0.021
PDmethodte							0.072**	0.023
PDassessme							0.114**	0.022
PDstandards							0.229**	0.023
Strongacc	0.230**	0.021	0.229**	0.021	0.182**	0.023	0.170**	0.023

(• Dependent Variable: To the extent teachers use state or district standards to guide your instructional practice. • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

When we include only teacher level variables (Model 1) with accountability variable, the coefficient size of the Strongacc variable is 0.230. Adding teacher perception variables (Model 2), however, decreases the coefficient very slightly to 0.229.

Adding school level variables reduces the accountability policy coefficient to 0.182, which means that including school level factors disentangles some magnitude of effect from the effect of accountability variable on the teachers' use of standards.

Finally, adding PD variables into the model reduces the coefficient size of accountability to 0.170. This illustrates that adding other variables such as teacher characteristic variables, school level variables and PD variables reduce some explanatory power of accountability policy in Models 1 and 2.

8.1.2 Increase of Sample Size by Dropping A Few Control Variables

We estimate the effects of teacher, school and accountability variables on teachers' use of state/district standards for instruction using three full models, which are linear regression, ordered probit, and ordered logit models. For comparison, we report three models' estimates in Table 2.8. Table 2.8 indicates that the signs and significance of most independent variables are same across the three models except for a PD variable, PDmethodte and certification variable. In the linear regression model, PDmethodte is significant, but becomes insignificant in the ordered probit and ordered logit models. Certification becomes more significant in ordered probit and ordered logit models. All estimates between ordered probit and ordered logit models appear to remain same in the direction and significance.

Table 2.8: Linear Regression, Ordered Probit, and Ordered Logit Estimates of Teacher's Use of State/District Standards for Instruction

Variables	Linear Regression (OLS)		Ordered Probit (MLE)		Ordered Logit (MLE)	
	b	s.e.	b	s.e.	b	s.e.
Age	0.006**	0.002	0.009**	0.002	0.016**	0.004
Salary	-0.000006**	0.000002	-0.000008**	0.000002	-0.000012**	0.000003
Remaassi	0.193**	0.052	0.252**	0.071	0.454**	0.120
HRSEng	-0.002	0.002	-0.003	0.002	-0.006	0.004
HRSMATH	0.011**	0.003	0.013**	0.004	0.023**	0.007
Mathalba	-0.304*	0.149	-0.377*	0.179	-0.726*	0.323
English	0.032	0.060	0.029	0.085	0.040	0.147
Male	-0.228**	0.039	-0.277**	0.045	-0.478**	0.078
Minority	0.029	0.036	0.028	0.048	0.056	0.083
Totexper	0.004	0.005	0.005	0.007	0.006	0.012
Sqtotexp	-0.00004	0.00014	-0.00006	0.00019	-0.00004	0.00032
Unionmem	-0.035	0.028	-0.048	0.038	-0.093	0.064
MA	-0.001	0.026	0.006	0.034	0.020	0.058
Verycomp	0.012	0.028	0.003	0.036	-0.001	0.062
Certrec	0.127*	0.050	0.166**	0.060	0.301**	0.105
Zclimate	-0.113**	0.015	-0.157**	0.020	-0.266**	0.035
Zcontrol	-0.1028**	0.016	-0.113**	0.019	-0.172**	0.033
Zinfluen	0.060**	0.015	0.069**	0.019	0.104**	0.033
Zstudent	0.059**	0.016	0.070**	0.021	0.113**	0.036
Totalenroll	-0.00006	0.00003	-0.00007	0.00004	-0.0001	0.0001
PerFRLkw	0.002**	0.001	0.003**	0.001	0.005**	0.001
NewminPER	0.001	0.001	0.001	0.001	0.001	0.001
Stutearatio	0.005*	0.002	0.009*	0.004	0.015*	0.007
Suburban	0.067**	0.025	0.083*	0.033	0.132*	0.056
PDdiscipline	0.014	0.025	0.039	0.032	0.079	0.055
PDindepth	0.115**	0.027	0.139**	0.034	0.225**	0.058
PDmethodte	0.065*	0.031	0.071	0.038	0.119	0.066
PDassessme	0.087**	0.029	0.112**	0.035	0.194**	0.061
PDstandards	0.233**	0.034	0.257**	0.040	0.425**	0.069
Strongacc	0.097**	0.027	0.130**	0.037	0.220**	0.062
cons	3.400**	0.101				

(• Dependent Variable: Useofstandard, to the extent teachers use state or district standards to guide your instructional practice. • Pweight: tfnlwgt. Number of observation: 12,833, Strata: state. Number of strata: 51, Population size = 1,259,363 • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

Among the PD variables, PDstandards has the largest effect on the use of standards as expected and discipline relevant PD program has the least effect. Teachers who have joined PD for learning performance or content standards are more likely to use state/district standards to guide their instruction than teachers who have joined other PD, holding other variables constant. The PDmethodte variable is significant at the 0.05 level in the linear regression model, however the significance disappears in the ordered probit and ordered logit model. Teachers from the states with strong accountability policy are more likely to use standards to guide their practice than teachers from the states with weak accountability policy, keeping other variables fixed. Thus, it seems that accountability policy has a positive effect on the teachers' use of standards as well as PD programs.

However, the coefficient size of the variable, PDstandards (0.233) is twice larger than the variable, Strongacc (0.097). That is, holding other variables including Strongacc constant, teachers who joined PD to learn standards are more likely to use state standards by 0.233 on the scale score than those who did not join on average (from the linear regression model). Holding other variables constant, teachers from the states implementing strong accountability are more likely to use state standards by 0.097 on the score than those working at the states with weak accountability policy, which is relatively smaller effect than the PD program for standards. This indicates that opportunity to learn the state standards by joining PD will have more effect on the change of teachers' practice than accountability policy.

To check whether the difference of the coefficient size on these two variables is statistically significant, the Wald test was executed. The null hypothesis for this test is

$B_{PDstandards} = B_{Strongacc}$ against the alternative hypothesis, $B_{PDstandards} > B_{Strongacc}$. The test result ($F(1, 12782) = 5.5$, $Prob > F = 0.0191$ (P-value = 0.0096 for one tail test, thus we reject the null hypothesis) shows that the effect size is significantly different at the 0.01 level. That is, the effect of PDstandards is significantly larger than that of Strongacc: opportunity to learn will make teachers use standards to guide their instruction more effectively than accountability policy. However, the population size that the models in Table 2.8 represent is only 1,259,363.

Now, we exclude the HRSmath, and HRSread variables to increase sample size. Also age variable was dropped in order to see whether there is an effect of experience since it could cause multicollinearity with experience variable. Table 2.9 exhibits the results after dropping the three independent variables. The sample size increases from 12,843 to 34,109 and the represented population size from 1,259,363 to 2,440,181.

Generally the significance and sign of the coefficients in Table 2.9 do not change except for few variables. After we drop three variables (age, hours of teaching in math and reading), the experience variable becomes significant. This could mean that there is multicollinearity between age and experience. The experience has decreasing marginal effect on the use of standards as we can see; coefficient of the square of experience is negative. One perception variable, influence becomes insignificant in the ordered logit model. The PD variable, PDmethode, and school size variable, Totalenroll, become significant in this model.

Table 2.9: Linear Regression, Ordered Probit, and Ordered Logit Estimates of Teacher's Use of State/District Standards for Instruction after Dropping Three Variables, Age, HRSmath, and HRSread.

Models	Linear Regression (OLS)		Oprobit (MLE)		Ordered Logit (MLE)	
Variables	b	s.e.	b	s.e.	b	s.e.
Salary	-0.000005**	0.000001	-0.000005**	0.000001	-0.000009**	0.000002
Remaassi	0.166**	0.023	0.186**	0.029	0.297**	0.050
Mathalba	-0.090*	0.036	-0.115**	0.042	-0.197**	0.073
English	-0.036	0.031	-0.056	0.039	-0.104	0.067
Male	-0.248**	0.019	-0.289**	0.021	-0.492**	0.036
Minority	0.048	0.025	0.058	0.032	0.103	0.055
Totexper	0.011**	0.003	0.015**	0.004	0.024**	0.007
Sqtotexp	-0.0002	0.0001	-0.0002*	0.0001	-0.0004	0.000
Unionmem	-0.030	0.019	-0.035	0.023	-0.062	0.040
MA	0.026	0.018	0.036	0.022	0.066	0.037
Verycomp	-0.014	0.018	-0.024	0.022	-0.048	0.038
Certrec	0.088**	0.032	0.106**	0.038	0.178**	0.065
Zclimate	-0.146**	0.011	-0.183**	0.014	-0.311**	0.023
Zcontrol	-0.100**	0.010	-0.105*	0.011	-0.163**	0.019
Zinfluen	0.034**	0.010	0.030*	0.012	0.035	0.021
Zstudent	0.030**	0.011	0.034*	0.013	0.049*	0.023
Totalenroll	-0.00005**	0.00001	-0.00006**	0.00002	-0.00011**	0.00003
PerFRLkw	0.0019**	0.0004	0.002**	0.001	0.004**	0.001
NewminPER	0.0011**	0.0004	0.0014**	0.0005	0.002**	0.001
Stutearatio	0.008**	0.002	0.011**	0.003	0.018**	0.004
Suburban	0.061**	0.017	0.072**	0.021	0.120**	0.036
PDdiscipline	0.006	0.017	0.015	0.021	0.032	0.035
PDindepth	0.099**	0.018	0.118**	0.021	0.195**	0.036
PDmethodte	0.068**	0.020	0.072**	0.023	0.123**	0.039
PDassessme	0.099**	0.019	0.114**	0.022	0.194**	0.037
PDstandards	0.213**	0.021	0.229**	0.023	0.384**	0.040
Strongacc	0.139**	0.019	0.170**	0.023	0.287**	0.039
_cons	3.561**	0.058				

(• Dependent Variable: Useofstandard, to the extent teachers use state or district standards to guide your instructional practice. • pweight: tfnlwgt, Number of observation = 34,109, Strata: state, Number of strata = 51, Population size = 2,440,181, • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

Interestingly, still the college major variable, Mathalba, remains as significantly negative. In other words, teachers whose college major is math or math education are

less likely to use standards to guide their practice than those whose major is not math or math education. Teachers working at suburban schools are more likely to use standards than teachers working at other areas, holding other factors constant. Teachers from schools serving more minority or economically disadvantaged students are more likely to use standards guidance, keeping other variables same. Teachers in schools with larger student teacher ratio also more likely to use state/district standards to guide their teaching practice.

Another interesting point is that teachers who report that they have more control in the classroom are less likely to use the state standards. This might mean that teachers' classroom instruction could be external policy proof. It would not be easy for an external policy to penetrate into the classroom. The variable, $Z_{influence}$, shows that teachers who recognize that they have more influence on school policy such as setting performance standards for students, evaluating teachers, hiring new full-time teachers are more likely to accept the state standards as their instructional guidance. In addition, negative sign of $Z_{climate}$ variable also illustrates that teachers working at more supportive school climate will be more likely to use state standards (remember that higher score in $Z_{climate}$ means worse school circumstance). This could imply that stronger internal accountability system will have a positive effect on the teachers' use of state standards.

It seems that the difference of coefficient size between two variables, $PD_{standards}$ and $Strong_{acc}$, becomes smaller. However, still the coefficient size of the PD variable, $PD_{standards}$, is much larger than that of the $strong_{acc}$. And this is statistically significant when we test the null hypothesis, $B_{Strong_{acc}} = B_{PD_{standards}}$ against the

alternative hypothesis, $B_{\text{Strongacc}} < B_{\text{PDstandards}}$ using the Wald test ($F(1,34058) = 3.14$, p-value is 0.0383 at the one tail test, so we reject the null hypothesis at the 0.05 level). Thus, same conclusion, the size of effect of PD of standards is larger than the size of effect of accountability, can be drawn. And the increased sample size and its representing population size approaches to the real population size closely without changing critical difference in the analysis by dropping the three variables, age, and hours of teaching math and reading.

8.1.3 Checking Instrumental Variables Estimation

In the method section, we discussed the self-selection problem that could be caused by teachers' voluntary decision of the PD program participation. Table 2.6 and Table 2.9 exhibit that the coefficient of PD program related to state or district standards decreases significantly as we control more teacher and school level variables. This implies that certainly self-selection problem exists and the problem is alleviated as we control more variables. Although we alleviated (or hopefully solve) self-selection problem by controlling enough independent variables in the equation, it would be worth trying to find a good IV to check whether the IV will provide a different estimation.

Table 2.10: The Partial Effect of Expenditure of Instructional Staff Support on Teachers' Standards-related Professional Development Program Participation in Michigan (Probit Model)

	Model 1			Model 2		
	b	se	P>t	b	se	P>t
<i>Inssup</i>	-0.00000002	0.00000002	0.379	-0.00000003	0.00000002	0.123
Salary	-0.00002	0.00001	0.109	-0.00002	0.00001	0.102
Remaassi	0.187	0.249	0.453	-0.003	0.234	0.988
Mathalba	0.278	0.336	0.409	0.112	0.299	0.708
English	-0.453	0.318	0.155	-0.318	0.296	0.282
Male	0.054	0.161	0.736	-0.183	0.155	0.236
Minority	-0.204	0.292	0.485	0.080	0.277	0.774
Totexper	0.053	0.038	0.161	0.062	0.032	0.055
Sqtotexp	-0.001	0.001	0.376	-0.001	0.001	0.175
Unionmem	0.011	0.661	0.987	-0.254	0.466	0.585
MA	-0.020	0.197	0.92	0.292	0.187	0.119
Verycomp	-0.072	0.169	0.672	-0.076	0.168	0.65
Certrec	-0.791	0.425	0.063	-0.316	0.327	0.334
Zclimate	-0.108	0.123	0.38	-0.191	0.103	0.063
Zcontrol	-0.059	0.096	0.539	-0.047	0.091	0.603
Zinfluen	0.103	0.107	0.334	0.206	0.099	0.037
Zstudent	0.059	0.102	0.564	-0.001	0.099	0.991
Totalenroll	-0.0001	0.0002	0.74	0.000	0.000	0.524
PerFRLkw	0.005	0.005	0.277	0.003	0.004	0.454
NewminPER	0.001	0.006	0.82	0.007	0.006	0.215
Stutearatio	0.019	0.022	0.378	0.024	0.021	0.246
Suburban	0.353	0.203	0.082	0.305	0.180	0.091
PDdiscipline	-0.018	0.168	0.913			
PDindepth	1.055	0.164	0			
PDmethodte	-0.030	0.172	0.863			
PDassessme	0.856	0.165	0			
_cons	0.247	0.971	0.8	0.715	0.783	0.362

* Dependent variable: PDstandards. Note: pweight: tfnlwgt, strata: state, number of strata: 1.
Observation: 556. Population size: 68,734

First candidate as an IV we mentioned in the method section is expenditure on instructional staff support services. Unfortunately, SASS database does not provide expenditure of instructional staff support. Using Michigan K-12 finance database, we obtain the information regarding instructional staff support expenditure and run the

regression equation (2) to see there is correlation between standards-related PD program participation and the expenditure. Table 2.10 provides two probit models which show whether there is a partial effect of expenditures of instructional supports on teachers' standards-related PD program using Michigan portion of SASS samples. Difference between Model 1 and Model 2 is that Model 2 excludes other PD variables, because other PD variables would be endogenous. Inssup is the instrumental variable, expenditures of instructional staff supports. Accountability policy variable is dropped in the model since we are using only Michigan teachers. Analysis using MI samples shows that the expenditure of instructional staff support services does not have any partial effect on teachers' participation on standards-related PD activities. Since the expenditure of instructional staff support includes other categories in addition to PD program support, this result is not arbitrary. Thus, expenditure of instructional staff support does not appear to be an appropriate IV.

Second candidate for an IV is the number of PD programs provided to school or district administrators. We run the probit model and found that the number of PD programs for school or district administrators has a partial effect on teachers' standards-related PD program participation. Table 2.11 displays the partial effect of the variable, the number of PD programs for administrators, PDadminitrator. The coefficient of PDadminitrator is 0.026 and it is significant at the 0.01 level in Model 1. And in Model 2, the coefficient of PD administrator is still significant at the 0.01 level. This is somewhat expected result. We can think that districts which provide more PD opportunities to administrators also provide more PD programs to teachers, thus teachers working in such districts are more likely to participate in standards-related PD programs.

Table 2.11: The Partial Effect of the Number of PD for Administrators on Teachers' Standards-related Professional Development Program Participation (Probit Model)

	Model 1			Model 2		
	b	s.e.	P>t	b	s.e.	P>t
PDadministrator	0.026	0.006	0	0.034	0.005	0
Salary	0.000001	0.000002	0.524	-0.000001	0.000001	0.468
Remaassi	0.050	0.043	0.244	0.046	0.040	0.241
Mathalba	0.120	0.058	0.038	0.019	0.054	0.723
English	-0.014	0.057	0.807	-0.012	0.053	0.821
Male	-0.166	0.028	0	-0.266	0.026	0
Minority	0.020	0.044	0.648	0.057	0.040	0.154
Totexper	0.016	0.005	0.002	0.033	0.005	0
Sqtotexp	-0.0004	0.0001	0.008	-0.001	0.000	0
unionmem	0.072	0.031	0.022	0.128	0.030	0
MA	-0.010	0.030	0.742	0.016	0.028	0.571
Verycomp	0.061	0.032	0.054	0.065	0.029	0.024
Certrec	0.159	0.053	0.002	0.174	0.051	0.001
Zclimate	-0.037	0.018	0.046	-0.081	0.017	0
Zcontrol	-0.077	0.014	0	-0.091	0.014	0
Zinfluen	0.035	0.016	0.026	0.080	0.015	0
Zstudent	0.038	0.019	0.043	0.052	0.017	0.003
Totalenroll	-0.00005	0.00002	0.031	0.000	0.000	0
PerFRLkw	0.001	0.001	0.145	0.002	0.001	0.004
NewminPER	0.0003	0.0007	0.651	0.002	0.001	0.002
Stutearatio	0.002	0.002	0.31	0.005	0.003	0.132
Suburban	-0.013	0.029	0.66	0.016	0.027	0.549
PDdiscipline	0.051	0.029	0.083			
PDindepth	0.934	0.027	0			
PDmethodte	0.220	0.030	0			
PDassessme	0.579	0.028	0			
Strongacc	-0.017	0.032	0.59	0.064	0.030	0.032
_cons	-0.952	0.094	0	-0.241	0.090	0.007

* Note: Dependent variable: PDstandards. Pweight: tfnlwgt, strata: state, number of strata: 51.
Observation: 30,272 . Population size: 2,160,346. F(27,30,195)=83.33, Prob>F=0.0000

Thus, we run IV regression using the number of PD opportunities for administrators as an IV. For convenience, we run the usual IV regression assuming that the dependent variable, use of standards, has a numeric meaning, although it is ordered response. Following table 2.12 is the result of IV estimation.

Table 2.12: IV Estimation for the Effect on Teachers' Use of State/District Standards

	<u>Model 1</u>			<u>Model 2</u>		
	b	s.e.	P>t	b	s.e.	P>t
Salary	-0.000004**	0.000001	0.001	-0.000004**	0.000001	0.001
Remaassi	0.146**	0.032	0	0.150**	0.029	0
Mathalba	-0.138**	0.053	0.009	-0.101*	0.044	0.021
English	-0.023	0.043	0.597	-0.025	0.039	0.529
Male	-0.166**	0.042	0	-0.157**	0.043	0
Minority	0.047	0.035	0.17	0.039	0.032	0.227
Totexper	0.006	0.005	0.272	0.002	0.006	0.688
Sqtotexp	-0.0001	0.0001	0.649	0.00002	0.00015	0.89
unionmem	-0.063*	0.029	0.028	-0.070*	0.028	0.013
MA	0.023	0.023	0.334	0.018	0.022	0.419
Verycomp	-0.042	0.027	0.118	-0.039	0.024	0.109
Certrec	0.003	0.053	0.952	0.013	0.046	0.777
Zclimate	-0.130**	0.015	0	-0.122**	0.016	0
Zcontrol	-0.052**	0.020	0.008	-0.056**	0.017	0.001
Zinfluen	0.012	0.015	0.409	0.003	0.016	0.837
Zstudent	0.014	0.016	0.38	0.012	0.015	0.423
Totalenroll	-0.00003	0.00002	0.085	-0.00003	0.00002	0.129
PerFRLkw	0.002**	0.001	0.006	0.001*	0.001	0.01
NewminPER	0.001	0.001	0.145	0.0001	0.001	0.521
Stutearatio	0.006**	0.002	0.006	0.005*	0.002	0.01
Suburban	0.076**	0.023	0.001	0.066**	0.021	0.002
PDdiscipline	0.005	0.024	0.829			
PDindepth	-0.400*	0.196	0.041			
PDmethode	-0.057	0.054	0.286			
PDassessme	-0.217*	0.122	0.074			
PDstandards	1.982**	0.680	0.004	1.508**	0.399	0
Strongacc	0.150**	0.025	0	0.129**	0.024	0
_cons	3.033**	0.209	0	3.000**	0.217	0

(• Dependent Variable: Useofstandard, To the extent teachers use state or district standards to guide your instructional practice. Instrumented: PDstandards. IV: number of PD programs offered to administrators by district. • pweight: tfnlwgt, Number of observation = 30,272. Strata: state, Number of strata = 51, Population size = 2,160,347 • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

Table 2.12 exhibits that the coefficient of the instrumented variable, PDstandards, becomes very large and its standard error also increases compared to the coefficient of PDstandard (0.233) of the OLS in Table 2.9. Coefficients of other independent variables

have the same signs and do not change much. However, some independent variables such as Z_{influen} and Z_{student} become insignificant because their coefficients decrease. Other four PD participation variables in Model 1 become insignificant and even the signs of coefficients changed from positive to negative. Model 2 in Table 2.12 excludes other PD participation variables in the equation since we are interested in the effect of PD participation in standards on teachers' use standards and other four PD variables are endogenous. Dropping four other PD variables reduces the coefficient and standard error at the same time, however, the standards-related PD program participation is still significant. Large standard error is expected since IV estimation has a tendency to make the standard error large. The effect of standards-related PD program has more significant effect on teachers' instructional practice (The Wald test, $H_0: B_{\text{PDstandard}} = B_{\text{Strongacc}}$, $H_1: B_{\text{PDstandard}} > B_{\text{Strongacc}}$. $F(1,30221) = 7.29$, and $P\text{-value} = 0.0034$), and we confirm that the conceptual framework of opportunities to learn works better than principal-agent model.

In sum, we found that both standards-related PD program participation and strong accountability policy spur teachers' use of state or district standards in their teaching practice. However, it appears that standards-related PD program participation has more effect on teaching practice, opportunities to learn work better than principal-agent model. Possible self-selection problem and the quality of PD program were concerns in the analysis. Self-selection problem is somewhat alleviated by including enough control variables, however IV estimation was used to solve the problem. IV estimation further confirms that standards-related PD program participation has larger effect on teachers' instructional change than accountability policy.

It also needs to be noted that the quality of PD program is weakly moderate, so if

teachers are provided with more effective PD programs, then the effect size of PD program participation would increase. Thus, we conclude that theory of opportunities to learn would be more effective in changing the teachers' instruction than accountability policy framed by principal-agent model.

8. 2. The Effect on the Teacher's Use of Information from State or Local Achievement Tests

8.2.1 Teachers' Use of Student Test Score to Strengthen Their Content Knowledge and Teaching Practice²⁴

First, we examine the effects of PD variables and accountability variables on the teachers' use of the student test information to strengthen their content knowledge or teaching practice. We expect that the PD on in-depth study may have a large effect on teacher's practice of using test score to strengthen their content knowledge, thus Table 2.13 presents simple models by including in-depth study PD program participation and accountability policy. Model 1 in Table 2.13 only includes two independent variables we want to compare directly. Participation in in-depth study PD program has more effect than accountability policy. However, as we include more control variables, their effect size became similar.

²⁴ Using student test score to check what areas the teacher needs to improve their subject knowledge and teaching practice could be regarded as a pedagogical aspect of the test-driven accountability policy, if we use the framework of Cohen's pedagogy of policy.

Table 2.13: Four Ordered Probit Models of the Effect on Teachers' Use of Test to Strengthen their Content Knowledge and Teaching Practice

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>		<u>Model 4</u>	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Salary					-0.000006**	0.000001	-0.000003*	0.000002
Remaassi					0.083*	0.033	0.106**	0.034
Mathalba					-0.247**	0.050	-0.204**	0.053
English					-0.069	0.045	-0.077	0.047
Male					-0.314**	0.026	-0.296**	0.027
Minority					0.185**	0.034	0.171**	0.041
Totexper					0.011*	0.004	0.010*	0.005
Sqtotexp					-0.0002	0.0001	0.000	0.000
Unionmem					-0.036	0.028	-0.036	0.029
MA					-0.029	0.026	-0.043	0.027
Verycomp					-0.099**	0.027	-0.094**	0.028
Certrec					0.053	0.046	0.066	0.049
Totalenroll							0.000**	0.000
PerFRLkw							0.001	0.001
NewminPER							0.000	0.001
Stutearatio							0.003	0.003
Suburban							-0.019	0.026
PDdiscipline			0.130**	0.024	0.109**	0.024	0.106**	0.025
PDmethodte			0.065*	0.027	0.053	0.028	0.059	0.029
PDassessme			0.130**	0.026	0.116**	0.026	0.118**	0.028
PDstandard			0.115**	0.029	0.107**	0.029	0.094**	0.031
PDindepth	0.251**	0.023	0.169**	0.025	0.149**	0.025	0.139**	0.026
Strongacc	0.102**	0.025	0.096**	0.025	0.077**	0.026	0.120**	0.028

Note: • Dependent Variable: Useforarea, to the extent teachers use test results to strengthen their content knowledge and teaching practice. • Pweight: tfnlwgt. Strata: state. Number of strata: 51. For model 1, observation number is 22,115. Represented population size is 1,722,596. For model 2, observation number is 22,115. Population size is 1,722,596. For model 3, observation number is 21,990. Population size is 1,715,663. For model 4, observation number is 19,873. Population size is 1,549,867.

• b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

Results from linear regression, ordered probit, and ordered logit models which exclude three variables, age and hours of teaching math and reading from the full model are provided in Table 2.14 because the result can represent more population. Table 2.14 displays that all PD variables except for PDmethodte have significant and positive effects on teachers' use of student test score to assess the areas where teachers need to

strengthen their content knowledge and instructional practice. Zinfluence variable also has significant and positive effect. However, Zcontrol variable is not significant any more in the ordered probit and ordered logit model.

Table 2.14: Effect on the Teachers' Use of Information from State or Local Test Scores to Strengthen Subject Area and Practice.

	<u>Linear Regression</u>		<u>Ordered Probit</u>		<u>Ordered Logit</u>	
	b	s.e.	b	s.e.	b	s.e.
Salary	-0.000004*	0.000002	-0.000004*	0.000002	-0.000007*	0.000003
Remaassi	0.131**	0.036	0.117**	0.035	0.196**	0.059
Mathalba	-0.220**	0.060	-0.199**	0.053	-0.324**	0.092
English	-0.058	0.052	-0.062	0.048	-0.109	0.082
Male	-0.307**	0.031	-0.286**	0.028	-0.496**	0.048
Minority	0.151**	0.042	0.157**	0.041	0.267**	0.069
Totexper	0.011*	0.005	0.012*	0.005	0.023**	0.008
Sqtotexp	-0.0003	0.0001	-0.0003**	0.0001	-0.0005*	0.0002
Unionmem	-0.025	0.030	-0.026	0.029	-0.045	0.050
MA	-0.046	0.029	-0.038	0.027	-0.060	0.047
Verycomp	-0.092**	0.030	-0.088**	0.028	-0.160**	0.048
Certrec	0.056	0.053	0.062	0.050	0.104	0.084
Zclimate	-0.062**	0.018	-0.061**	0.017	-0.109**	0.029
Zcontrol	-0.036*	0.016	-0.015	0.015	-0.013	0.027
Zinfluen	0.064**	0.016	0.055**	0.015	0.091**	0.026
Zstudent	0.028	0.019	0.024	0.018	0.037	0.030
Totalenroll	-0.00017**	0.00003	-0.00016**	0.00002	-0.00027**	0.00004
PerFRLkw	0.002**	0.001	0.002**	0.001	0.003**	0.001
NewminPER	0.0004	0.0006	0.001	0.001	0.001	0.001
Stutearatio	0.002	0.002	0.002	0.003	0.005	0.005
Suburban	-0.033	0.028	-0.020	0.027	-0.020	0.045
PDdiscipline	0.103**	0.027	0.097**	0.026	0.159**	0.044
PDmethodte	0.056	0.032	0.049	0.029	0.082	0.050
PDassessme	0.117**	0.030	0.108**	0.028	0.180**	0.047
PDstandards	0.092**	0.034	0.086**	0.031	0.155**	0.053
PDindepth	0.131**	0.029	0.123**	0.027	0.210**	0.046
Strongacc	0.120**	0.029	0.113**	0.028	0.189**	0.048
_cons	3.419**	0.088				

(• Dependent Variable: Useforarea, to the extent teachers use state or local tests to assess areas where they need to strengthen their content knowledge or teaching practice • pweight:tfnlwtg, Number of obs=19,873. Strata: state. Number of strata=51. Population size= 1,549,868 • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

Interestingly, college major in math or math education and selectivity of college (Verycomp) has significant negative effect. Teachers who graduated from the college which is the most highly or very competitive institution are less likely to use students' test score to assess areas whether they may want to strengthen content knowledge or practice. Male teachers are less likely to use student test score to find areas where they have to enhance content knowledge or teaching practice than female teachers.

Among the PD variables, PDindepth which means teachers joined PD regarding in-depth study of the content in his or her main teaching assignment field appears to have the largest positive effect on the teachers' use of student test scores to assess their content knowledge and instructional practice. In addition, the PDindepth variable appears to have larger impact size than the strong accountability variable. The Wald test, $H_0: B_{PDindepth} = B_{Strongacc}$, $H_1: B_{PDindepth} > B_{Strongacc}$, was executed to see whether the impact size of PDindepth is significantly larger than the impact size of Strongacc. Test result shows that the difference is insignificant ($F(1, 19822) = 0.06$, $\text{Prob} > F = 0.407$ at the one tail test in ordered probit model). Also when we check whether the difference between the coefficient size of other PD participation variables and that of accountability dummy variable is significant, no significant difference can be found.

Thus, while can see that the accountability policy have a certain positive effect on the teachers' use of test scores to assess their content knowledge and practice, the effect is not significantly different from those of PD variables. And we found that the PD program participation to do in-depth study for their main teaching field has a little bit larger positive effect on the teachers' use of test scores, although the difference is not significant.

8. 2. 2. Teachers' Use of Student Test Scores to Adjust Their Curriculum in Areas Where Their Students Encountered Problems.

Table 2.15: Four Ordered Probit Models of the Effect on Teachers' Use of Test to Adjust Their Curriculum in Areas Where Their Student Encountered Problems

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>		<u>Model 4</u>	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Salary					-0.000005**	0.000001	-0.000003	0.000002
Remaassi					0.142**	0.033	0.161**	0.034
Mathalba					-0.248**	0.049	-0.190**	0.052
English					-0.009	0.046	0.004	0.048
Male					-0.312**	0.026	-0.290**	0.028
Minority					0.124**	0.034	0.096*	0.041
Totexper					0.012*	0.005	0.013*	0.005
Sqtotexp					0.000	0.000	-0.0002	0.0001
Unionmem					-0.050	0.029	-0.047	0.030
MA					-0.010	0.026	-0.008	0.027
Verycomp					-0.083**	0.027	-0.074**	0.028
Certrec					0.077	0.048	0.080	0.051
Totalenroll							-0.00016**	0.00002
PerFRLkw							0.002**	0.001
NewminPER							0.0003	0.001
Stutearatio							0.003	0.003
Suburban							0.010	0.027
PDdiscipline			0.098**	0.024	0.091**	0.025	0.087**	0.026
PDindepth			0.162**	0.026	0.145**	0.026	0.131**	0.027
PDmethodte			0.073*	0.028	0.066*	0.029	0.078*	0.030
PDstandard			0.134**	0.030	0.123**	0.030	0.120**	0.031
PDassessme	0.255**	0.025	0.163**	0.027	0.144**	0.027	0.146**	0.028
Strongacc	0.153**	0.025	0.134**	0.025	0.127**	0.026	0.156**	0.029

Note: • Dependent Variable: UseTcurri, to the extent teachers use test results to adjust curriculum. • Pweight: tfnlwgt. Strata: state. Number of strata: 51. For model 1, observation number is 22,115, Population size is 1,722,596. For model 2, observation number is 22,115, Population size is 1,722,596. For model 3, observation number is 21,990, Population size is 1,715,663. For model 4, observation number is 19,873, Population size is 1,549,867. • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

Using student test scores to adjust curriculum in areas where students encountered problem would certainly help students progress in learning²⁵. Table 2.15 provides four ordered probit models. Since we expect that PD program related to assessment could have a larger effect on the use of test to adjust curriculum, we start to compare the assessment-related PD program participation with the accountability policy variable. In Model 1 of Table 2.15, we found that the assessment-related PD program participation has larger effect than the accountability policy. However, in Model 4, coefficient of accountability policy variable becomes larger than that of assessment-related PD program participation.

Table 2.16 shows similar result to the Table 2.15 and accountability policy variable appears to have slightly large effect than the PD program participation variable, PDassessme. The Wald test does not provides any evidence that strong accountability policy variable has larger effect than the effects of other three PD program participation variables, PDassessme, PDstandards, and PDindepth.

²⁵ However, we may need to interpret this item cautiously, since this practice can be reduced to teaching test.

Table 2.16: Effect on the Teachers' Use of Information from State or Local Test Scores to Adjust Their Curriculum in Areas Where Their Student Encountered Problems.

	Linear Regression		Ordered Probit		Ordered Logit	
	b	s.e.	b	s.e.	b	s.e.
Salary	-0.000004*	0.000002	-0.000003*	0.000002	-0.000005*	0.000003
Remaassi	0.179**	0.033	0.169**	0.034	0.274**	0.058
Mathalba	-0.193**	0.054	-0.189**	0.052	-0.323**	0.090
English	0.022	0.048	0.016	0.049	0.021	0.083
Male	-0.293**	0.031	-0.283**	0.029	-0.487**	0.049
Minority	0.071	0.041	0.081	0.042	0.129	0.071
Totexper	0.013**	0.005	0.015**	0.005	0.026**	0.008
Sqtotexp	-0.0002	0.0001	-0.0003	0.0001	-0.0005*	0.0002
Unionmem	-0.027	0.029	-0.036	0.030	-0.061	0.052
MA	-0.007	0.028	-0.002	0.028	-0.003	0.047
Verycomp	-0.063*	0.029	-0.067*	0.028	-0.129**	0.048
Certrec	0.071	0.052	0.078	0.051	0.145	0.089
Zclimate	-0.074**	0.018	-0.079**	0.018	-0.144**	0.030
Zcontrol	-0.030	0.016	-0.007	0.016	0.006	0.028
Zinfluen	0.054**	0.015	0.046**	0.015	0.073**	0.026
Zstudents	0.007	0.018	0.005	0.017	0.002	0.030
Totalenroll	-0.00014**	0.00002	-0.00014**	0.00002	-0.00025**	0.00004
PerFRLkw	0.002**	0.001	0.002**	0.001	0.004**	0.001
NewminPER	0.001	0.001	0.001	0.001	0.001	0.001
Stutearatio	0.002	0.002	0.003	0.003	0.007	0.005
Suburban	0.008	0.027	0.012	0.027	0.026	0.046
PDdiscipline	0.066*	0.026	0.075**	0.026	0.118**	0.045
PDindepth	0.118**	0.028	0.115**	0.027	0.192**	0.046
PDmethodte	0.074*	0.032	0.069*	0.030	0.126*	0.052
PDstandards	0.121**	0.033	0.114**	0.031	0.193**	0.053
<i>PDassessme</i>	0.143**	0.030	0.136**	0.028	0.230**	0.049
<i>Strongacc</i>	0.142**	0.028	0.150**	0.029	0.261**	0.048
_cons	3.362**	0.084				

(• Dependent Variable: UseTcurri, to the extent teachers use state or local tests to adjust curriculum • pweight:tfnlwgt, Number of obs=19,873. Population size= 1,549,868. Strata: state, Number of strata=51, • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

8.2.3 Teachers' Use of Student Test Scores to Group Students into Different Instructional Groups by Achievement or Ability.

Table 2.17: Three Ordered Probit Models on the Effect on Teachers' Use of Test to Group Students by Achievement or Ability

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>	
	b	s.e.	b	s.e.	b	s.e.
Salary			0.0000001	0.0000013	0.00000002	0.00000145
Remaassi			0.037	0.035	0.063	0.037
Mathalba			-0.289**	0.054	-0.296**	0.056
English			-0.179**	0.046	-0.192**	0.048
Male			-0.153**	0.027	-0.137**	0.028
Minority			0.253**	0.034	0.139**	0.039
Totexper			0.001	0.004	0.003	0.005
Sqtotexp			0.000	0.000	0.000	0.000
unionmem			-0.036	0.029	-0.027	0.030
MA			0.003	0.026	0.001	0.028
Verycomp			-0.046	0.028	-0.040	0.029
Certrec			-0.109*	0.049	-0.089	0.051
Totalenroll					0.00007**	0.00001
PerFRLkw					0.002**	0.001
NewminPER					0.002**	0.001
Stuteratio					0.001	0.002
Suburban					-0.015	0.027
PDdiscipline	0.163**	0.024	0.148**	0.025	0.128**	0.026
PDindepth	0.127**	0.026	0.111**	0.026	0.116**	0.027
PDmethodte	0.055	0.029	0.055	0.029	0.057	0.030
PDassessme	0.148**	0.027	0.127**	0.027	0.117**	0.029
PDstandards	0.013	0.030	0.008	0.030	-0.013	0.032
Strongacc	0.183**	0.025	0.156**	0.026	0.146**	0.029

Note: • Dependent Variable: Usegrouping, to the extent teachers use test results to group student by achievement or ability. • Pweight: tfnlwgt. Strata: state. Number of strata: 51. For model 1, observation number is 22115, population size is 1,722,596. For model 2, observation number is 21,990, population size is 1,715,663. For model 3, observation number is 19,873, population size is 1,549,867. • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

Finally, we examine the effects on the teachers' use of student test scores to group students into different instructional groups by achievement or ability within classroom.

Table 2.17 provides three probit models to see the effects of PD program participation

variables and accountability policy variable. Since we are not sure which PD program participation variable will have more effect on the grouping practice, we begin to compare five PD program participation variables with accountability variable in Model 1.

Interestingly, Table 2.17 exhibits that as we control more variables, the coefficient of standards-related PD program participation variable reduces and even the sign becomes negative. The discipline-related PD program participation variable has largest coefficient among PD program participation variables. Teachers who joined discipline-related PD program last twelve months before the survey are more likely to use test results to group students by achievement or ability. Accountability policy variable has larger effect than other PD program participation variables.

Now we provide three full models. The result of Table 2.18 is similar to that of model 3 of Table 2.17. Teachers teaching at the states with strong accountability policy are more likely to use this grouping practice more than other states' teachers. This means that accountability policy rather works in an opposition direction from the original intention. Since teachers feel some pressure to make students receive higher scores in state test under accountability policy, they might be tempted to group students by achievement or ability level. The professional variable, PDstandards, has a negative effect, although it is not significant. That is, teachers who joined PD regarding standards are less likely to use test result to group student by achievement or ability level. This is expectable since the state standards emphasize that all student must achieve high level of learning and discourage teachers to group students by ability.

Table 2.18: Effect on the Teachers' Use of Information from State or Local Test Scores to Group Students into Different Instructional Groups by Achievement or Ability.

	Linear Regression		Ordered Probit		Ordered Logit	
	b	s.e.	B	s.e.	b	s.e.
Salary	-0.000001	0.000002	-0.0000004	0.0000015	-0.000001	0.000002
Remaassi	0.087	0.045	0.067	0.037	0.123*	0.062
Mathalba	-0.350**	0.065	-0.300**	0.056	-0.508**	0.093
English	-0.222**	0.057	-0.183**	0.048	-0.298**	0.080
Male	-0.163**	0.034	-0.134**	0.029	-0.226**	0.048
Minority	0.173**	0.049	0.132**	0.040	0.231**	0.068
Totexper	0.006	0.006	0.005	0.005	0.007	0.008
Sqtotexp	0.00004	0.00016	0.00004	0.00013	0.0001	0.0002
Unionmem	-0.021	0.038	-0.022	0.031	-0.028	0.052
MA	0.009	0.034	0.007	0.028	0.002	0.047
Verycomp	-0.043	0.036	-0.036	0.029	-0.065	0.049
Certrec	-0.109	0.063	-0.087	0.051	-0.125	0.086
Zclimate	-0.050*	0.022	-0.038*	0.018	-0.062*	0.031
Zcontrol	-0.047*	0.019	-0.039*	0.016	-0.076**	0.026
Zinfluen	0.096**	0.020	0.080**	0.016	0.147**	0.028
Zstudent	-0.011	0.021	-0.008	0.018	-0.016	0.031
Totalenroll	-0.00007*	0.00003	-0.00006**	0.00002	-0.00011*	0.00004
PerFRLkw	0.002**	0.001	0.002**	0.001	0.003**	0.001
NewminPER	0.003**	0.001	0.003**	0.001	0.004**	0.001
Stutearatio	0.001	0.003	0.001	0.002	0.000	0.006
Suburban	-0.019	0.033	-0.013	0.027	-0.029	0.046
PDdiscipline	0.141**	0.032	0.117**	0.026	0.188**	0.044
PDindepth	0.127**	0.033	0.104**	0.027	0.169**	0.045
PDmethode	0.052	0.037	0.047	0.030	0.081	0.051
PDassessme	0.129**	0.035	0.107**	0.029	0.189**	0.048
PDstandards	-0.024	0.038	-0.020	0.032	-0.039	0.053
Strongacc	0.174**	0.035	0.143**	0.029	0.241**	0.048
_cons	2.168	0.103				

(• Dependent Variable: : Usegrouping, to the extent teachers use test results to group student by achievement or ability • pweight:tfnlwgt, Number of obs=19,873. Strata: state, Number of strata=51, Population size= 1,549,868. • b: coefficient. s.e: standard error. • ** means the coefficient is significant at the 0.01 level. * means the coefficient is significant at the 0.05 level.)

9. Conclusion

The survey data do not illuminate all submarine issues and practices of teacher behavior. We should be careful in interpreting the results presented so far. There is the self-selection problem and the qualities or usefulness of PD programs are weakly moderate. Despite such limitations, we can find some consistent patterns on the effect of PD and accountability policy from the analysis. We can summarize the findings as following:

- Both accountability policy and opportunity to learn (PD programs) have positive effects on the teachers' use of state/district standards to guide their instructional practice and on the teachers' use of student test information to assess areas to strengthen their content knowledge and adjust curriculum to help students.
- The effect of accountability policy does not overwhelm that of PD programs or opportunities to learn: Opportunity to learn appears to have larger effect on the teachers' use of standards than accountability policy. Both opportunity to learn and accountability policy have similar size of positive effects on the teachers' use of student test information to assess areas to strengthen their content knowledge and adjust curriculum to help students.
- States' academic standard guidance discourages teachers to group students by ability or achievement level within classroom, however, teachers working under the strong accountability policy are more likely to use grouping

practice by achievement or ability level than those working at the state with weak or no accountability policy. Teachers' opportunity to learn the content and performance standards has a negative effect on this grouping practice although the effect is not significant. Thus, accountability policy boosts grouping practice within classroom which is discouraged by state standard.

Therefore, this study provides some moderate evidence that priority in education policy may need to be given to the teachers' opportunity to learn rather than accountability policy. We must note that we do not have enough resource to invest in all policy options. Specifically, nowadays the U.S. is facing budget constraints for education, so more effective and wise use of education fund is needed. Rather than using funds as incentive rewards or sanctions to hold schools/teachers accountable for student achievement, investing education funds to support teachers' opportunity to learn standards might work better in order to help teachers improve their teaching practice. This can be done in various ways. For example, teachers may need to receive financial support when they join PD programs.

APPENDIX

1. Accountability Index, by State, 1999-2000

State	Grades with state testing in 1999-2000	School accountability 1999-2000	Repercussion for schools 1999-2000	Strength of repercussion for schools 1999-2000	HS exit test in 2000	Grade HS test first given	First grad class	Index
Alabama	3-11	School report cards	Ratings, intervention	Strong	Yes	10	2001	4
Alaska	4-7	None	None	None	Yes	10	2002	1
Arizona	3,5,8,10	Report cards	Public Shame	Weak	Yes	10	2002	2
Arkansas	4,6	None	None	None	No			1
California	2-11	Report cards	Ratings, awards, intervention	Strong	No	10	2004	4
Colorado	3, literacy	None	None	None	No			1
Connecticut	4,6,8,10	Reporting scores to state	Identify schools with needs	Weak	No			1
Delaware	3,5,8,10,11	None	None	None	No	10	2004	1
Florida	4,5,8,10	Report cards	Ratings, subject to vouchers	Strong	Yes	10	1988	5
Georgia	3,4,5,8,11	School reports	None	None	Yes	11	1995	2
Hawaii	3,5,8,10	None	None	None	No			1
Idaho	ITBS,3-8	None	None	None	No			1
Illinois	3,4,5,8,10	Academic improvement	Watch lists, warnings, intervention	Moderate	No			2.5
Indiana	3,6,8,10	Performance assessment	Accreditation	Moderate	Yes	10	1999	3
Iowa	None	None	None	None	No			0
Kansas	3,4,5,8,10	School Reports	Accreditation	Weak	No			1
Kentucky	4,5,7,8,10-12	Meeting state improvement goals	Monetary rewards, intervention	Strong	No			4
Louisiana	LEAP,4,8,	Report cards, growth targets	Intervention	Moderate	Yes	10	1991	3
Maine	4,8,11	None	None	None	No			1
Maryland	3,5,8	School performance index	Monetary rewards, reconstitution	Strong	Yes	10,11,12	2001	4
Massachusetts	4,8,10	Students only	Student promotions	Implicitly only	Yes	10	2003	2

Michigan	4,5,7,8	School rating	Accreditation	Weak	No			1
Minnesota	3,5,8,10	School reports	None	None	Yes	8,10		2
Mississippi	2-8	Only districts accountable, based on test scores	Public recognition, loss of accreditation	Moderate to strong at district level	Yes	11	1994	3
Missouri	3-11	School can be deemed academically deficient	Possible audit	Weak	No			1.5
Montana	4,8,11	None	None	None	No			1
Nebraska	None	None	None	None	No			0
Nevada	4,8,10	School reports	None	Weak	Yes	11	1999	1.5
New Hampshire	3,6,10	None	None	None	No			1
New Jersey	4,5,11	Mostly district level, 75% pass rate	Audits, possible state takeover	Strong	Yes	11		5
New Mexico	1-9	School ratings and district rankings	Some money rewards, probation	Moderate to strong	Yes	10	1990	4
New York	4,5,8,11	State review of school performance	Freeze on pupil registration	Strong	Yes	10	1998	5
North Carolina	3-8	School ratings	Money rewards, intervention	Strong	Yes	9	1994	5
North Dakota	4,8,12	Improve student learning	Accreditation	Weak	No			1
Ohio	4,6,9,12	Report cards, but mainly district level	Money for schools, sanctions for districts	Moderate	Yes	9		3
Oklahoma	5,8	Reports to state	Accreditation	Weak	No			1
Oregon	3,5,8,10	School performance ratings	Write school improvement plans	Weak to moderate	Yes	10	1991	2.5
Pennsylvania	5,6,8,9,11	High schools have ratings	Money for HS improvement	Weak	No			1
Rhode Island	3,4,7,8,10	Yearly progress on test results	Reconstitution	Weak implementation	No			1
South Carolina	3-8,10	District only	District defined as impaired	Moderate	Yes	10	1990	3
South Dakota	2,4,5,8,9,11	Test reports	None	None	No			1

Tennessee	3-8,9	Test reports	Accreditation	Weak	Yes	9		1.5
Texas	3-9,10	Reports cards	School ratings, interventions	Strong	Yes	10	1991	5
Utah	3,5,8,11	None	Accreditation	Weak	No	10	2007	1
Vermont	2,4,8,10	School reports	Identify schools for assistance	Weak	No			1
Virginia	3,4,5,6,8,9	Report tests, other data	Standards of Accreditation	Weak to moderate	No			2
Washington	2-10	School reports	Accreditation	Weak	No	10	2008	1
West Virginia	3-8	Performance audits	Intervention	Strong	No			3.5
Wisconsin	3,4,8,10	Continuous progress indicator	Ratings of schools	Weak to moderate	No	11	2004	2
Wyoming	4,8,11	Only district	Accreditation	Weak	No		2001	1

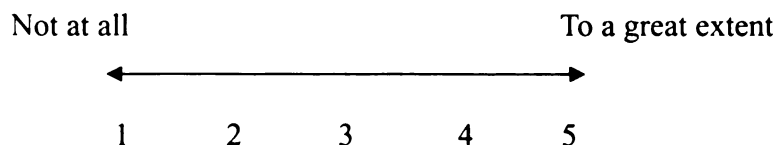
[Source : Carnoy and Loeb (2002)]

2. Number of Sample on Full-time Teachers by State

State ID	State	Number of Teacher	State ID		Number of Teacher
1	Alabama	935	30	Montana	957
2	Alaska	661	31	Nebraska	723
4	Arizona	828	32	Nevada	396
5	Arkansas	710	33	New Hampshire	465
6	California	2,031	34	New Jersey	618
8	Colorado	730	35	New Mexico	621
9	Connecticut	570	36	New York	1,117
10	Delaware	209	37	North Carolina	687
11	District of Columbia	214	38	North Dakota	751
12	Florida	935	39	Ohio	756
13	Georgia	687	40	Oklahoma	1,593
15	Hawaii	403	41	Oregon	632
16	Idaho	661	42	Pennsylvania	739
17	Illinois	818	44	Rhode Island	281
18	Indiana	683	45	South Carolina	623
19	Iowa	697	46	South Dakota	941
20	Kansas	691	47	Tennessee	1,030
21	Kentucky	645	48	Texas	2,105
22	Louisiana	754	49	Utah	619
23	Maine	618	50	Vermont	358
24	Maryland	522	51	Virginia	1,042
25	Massachusetts	614	53	Washington	734
26	Michigan	723	54	West Virginia	613
27	Minnesota	742	55	Wisconsin	735
28	Mississippi	838	56	Wyoming	547
29	Missouri	773			
Total					38,375

3. Questionnaires of Public Teacher Survey on the Use of State or District Standards and Student Test Scores, Which are Employed as Dependent Variables in Analysis.

44. Using the scale 1-5 where 1 is “Not at all” and 5 is “To a great extent,” to what extent do you use state or district standards to guide your instructional practice in your main teaching assignment field?

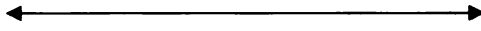


47-B Using the scale 1 – 5, where 1 is “Not at all” and 5 is “To a great extent,” to what extent do you use the information from your students’ test scores.

(1) To group students into different instructional groups by achievement or ability?

Not at all

To a great extent



1 2 3 4 5

(2) To assess areas where you need to strengthen your content knowledge or teaching practice

Not at all

To a great extent

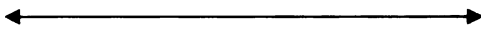


1 2 3 4 5

(3) To adjust your curriculum in areas where your students encountered problems.

Not at all

To a great extent



1 2 3 4 5

4. Ordered Probit Regression and Ordered Logit Regression Model

The ordered probit regression model can be derived from a latent variable model.

$$y^* = XB + e, \quad e|X \sim \text{Normal}(0,1)$$

where B is $k \times 1$ vector. And X is data matrix and it does not contain a constant.

For the five ordered response (from one to five) case used in this paper, let $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$ be unknown cut points, and define

$$y = 1 \quad \text{if } y^* \leq \alpha_1$$

$$y = 2 \quad \text{if } \alpha_1 < y^* \leq \alpha_2$$

$$y = 3 \quad \text{if } \alpha_2 < y^* \leq \alpha_3$$

$$y = 4 \quad \text{if } \alpha_3 < y^* \leq \alpha_4$$

$$y = 5 \quad \text{if } y^* > \alpha_4$$

$$P(y=1|X) = P(y^* \leq \alpha_1|X) = P(XB + e \leq \alpha_1|X) = \Phi(\alpha_1 - XB)$$

$$P(y=2|X) = P(\alpha_1 < y^* \leq \alpha_2|X) = \Phi(\alpha_2 - XB) - \Phi(\alpha_1 - XB)$$

.....

$$P(y=5|X) = P(y^* > \alpha_4|X) = 1 - \Phi(\alpha_4 - XB),$$

where, Φ is the normal cumulative density function. If we replace this normal

cumulative density function with the logit function, we will get ordered logit regression model. The parameters in vector B, can be estimated by maximum likelihood.

If we assign the numeric value on the response, the expected value of the probability of Y can be obtained. That is, $E(Y|X) = 1 P(y=1|X) + 2 P(y=2|X) + 3 P(y=3|X) + 4 P(y=4|X) + 5 P(y=5|X)$

Wooldridge (2001, ch.15) provides more specific explanation on these models.

5. Professional Development Program Participation Rates by State

State	PDStandards	PDdiscipline	PDindepth	PDmethodte	PDassessme
Alabama	0.782062	0.537209	0.665487	0.80926	0.662005
Alaska	0.741954	0.357978	0.552259	0.6599	0.655642
Arizona	0.749077	0.442775	0.574965	0.783267	0.717698
Arkansas	0.744372	0.486365	0.597104	0.755813	0.808584
California	0.780092	0.338789	0.660566	0.766206	0.725105
Colorado	0.82527	0.292052	0.636656	0.714449	0.744196
Connecticut	0.729382	0.341797	0.662089	0.768497	0.728076
Delaware	0.757637	0.395464	0.58364	0.659911	0.685935
District of Columbia	0.903313	0.410614	0.738428	0.811612	0.688479
Florida	0.772308	0.463126	0.662123	0.780923	0.668245
Georgia	0.645933	0.417571	0.571326	0.772687	0.537901
Hawaii	0.820235	0.356009	0.590875	0.666098	0.68911
Idaho	0.597948	0.403119	0.566182	0.680097	0.52658
Illinois	0.707986	0.366366	0.56723	0.69088	0.593003
Indiana	0.614634	0.415817	0.476677	0.745063	0.519365
Iowa	0.723277	0.437334	0.522414	0.655708	0.683066
Kansas	0.739952	0.442166	0.560991	0.774278	0.694007
Kentucky	0.826726	0.541575	0.69714	0.718838	0.743803
Louisiana	0.789728	0.478478	0.590231	0.794669	0.709981
Maine	0.818092	0.329082	0.60739	0.659607	0.810644
Maryland	0.782272	0.40986	0.655762	0.791166	0.780503
Massachusetts	0.78383	0.353279	0.666592	0.729783	0.630786
Michigan	0.705047	0.425784	0.579279	0.75978	0.548377
Minnesota	0.812535	0.314388	0.50832	0.605754	0.667427
Mississippi	0.641922	0.607381	0.557619	0.72587	0.600857
Missouri	0.77951	0.451347	0.5893	0.786196	0.81657
Montana	0.601887	0.479964	0.51549	0.655028	0.496681
Nebraska	0.670027	0.45801	0.488608	0.64939	0.600758
Nevada	0.740025	0.360122	0.59946	0.685318	0.514021
New Hampshire	0.789046	0.423121	0.670842	0.743294	0.672197
New Jersey	0.660493	0.387786	0.51487	0.698954	0.529607
New Mexico	0.654716	0.354146	0.487943	0.656979	0.535901
New York	0.771186	0.308621	0.587907	0.648815	0.658986
North Carolina	0.718554	0.413199	0.567725	0.785353	0.731348

North Dakota	0.532366	0.444193	0.480339	0.622475	0.416043
Ohio	0.673291	0.354036	0.515735	0.719386	0.560769
Oklahoma	0.650936	0.658482	0.597347	0.813363	0.459393
Oregon	0.831704	0.296049	0.608983	0.660007	0.809844
Pennsylvania	0.635968	0.431584	0.493968	0.652567	0.588301
Rhode Island	0.735149	0.171807	0.537676	0.593723	0.580272
South Carolina	0.701022	0.380441	0.566297	0.714685	0.568509
South Dakota	0.668356	0.384034	0.494898	0.598589	0.527814
Tennessee	0.693629	0.540773	0.545347	0.713594	0.585816
Texas	0.751895	0.548769	0.685842	0.824888	0.607998
Utah	0.698101	0.474218	0.667538	0.813867	0.548337
Vermont	0.805801	0.352522	0.615089	0.680395	0.695254
Virginia	0.761481	0.373208	0.560009	0.723452	0.612497
Washington	0.789302	0.286747	0.595062	0.710435	0.761914
West Virginia	0.665472	0.428743	0.513166	0.721719	0.661932
Wisconsin	0.752053	0.291096	0.494367	0.602586	0.586409
Wyoming	0.827571	0.408892	0.55552	0.689378	0.7431

6. Questionnaire used to make the number of professional development programs for school or district administrators.

Does this district provide the following professional development opportunities for school or district administrators? (Include coordinators, supervisors, principals, directors, superintendents, and school board members.)

a. Administrative internships.

1. Yes 2. No.

b. Training in management techniques

1. Yes 2. No.

c. Training in evaluation and supervision

1. Yes 2. No.

d. Training to use technology for planning, budgeting, decision-making, and reporting

1. Yes 2. No.

e. Training about advances in curriculum, teaching and assessment

1. Yes 2. No.

f. Formal networking opportunities for personnel with similar responsibilities

1. Yes 2. No.

g. Reimbursement to attend local, state, and national conferences

1. Yes 2. No.

h. Funding for university or college course work

1. Yes 2. No.

i. Opportunities to serve as mentors within the district

1. Yes 2. No.

j. Strategic planning retreats

1. Yes 2. No.

k. Opportunities to visit schools and districts within and outside of the immediate community

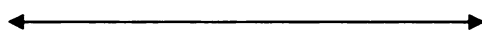
1. Yes 2. No.

7. Questionnaire on the Usefulness of Professional Development Programs

Overall, how useful were these activities (professional development activities) to you?

Not useful at all

Very useful



1

2

3

4

5

****SASS questionnaire asks teachers whether the professional development activities that focus on in-depth studies of the content, content or performance standards, method of teaching, assessment, and discipline were useful respectively.****

REFERENCES

- Abelmann, C. & Elmore, R. with Even, J, Kenyon, S, & Marshall, J. (1999). When accountability knocks, will anyone answer? Consortium for Policy Research in Education *Research Report Series, RR-42*. Graduate School of Education, University of Pennsylvania.
- Amrein, A.L. & Berliner, D.C. (2002a, March 28). High-stakes testing, uncertainty, and student learning *Education Policy Analysis Archives, 10*(18). Retrieved July 18, 2003 from <http://epaa.asu.edu/epaa/v10n18/>.
- Amrein, A.L. & Berliner, D.C. (2002b). The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP Test results in states with high school graduation exams . Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved July 18, 2003 from <http://www.asu.edu/educ/eps/EPRU/documents/EPsL-0211-126-EPRU.pdf>.
- Amrein, A.L. & Berliner, D.C. (2002c). An analysis of some unintended and negative consequences of high-stakes testing. Tempe, AZ: Education Policy Studies Laboratory, Arizona State University. Retrieved July 18, 2003 from <http://www.asu.edu/educ/eps/EPRU/documents/EPsL-0211-125-EPRU.pdf>.
- Barns, C.A. (2002). Standards reform in high-poverty schools : managing conflict and building capacity. New York, Teachers College Press. Braun, H. (2004, January 5). Reconsidering the impact of high-stakes testing, *Education Policy Analysis Archives, 12*(1). Retrieved [Date] from <http://epaa.asu.edu/epaa/v12n1/>.
- Carnoy, M. and Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis, 24*, 305-331.
- Clotfelter, C.T. & Ladd, H.F.(1996). Recognizing and rewarding success in public schools, In H.F. Ladd (ed), *Holding Schools Accountable*, Brookings.
- Clotfelter, C.T., Ladd, H.F., Vigdor, J.L., & Diaz, R.A.(2003). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers? Paper presented at American Economic Association. January, Washington D.C.
- Cohen, D.K. (1990). A revolution in one classroom: the case of Mrs. Oublier, *Educational Evaluation and Policy Analysis*, Fall 1999, v.12,n3, pp.311-329.
- Cohen, D.K. (1996). Standards-based school reform: policy, practice, and performance, In H.F. Ladd (ed), *Holding Schools Accountable*, Brookings.

Cohen, D.K. & Barns, C.A. (1993). Pedagogy and policy, In D. K. Cohen, M. W. McLaughlin, and J. E. Talbert, (Eds). *Teaching for Understanding : Challenges for Policy and Practice*. San Francisco : Jossey-Bass.

Cohen, D.K & Hill, H.C. (2001). *Learning policy: when state education reform works*. New Haven, Yale University Press.

Debray, E., Parson, G., & Woodworth, K. (2001). Patterns of response in four high schools under state accountability policies in Vermont and New York. In S. H. Fuhrman (ed), *From the capitol to the classroom: standards-based reform in the states*. University of Chicago Press.

Education Commission of the States (2000). Informing practices & improving results with data-driven decisions. <http://www.ecs.org/clearinghouse/31/12/3112.htm>

Education Commission of the States (2001). Rewards and Sanctions for School Districts and Schools, Last Updated in March 2001.
<http://www.ecs.org/clearinghouse/18/24/1824.htm>

Elmore, R.F., Abelman, C.H. & Fuhrman, S.H.(1996). The new accountability in state education reform: from process to performance, In H.F. Ladd (ed.), *Holding schools accountable*, Brookings.

Fuhrman, S.H. (2001). *From the capitol to the classroom : standards-based reform in the states*. University of Chicago Press.

Good, T.M. & Brophy, J.E. (2002). *Looking in Classrooms*, 9th Edition. Person Education inc.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). Improving student achievement: What NAEP test scores tell us. Santa Monica, CA: RAND Corporation. Available at <http://www.rand.org/publications/MR/MR924>

Hanushek, E. et al., (1994). *Making schools work: improving performance and controlling costs*. Brookings.

Hanushek, E.& Raymond, (2001). The Confusing World of Educational Accountability, National Tax Journal, Vol. LIV, No2.

Harris, D.C. (2002). Lowering the bar or moving the target: A wage decomposition of Michigan's charter and traditional public school teacher. Working Paper. The Education Policy Center, Michigan State University. Available at www.epc.msu.edu

Hawley, W.D. and Valli, L. (1999). The essentials of effective professional development: A new consensus, in L. Darling-Hammond and G. Sykes (coed.), *Teaching as the Learning Profession: Handbook of Policy and Practice*. Jossey-Bass. San Francisco.

Kanstoroom, M. (2000). Value-added assessment: ready for prime time? Education Leaders Council. *Newsletter*, Summer 2000.

Ladd, H. (1999). The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes. *Economics of Education Review* 18. pp.1-16.

Lankford, Hamilton, Susanna Loeb, Susanna, and Wykoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), Spring, pp. 37-62.

Massell, D. (2001). The theory and practice of using data to build capacity: state and local strategies and their effects, In S. H. Fuhrman (ed), *From the capitol to the classroom: standards-based reform in the states*. University of Chicago Press.

Newmann, F.M., King, M.B., & Rigdon, M., (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review*, V67, N1. pp 41-69.

North Carolina Department of Public Instruction, School Improvement Division, Strategies to improve instructions
<http://www.ncpublicschools.org/schoolimprovement/closingthegap/strategies/movement/schools.shtml>

Roderick, M., Jacob, B.A., & Bryk, A.S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis* (2002), V24, N4. pp.333-357

Rosenshine, B. (2003, August 4). High-stakes testing: Another analysis. *Education Policy Analysis Archives*, 11(24). Retrieved August 4, 2003 from <http://epaa.asu.edu/epaa/v11n24/>.

Sanders, W. L. (2000). Value-added assessment from student achievement data: opportunities and hurdles, SASinSchool, SAS Institute, Inc. Cary, NC.

Smith, M. & O'Day, J. (1990). Systemic school reform, In *Politics of Education Association Yearbook*.

Stone, J.E., Value-added assessment: an accountability revolution,
<http://edexcellence.net/better/tchrs/16.htm>

Supovitz, J.A. & Klein, V. (2003). Mapping a course for improved student learning: how innovative schools systematically use student performance data to guide improvement. Consortium for Policy Research in Education, University of Pennsylvania.

Swanson, C.B., & Stevenson, D.L. (2002). Standards based reform in practice: Evidence

on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis*, V24. N1. 1-27

Wolfe, E. W., Ray, L. M., & Harris, D. C. (in press). A Rasch analysis of three measures of teacher perception. *Educational and Psychological Measurement*.

Wooldridge, J. (1999). *Introductory Econometrics: A Modern Approach*. South-Western.

Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

CHAPTER III

DOES ACCOUNTABILITY POLICY DIMINISH TEACHERS' INTRINSIC MOTIVATION? EVIDENCE FROM THE 2000 SASS DATABASE

1. Introduction

Performance-based accountability policy has been implemented in several states since 1990. Current federal education policy, embodied in the No Child Left Behind Act, calls for the enactment of strong accountability policies in all states. The NCLB requires implementation of annual student testing and a series of increasingly severe sanctions for the schools that do not meet their annual yearly progress goals. Certainly the new federal law brings more pressure and centralized control to K-12 public education. However, little research has examined the psychological effect of this pressure on teachers.

Most research about the effect of accountability policy has focused on the evaluation of whether states adopting such policies have improved their student achievement on the National Assessment of Educational Program (NAEP) or other standardized tests (Grissmer, et al., 2000; Roderick, et al., 2002). Such evaluations adopt a top-down perspective based on the rationale of accountability policy and check whether the policy has yielded expected results. They do not investigate the policy's effect on actual instructional practice in classrooms and schools. Other researchers focus on how the accountability policies will work inside schools and provide some useful lessons about their effects (Abelmann, et al., 1999; Barns, 2002; Newmann, et al., 1997).

Nevertheless, possible psychological effects of accountability policies on teaching have not been subject to systematic empirical study.

Actually, there are relatively few studies about the psychological effect of any education policy on teachers. This is expected since there is little research on the teachers' learning or motivation from the psychology discipline. The well known textbook about motivation theory, *Motivation in Education* (Pintrich and Schunk, 2002) only deals with students' motivation and learning processes and provides suggestions on how to motivate students to learn. Application of motivation theory to teachers –for instance, how to cultivate teachers' self-determination to teach, or how policy can elicit teachers' commitment in multicultural situations – is rare²⁶. Since teachers' learning has been ignored by many education scholars (Cohen, 1990), the psychological approach to teachers' motivation to teach or learn has been largely disregarded. Bandura's self-efficacy theory has received some attention. However, his theory was limited to illuminating the effect of teacher self-efficacy on students' learning or the effect of organizational factors on teachers' self-efficacy (Goddard, et al., 2000; Tschannen-Moran, et al., 1998). The psychological effect of policy on teachers has been largely neglected by scholars of self-efficacy theory.

Recently, Sheldon and Biddle (1998) introduced self-determination theory to argue that the accountability policy will have detrimental effects on teachers and student learning. They hypothesize that rigid standards and accountability guidelines, and tangible sanctions may diminish the motivation and performance of teachers and students.

²⁶ Bess, J.L. (1997) collects motivation theories to discuss how to motivate faculty to research and teach intrinsically. However, the book does not provide any empirical research and only puts forth general arguments. The book also does not deal with K-12 teachers.

Unfortunately, they did not provide any empirical evidence directly related to the effect of the accountability policy on teacher motivation. They only provide some research concerning the negative consequences of teachers' controlling-style instructional practices on student learning or motivation. However, their argument is significant since it addresses the issue of psychological effect of education policy on teachers and students.

Self-determination theory is one of the most comprehensive motivation theories which has been reinforced by empirical evidence (Pintrich and Schunk, 2002). This paper adopts self-determination theory to examine the effect of accountability policy on teacher motivation empirically. Specifically, we will analyze the 2000 National Center for Educational Statistics (NCES) Schools and Staffing Surveys (SASS) database to evaluate the prediction of self-determination theory that current accountability policies with performance-contingent rewards and sanctions will undermine teachers' intrinsic motivation to teach.

2. Accountability Framed by Principal-Agent Model

Current accountability policy is implicitly framed by a naive principal-agent model. Test-score-contingent rewards for schools and teachers and sanctions on failing schools are the main features of the policy. That is, the accountability policy proceeds from an assumption that teachers and schools will improve test scores when monetary incentives and sanctions are provided. This perspective has two assumptions. First, teachers are not intrinsically motivated to exert high levels of effort, and monetary incentives or punishment will elicit increased effort. By this perspective, changes in

teachers' behavior can be reliably spurred only by external incentives. Second, teachers have different goals than the state. The principal-agent model presumes that agents are working for their own interests which deviate from the principal's goals. Teachers, for example, may care more about maximizing their leisure than increasing student achievement. So, a monetary incentive system or linear salary payment contract is necessary to induce teachers into working for the state's (i.e., the principal's) goal of increasing students' test scores. Thus, the policy consists of deadlines, performance-contingent rewards and punishments.

Previous research, however, suggests that these assumptions do not correctly reflect teachers' decision to teach. Teachers have intrinsic reasons or motivation to select their jobs (Feiman-Nemser and Floden, 1986; Lortie, 2002). Lortie (2002) provides some reasons why teachers choose to teach. From the national survey conducted by the National Education Association and intensive interviews with teachers in the Boston metropolitan area, Lortie (2002) found that one of the main reasons that teachers choose their job is their "desire to work with young people." Teachers value interpersonal work and caring for youngsters. The idea that teaching is "a valuable service of special moral worth" is another reason for teachers to select their job. That is, teachers respond that the opportunity to render an important service is one of their main reasons for teaching. More than half of teachers chose these two reasons for teaching. Certainly some choose the teaching occupation, because it offers relatively secure employment with regular hours and summer vacation. However, these reasons are not used by those who recruit teachers, and at least by teachers' accounts they represent secondary considerations (Lortie, 2002).

Another special characteristic of teaching is that the main reward is psychic or intrinsic, not extrinsic. The culture of teachers and structure of teaching rewards favors emphasis on psychic rewards (Lortie, 2002). Historically, teachers have favored egalitarian compensation systems and they continue to oppose differentiation in salary on grounds other than seniority or education (Lortie, 2002; Tyack and Cuban, 1995). If teachers' primary motivation for teaching is intrinsic, then external rewards may not affect teachers' effort directly. Some of the psychic rewards of teaching include the chance to study, read, and plan for classes; classroom management; and the chance to associate with young people and other teachers (Lortie, 2002). Among the reasons, most teachers (86.1%) claim to receive psychic rewards from "knowing that I have reached students and they have learned" (Lortie, 2002). If we can accept their self-reported answers, teachers certainly receive psychic or intrinsic rewards from their work.

Therefore, at least in terms of their own accounts of what motivates them, the pleasure of working with young students is teachers' main motivation to teach. Teachers obtain intrinsic reward from knowing that they have reached students and students have learned. The assumptions underlying accountability policy appear at odds with this fundamental dimension of teachers' work experience.

3. Self-Determination Theory

Self-determination theory starts with the assumption that a person has innate and constructive tendencies to develop a more elaborated and unified sense of self (Deci and Ryan, 1985; Ryan and Deci, 2002). Intrinsic motivation is based on a basic human

need to be competent and self-determining. To be intrinsically motivated, a person must feel free from pressures and experience his/her action as autonomous (Deci and Ryan, 1985).

One of the main questions that self-determination theory intends to answer is: If a person who is involved in an intrinsically motivated activity begins to receive an extrinsic reward for doing it, will his intrinsic motivation be enhanced or decreased? (Deci and Ryan, 1985). Self-determination theory states that:

External events relevant to the initiation or regulation of behavior will affect a person's intrinsic motivation to the extent that they influence the perceived locus of causality for that behavior. Events that promote a more external perceived locus of causality will undermine intrinsic motivation, whereas those that promote a more internal perceived locus of causality will enhance intrinsic motivation. (Deci and Ryan, 1985, p. 62)

For instance, imagine that a teacher originally likes to teach students and receives psychic rewards from improving his/her students' achievement. One day a school principal or the state imposes a contingent reward/sanction for teachers based on student achievement. Then, what will happen to the teacher's original enjoyment of teaching or intrinsic motivation and psychic reward? According to the above statement of self-determination theory, one would predict that teachers' enjoyment of teaching and the psychic reward will be replaced by enjoyment of receiving the monetary reward. That is, the perceived locus of causality to teach will shift from internal psychic reward to

external monetary rewards. As a monetary incentive is provided, the internal reason for teaching is replaced by the monetary reason, and the intrinsic motivation is diminished.

However, self-determination theory further proposes that the external events do not always have a detrimental effect on the intrinsic motivation. The theory specifies conditions under which outside events such as reward and surveillance will suppress intrinsic motivation as follows:

Events relevant to the initiation and regulation of behavior have three potential aspects, each with a functional significance. The informational aspect facilitates an internal perceived locus of causality and perceived competence, thus enhancing intrinsic motivation. The controlling aspect facilitates an external perceived locus of causality, thus undermining intrinsic motivation and promoting extrinsic compliance or defiance. The amotivating aspect facilitates perceived incompetence, thus undermining intrinsic motivation and promoting amotivation. The relative salience of these three aspects to a person determines the functional significance of the event. (Deci and Ryan, 1985, p. 64)

Research on self-determination has shown that positive feedback, encouragement of autonomy and choice are informational (Deci, 1971, 1995; Grolnick, & Ryan, 1987), while performance-contingent rewards, deadlines, surveillance are regarded as controlling by agents (Amabile, 1979; Amabile, et al., 1976; Deci, et al., 1981; Enzle and Anderson, 1993; Lepper and Greene, 1975).

Monetary reward need not be contrary to internal motivation so long as it is not attached to the performance and does not contain a controlling aspect. Sometimes the monetary incentive can increase job satisfaction. However, Deci and Ryan (1985) mention that increased job satisfaction from more monetary reward or extrinsic reward is not identical with increased intrinsic motivation. When people experience a sense of choice in initiating and regulating their own actions and feel internal causality of locus for their work, they are self-determined. Self-determination theory predicts that when teachers perceive the locus of causality for their work as internal, they are intrinsically motivated and commit themselves to teaching and consequently their students will benefit. Task-contingent rewards or sanctions and other mechanisms will undermine teacher's intrinsic motivation and make them alienated from their work.

Deci and Ryan's self-determination theory has received some attention from economists. Kreps (1997), for example, draws on intrinsic motivation theory to argue that simplistic application of monetary incentives to employees must be considered carefully. He notes that jobs high in intrinsic motivation often involve implementation of ambiguous tasks. Creativity is required to effectively perform tasks involving ambiguity. In this situation it would be difficult to get incentives right. People work hard when they really enjoy it. However if extrinsic incentives are imposed, people will attribute his efforts to those incentives, developing a distaste for the required efforts. Thus, to complement intrinsic motivation, economic incentives should emphasize the voluntary nature of the desired behavior (Kreps, 1997).

4. Literature Review

There has been some research in psychology, which examines the effect of external reward or standards on teaching and learning. Garbarino (1975) explored the effect of the imposition of anticipated and contingent reward on the interaction style of an older child acting as a tutor for a younger one. Two groups of fifth and sixth graders were trained to help teach first and second graders. One group was told that they would be given a free ticket to the movies only if the younger children learn how to play well. The other group received no statement about rewards. Tutor behavior, student performance, and interaction context were measured. The results indicated that tutors in the reward condition evaluated the younger children and their performance more negatively. The children taught by the tutors who received rewards displayed less learning and more errors in their performance. Measures of interaction showed that tutoring in the no-reward condition was rated as significantly more positive in emotional tone than in the rewarded condition, and there were significantly more instances of laughter in the no-reward condition. The rate of learning per unit of time was also higher for the no-reward condition.

Deci, et al. (1982) examined what conditions make teachers more controlling or more autonomy-oriented with students. Self-determination theory implies that when pressured toward particular outcomes, teachers may become more controlling with their students, which could diminish the intrinsic motivation of those students. Deci, et al. (1982) test the hypothesis that imposing responsibility to teachers for their students' performing up to standards will impose more pressure on teachers and make them more

controlling with their students. Their experiment shows that teachers who had been given the performance standards induction were much more demanding and controlling than teachers in the no-performance-standards condition. These controlling teachers made twice as many utterances: they allowed student to work alone much less and they gave three times as many directives and should-type statements. The experiment illustrates that teachers, when they feel pressure, tend to lecture and explain more and provide less choice and less opportunity for independent or autonomous students learning. Deci, et al. (1982) concludes that performance standards need to be communicated in an informational way, otherwise the standards could be experienced as pressure by teachers and negatively affect teaching and learning. These two studies (Deci, et al., 1982; Garbarino, 1975) directly deal with the effect of externally imposed rewards and standards on teaching and learning. Both reinforce the view that reward and standards should be provided in an informational, not controlling, way.

More studies on teaching style show that autonomy-supportive teaching has positive effects on student learning. Benware and Deci (1984) explore rote and conceptual learning under active conditions and passive conditions. Active condition means that students learn materials in order to teach, while passive condition means that student learn materials to be tested. Their experiment assigned students to two groups, the experimental group (learning in order to teach) and the control group (learning in order to take an exam) and assessed intrinsic motivation of these two groups: how interesting subjects found the contents of the learning materials, how enjoyable they found the experiment, and how much additional time they were willing to volunteer for the experiment. The experimental group showed significantly higher interest and

enjoyment, and further participation. In addition, the conceptual learning score of the experimental group was much higher than in the control group. So, Benward and Deci's (1983) study illustrates that an active learning paradigm could enhance students' intrinsic motivation to learn and facilitate deeper learning.

Other research also examines the effects of autonomous teaching style and students' perceive autonomy on student performance (Deci, et al., 1981; Flink, et al., 1990; Miserandino, 1996; Reeve et al., 1999) and investigates whether dropping out of high school is correlated with students' low level of self-determination (Vallerand, et al., 1997). Although these studies do not explore the effects of external instruments on teachers and student learning, they provide an important lesson for the quality of education: student learning can be enforced within an autonomous environment. Thus, these studies suggest that accountability policies which create a more controlling environment for education and push teachers to use more controlling instructional practice would undermine the quality of student learning.

5. Study Hypotheses

Self-determination theory predicts that the teachers in states with strong accountability policies will be more likely to feel alienated from their work and find teaching to be less attractive. That is, they will be more likely to respond that they would not be a teacher again if they were to start over again and it is a waste of time to try to do best as a teacher.

6. Data

The National Center for Educational Statistics has conducted a national teacher and school staff survey, the Schools and Staffing Survey (SASS). SASS has been implemented in school years 1987-1988, 1990-1991, 1993-1994, and 1999-2000. SASS uses stratified random sampling to represent the national population. SASS surveys teachers, principals, administrators, district administrators and includes public, charter, private schools. We analyze data for 38,375 full-time teachers from the 1999-2000 SASS public school teacher survey. The SASS public school administrator survey provides many useful school-level variables. The STATA software provides information about the represented population number by the sample size, so we can see how much population is represented in the model.

The SASS survey includes three questions that provide proxies of teachers' intrinsic motivation, whether teachers:

- would not become a teacher again if they were to start over in college
- think that it is not waste of time to do their best as a teacher,
- are dissatisfied with being a teacher at their schools,

These dependent variables generally reflect teachers' intrinsic motivation and commitment on teaching (Appendix provides these questionnaires). If teachers lose intrinsic motivation to teach, they are less likely to say they would become a teacher again if they were to start their life again. Thus, as dependent variable, teachers' response

about whether they would become a teacher again serves as a good indicator of teachers' intrinsic motivation. The second dependent variable, the extent to which teachers think it is not waste of time to do their best as a teacher, would be the best proxy for teachers' intrinsic motivation. Teachers who value their teaching job highly and receive more fulfillment from it are regarded as motivated innately and we can expect that such highly intrinsically motivated teachers will try to do best as a teacher.

The third dependent variable measuring teachers' satisfaction could also reflect teachers' intrinsic motivation in some degree, however, not exclusively. For instance, a monetary reward provided under strong accountability policy could increase teachers' satisfaction. The increased satisfaction is not identical with enhancement of intrinsic motivation (Deci & Ryan, 1985). Despite this imprecision in interpretation, there is no reason to exclude the variable from analysis.

The explanatory variables include variables reflecting teachers' characteristics, school characteristic, professional development, and accountability policy. The accountability policy variable is adopted from Carnoy and Loeb's (2002) index of state accountability policy. If teachers are teaching in states with strong accountability policies such as Alabama, North Carolina, Texas, California, Florida, New Jersey, New York, New Mexico, Kentucky, or Maryland, the accountability policy variable is one, otherwise zero. These strong accountability stats have monetary rewards, and sanctions, while the other states do not (See Appendix Table 2). On Carnoy and Loeb's 0-5 scale, these stats had an average accountability index value of 4.6. If accountability policies cause teachers to feel pressure, it should be strongest in these states. Table 3.1 provides the definition of dependent and independent variables.

Table 3.1: Definitions of Dependent and Independent Variables Used in the Analysis

Independent Variables:

Teacher (Basic) Characteristics Variables

- Male : Dummy variable which takes on the value 1 if the teacher is male and 0 if the teacher is female.
- Minority: Dummy variable which takes on the value 1 if the teacher is minority and 0 if the teacher is white.
- Age: Continuous variable indicating the age of teacher.
- Sqage : Continuous variable, which is the square value of age
- TotExp: Continuous variable. Total teaching experience measured by years.
- Sqtotexp: Continuous variable.
Square of TotExp.
- Salary: Continuous variable. Teacher Annual Salary
- Unionmem: Dummy variable which takes on value 1 if the teacher is union member, otherwise 0
- Mathscie: Dummy variable which takes on value 1 if the teacher's main teaching assignment field is math or science.

Teacher Knowledge or Ability Variables

- MA: Dummy variable which takes on value 1 if the teacher has a master degree, otherwise 0.
- Mathalba: Dummy variable which takes on value 1 if the teacher's college major is math or math education, otherwise 0.
- SciBA: Dummy variable which takes on value 1 if the teacher's college major is science or science education, otherwise 0.
- Verycomp: Selectivity of undergraduate institution. Dummy variable which takes on value 1 if the teacher's undergraduate institution is very competitive, highly competitive or the most competitive, 0 if the teacher's undergraduate institution is competitive or less competitive, non competitive or special. This selectivity of undergraduate institution is from the ratings of Barron's 2001 Profiles of American Colleges. This variable can be a proxy for the teacher's innate ability.
- Certrec : Dummy variable which takes on value 1 if the teacher obtained teaching certification which is regular, advanced, provisional or probational in her/his main teaching assignment, 0 if the teacher reports that temporary, emergency or no certification..

Teacher Professional Development

- PDindepth : Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on in-depth study of the content in his or her main teaching assignment field in the past 12 months. 0 means the teacher did not participate.
- PDstandards: Dummy variable which takes on value 1 if the teacher participated

in any professional development activities that focused on content and performance standards in his or her main teaching assignment field in the past 12 months. Otherwise 0.

PDmethodte: Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on methods of teaching in the past 12 months. Otherwise 0.

PDassessme: Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on student assessment, such as methods of testing, evaluation, performance assessment, etc in the past 12 months. Otherwise 0.

PDdiscipline : Dummy variable which takes on value 1 if the teacher participated in any professional development activities that focused on student discipline and management in the classroom in the past 12 months. Otherwise 0.

Teacher Perception Variables²⁷

Zinfluence: Continuous (scaled) variable. Higher score indicates higher perception of influence in school policy such as setting performance standards for students, establishing curriculum, evaluating teachers, hiring new full-time teachers, setting discipline policy, deciding the usage of school budget, and determining the contents of in-service professional development program.

ZControl: Continuous (scaled) variable. Higher score indicates that the teachers perceive that they have much control over following areas such as selecting textbooks and other instructional materials, selecting content topics, and skills to be taught, selecting teaching techniques, evaluating and grading student, disciplining students, and determining the amount of homework to be assigned.

ZStudent: Continuous (scaled) variable. Higher score means that teachers perceive no serious student problem and low score means that teachers perceive serious student problem. Examples of student problems are: student tardiness, absenteeism, robbery of theft, pregnancy, alcohol, and so on.

ZClimate: Continuous (scaled) variable. Higher score means that teachers perceive a worse school climate and lower score means that teachers perceive a better school climate.

School variables

PerFRLkw: Continuous variable. Percentage of student receiving free or reduced lunch

NewminPER: Continuous variable. Percentage of student of color

Totalenroll: Continuous variable. School size. Total enrollment of student.

²⁷ Please see the appendix A in the working paper, Debbi Harris (2002), Lowering the bar or moving the target: A wage decomposition of Michigan's charter and traditional public school teacher, for more information about these scaled variables. The paper is available at www.epc.msu.edu.

Suburban: Dummy variable that takes value 1 if the school is located at suburban area.

Accountability Policy Variable

Strongacc: Dummy variable which takes the value 1 if the teacher is from the states with strong accountability policy (Alabama, North Carolina, Texas, California, Florida, New Jersey, New York, New Mexico, Kentucky, Maryland), 0 otherwise.

Dependent Variables

NotBeTeacher: Scale is one to five. One means that the teacher certainly would become a teacher if he/she could go back to his/her college days and start over again. Five means that the teacher certainly would not become a teacher. Higher scale means that it is less likely that the teacher would become a teacher again.

Notwasteoftime: Scale is one to four. One means that the teacher strongly agrees that he/she sometimes feels it is a waste of time to try to do his/her best as a teacher. Four means that the teacher strongly disagrees that he/she feels it is a waste of time to try to do his/her best as a teacher. Higher scale indicates that the teacher feels that it is not a waste of time to try to do his/her best as a teacher.

Notsatisfaction: Scale is one to four. One means that the teacher strongly agrees that he/she is generally satisfied with being a teacher at the school. Four means that the teacher strongly disagrees that he/she is generally satisfied with being a teacher at the school. Higher scale means less satisfaction.

The square term of age variable is included in the model since the relationship between age and the dependent variable, NotBeTeacher, could be U-shape. Math or science teachers are more likely to leave teaching since other job opportunities are more open to them (Ingersoll, 2001; Murnane, et al., 1991). Consequently, a dummy variable, whether teachers are math or science teachers is included. Table 3.2 provides basic statistics on all the dependent and independent variables. Average public school teacher age is forty two and a quarter of teachers is male. Only 4 percent of teach majored in math during college and only 5.3 percent majored in science. Almost half of teachers possesses masters degree. In 2000, 35 percent of teachers worked in states with strong

accountability policies.

Table 3.2: Summary Statistics

Mean	Estimate	Std. Err.	Observation	Pop. Size
NotBeTeacher*	2.136	0.010	38,375	2,727,067
Notsatisfaction*	1.600	0.006	38,375	2,727,067
Notwasteoftime*	3.374	0.007	38,375	2,727,067
Age	42.236	0.090	38,375	2,727,067
Sqage	1,897.322	7.556	38,375	2,727,067
Salary	39,928.240	99.506	38,375	2,727,067
Totexper	14.808	0.084	38,375	2,727,067
Sqtotexp	321.613	2.894	38,375	2,727,067
Male	0.255	0.003	38,375	2,727,067
Minority	0.160	0.003	38,375	2,727,067
Unionmem	0.797	0.003	38,375	2,727,067
Mathalba	0.040	0.001	38,375	2,727,067
SciBA	0.053	0.002	38,375	2,727,067
Mathscie	0.135	0.002	38,375	2,727,067
MA	0.459	0.004	37,994	2,709,439
Verycomp	0.269	0.004	38,375	2,727,067
Certrec	0.930	0.002	38,375	2,727,067
Zcontrol	-0.028	0.008	38,375	2,727,067
Zinfluen	-0.019	0.008	38,375	2,727,067
Zstudent	-0.032	0.008	38,375	2,727,067
Zclimate	0.026	0.008	38,375	2,727,067
NewminPER	34.977	0.270	38,214	2,718,586
PerFRLkw	38.582	0.256	34,421	2,455,204
Totalenroll	825.902	4.165	35,333	2,495,093
Stutearatio	15.830	0.031	35,333	2,495,093
Suburban	0.501	0.004	38,375	2,727,067
PDdiscipline	0.411	0.004	38,375	2,727,067
PDindepth	0.593	0.004	38,375	2,727,067
PDmethodte	0.733	0.004	38,375	2,727,067
PDassessme	0.640	0.004	38,375	2,727,067
PDstandards	0.734	0.004	38,375	2,727,067
Strongacc	0.351	0.003	38,375	2,727,067

* indicates dependent variable.

7. Method

The basic method to measure the effect of accountability policy on teachers' intrinsic motivation is regression analysis. The full regression model is:

$$\text{Teacher's intrinsic motivation} = B_0 + X_1B_1 + X_2B_2 + X_3B_3 + X_4B_4 + X_5B_5 + u$$

where X_1 is a vector of teachers' characteristics such gender, race, college major. X_2 is a vector of teachers' perceptions. X_3 is a vector of school characteristics and X_4 is a vector of teacher professional development relevant variables. X_5 is a dummy variable whether teachers are from the states of strong accountability policy.

An ordered probit regression model and an ordered logit regression model will be employed since the dependent variable is ordered response of teachers on the survey question⁴. For instance, to answer the survey question, "I sometimes feel it is waste of time to try to do my best as a teacher", which is used as a dependent variable, the teacher needs to choose one to five scale of answer where five means strongly disagree and one means strongly agree. So, five means that teachers strongly disagree with the statement that it is waste of time to try to do one's best as a teacher. It is hard to say that the scale exactly has the numeric mean. The difference between scale four and scale two does not necessarily mean that it is twice as influential as the difference between scale one and two. We can only know that five means more influence of state or district standards than four, and four means more influence than three, in other words, the response scale has ordinal meaning. However, linear regression result also will be provided to check

whether ordinary least square linear regression produces significantly different results compared to the ordered probit and ordered logit models. If so, ordered probit or ordered logit model estimates are preferred. Otherwise, looking at the results of linear regression model for the convenient interpretation of coefficient size will be fine.

8. Results

8.1 Are Teachers Under the Strong Accountability Policy Less Likely to Become a Teacher Again, If They Were to Start Over?

The dependent variable, whether teachers would become a teacher again, certainly captures teachers' feeling of their job's attractiveness and their current motivation or self-determination to teach. If teachers have lost their motivation or interest in teaching, they would answer that they would not become teachers again if they could start over again. If teachers experience self-determination in their work or received sufficient psychic reward, they are more likely to choose to become a teacher again. Thus, the dependent variable can be deemed as a good proxy of intrinsic motivation.

Certainly, organizational and individual factors influence teachers' perception of the attractiveness of teaching. Thus, other possible reasons must be controlled in the analysis in order to isolate the effect of accountability policy on teachers' perception on teaching again.

Table 3.3: Four Ordered Probit Models on the Teachers' Perception that They Would not Become a Teacher Again

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>		<u>Model 4</u>	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Age			0.025***	0.008	0.025***	0.008	0.025***	0.008
Sqage			0.00001***	0.000	0.000***	0.000	0.000***	0.000
Salary			0.000***	0.000	0.000***	0.000	0.000***	0.000
Totexper			0.044***	0.004	0.045***	0.005	0.045***	0.005
Sqtotexp			-0.001***	0.000	-0.001***	0.000	-0.001***	0.000
Male			0.175***	0.019	0.142***	0.020	0.114***	0.022
Minority			0.039	0.029	0.053*	0.029	0.058*	0.033
Unionmem					-0.069***	0.022	-0.115***	0.023
Mathscie					0.210***	0.034	0.174***	0.035
Mathalba					-0.073*	0.044	-0.097**	0.047
SciBA					0.092**	0.042	0.045	0.045
MA					0.052**	0.020	0.039*	0.022
Verycomp					0.134***	0.021	0.112***	0.022
Certrec					-0.063*	0.037	-0.034	0.040
Zcontrol							-0.084***	0.011
Zinfluen							-0.052***	0.012
Zstudent							-0.007	0.014
Zclimate							0.367***	0.014
NewminPER							0.000	0.000
PerFRLkw							-0.001*	0.001
Totalenroll							0.000**	0.000
Stutearatio							0.000	0.002
Suburban							-0.006	0.021
Strongacc	0.077***	0.020	0.109***	0.021	0.109***	0.021	0.103***	0.023

Dependent variable: NotBeTeacher. For model 1 and 2, Number of obs=38,375, Population size=2,727,066. For model 3, Number of obs = 37,994, Population size = 2,709,439. For model 4, Number of obs = 34,109, Population size = 2,440,181. For all model, number of strata = 51. *** means that the coefficient is statistically significant at the 0.01 level. ** means statistical significance at the 0.05 level. * means p-value < 0.10.

Table 3.3 displays four ordered probit models of the determinants of teachers' perception of whether they would not become a teacher again if they could start over in college. Model 1 in Table 3.3 includes only the accountability policy variable, while Models 2, 3, and 4 include other teacher and school variables. The significance and sign

of the accountability variable does not change across the four models. The size of coefficient on the accountability variable progressively increases as we specify more control variables in the Models 2, 3, and 4. Even in the probit model of Table 4, the size of the accountability variable coefficient increases further. Thus, across the five models, we can find strong and consistent evidence that teachers working under strong accountability policies are more likely to report that they would not become a teacher again.

Table 3.4 presents three full models, - linear OLS regression, ordered probit, and ordered logit models. Interestingly male teachers were more likely to respond that they would not become a teacher again. If they were given opportunities to start over again, male teachers would be more likely to choose other occupations. Math and science teachers and teachers from very, highly or mostly competitive colleges also responded that they would not become teachers again. An MA degree is marginally significant at the 0.1 level. Certainly teachers who have more capabilities and chances of other job opportunities are more likely to perceive that they would not become a teacher again, if they were to start over. Both experience and age variables have significant effects and the relationship is non-linear as expected. Zcontrol, which indicates the extent that teachers have control on classroom instruction and content and skills to be taught, has a negative and significant coefficient. That is, teachers who have more control in classroom instruction are more likely to become a teacher again. Zinfluen also has the expected coefficient sign. Teachers who are more influential in school policies respond that they would become a teacher again. Zclimate variable provides evidence that organizational factors have an effect on teachers' perception of becoming a teacher

again.

Table 3.4: Effects on the Teacher's Perception on Whether They Would Not Become a Teacher

	<u>Linear</u>		<u>Ordered Probit</u>		<u>Ordered Logit</u>	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Age	0.023***	0.008	0.025***	0.008	0.038***	0.014
Sqage	-0.0002***	0.0001	-0.0003***	0.0001	-0.0004***	0.0002
Salary	-0.00001***	0.000001	-0.000009***	0.000001	-0.000016***	0.000002
Totexper	0.043***	0.005	0.046***	0.005	0.078***	0.008
Sqtotexp	-0.001***	0.0001	-0.001***	0.0001	-0.001***	0.0002
Male	0.107***	0.022	0.102***	0.022	0.164***	0.037
Minority	0.075**	0.033	0.062*	0.033	0.090	0.056
Unionmem	-0.108***	0.024	-0.111***	0.023	-0.184***	0.040
Mathscie	0.185***	0.038	0.170***	0.036	0.288***	0.061
Mathalba	-0.121**	0.049	-0.101**	0.048	-0.168**	0.082
SciBA	0.029	0.047	0.042	0.045	0.068	0.077
MA	0.039*	0.022	0.041*	0.022	0.079**	0.037
Verycomp	0.098***	0.022	0.113***	0.022	0.186***	0.037
Certrec	-0.016	0.039	-0.028	0.040	-0.049	0.068
Zcontrol	-0.085***	0.011	-0.089***	0.011	-0.155***	0.019
Zinfluen	-0.056***	0.012	-0.047***	0.012	-0.077***	0.021
Zpercept	-0.004	0.013	-0.007	0.014	-0.013	0.023
Zclimate	0.343***	0.013	0.361***	0.014	0.618***	0.024
NewminPER	-0.0001	0.0005	0.000	0.0005	0.000	0.001
PerFRLkw	-0.001	0.001	-0.001	0.001	-0.002*	0.001
Totalenroll	-0.00004**	0.00002	-0.00004**	0.00002	-0.00006**	0.00003
Stutearatio	-0.0001	0.002	-0.00004	0.002	-0.001	0.003
Suburban	-0.005	0.021	-0.004	0.021	-0.018	0.036
PDdiscipline	-0.025	0.021	-0.029	0.021	-0.056	0.035
PDindepth	-0.042*	0.021	-0.043**	0.021	-0.071*	0.036
PDmethodte	-0.061***	0.023	-0.061***	0.023	-0.103***	0.039
PDassessme	-0.017	0.022	-0.016	0.022	-0.026	0.037
PDstandards	-0.051**	0.024	-0.048**	0.024	-0.074*	0.040
Strongacc	0.113***	0.023	0.110***	0.023	0.175***	0.039
cons	1.704***	0.159				

Note: Dependent Variable : NotBeTeacher. Number of obs = 34,109, Number of strata = 51, Population size = 2,440,181 *** means that the coefficient is statistically significant at the 0.01 level. ** means statistical significance at the 0.05 level. * means p-value < 0.10.

Among teacher professional development variables, participations in activities that focused on teaching methods and content and performance standards have significant effects on teachers' willingness to become a teacher again. That is, opportunities to learn would increase the likelihood that teachers would become a teacher again, if they start over again.

The policy variable, Strongacc, which we are interested in, has a very significant positive effect. That is, teachers who are working under the strong accountability policy respond that they would not become a teacher again if they can start over in college. This means that accountability policy undermines the teachers' perception of the attractiveness of teaching at public schools or teachers lose their intrinsic motivation for teaching when working under accountability policies.

8.2 Do Teachers Under Strong Accountability Policy Become More Likely to Think That It is Waste of Time to Try to Do Best as a Teacher?

Whether or not teachers feel it is not waste of time to try to do best as a teacher, is perhaps the best proxy variable reflecting teachers' intrinsic motivation to teach. Teachers who receive intrinsic reward from teaching would answer that they are more likely to try to do best as a teacher, while teachers who do not receive any intrinsic reward from teaching will think that it is waste of time to teach hard. Because other factors such as salary structures or union membership might influence such views, individual and school level characteristics are also controlled. Table 3.5 examines whether the models of various specification will change the sign and significance of the accountability policy.

Table 3.5: Four Ordered Probit Models of Effects on the Teachers' Perception that Teaching Hard is Not a Waste of Time

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>		<u>Model 4</u>	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Age			-0.005	0.008	-0.005	0.008	0.003	0.009
Sqage			0.0001	0.00009	-0.00009	0.00009	0.00002	0.000103
Salary			0.00001***	0.000001	0.00001***	0.000001	0.00001***	0.000001
Totexper			-0.014***	0.005	-0.014***	0.005	-0.016***	0.005
Sqtotexp			0.0002	0.0001	0.0002	0.0001	0.0002*	0.0001
Male			-0.138***	0.021	-0.114***	0.021	-0.048**	0.023
Minority			0.083***	0.029	0.075***	0.029	0.104***	0.035
Unionmem			0.039	0.024	0.040*	0.024	0.043*	0.026
Mathscie					-0.173***	0.035	-0.092**	0.037
Mathalba					0.039	0.044	0.063	0.049
SciBA					-0.003	0.044	0.014	0.045
MA					-0.038*	0.022	-0.027	0.024
Verycomp					-0.048**	0.023	-0.053**	0.025
Certrec					0.020	0.040	-0.031	0.042
Zcontrol							0.141***	0.013
Zinfluen							0.157***	0.013
Zpercept							0.353***	0.016
NewminPER							0.001***	0.000
PerFRLkw							0.001	0.001
Totalenroll							0.00007***	0.00002
Stutearatio							-0.002	0.002
Suburban							0.041*	0.024
Strongacc	-0.050**	0.022	-0.071***	0.022	-0.076***	0.022	-0.089***	0.025

Note: Dependent Variable : Notwasteoftime.

For model 1 and 2, Number of obs=38,375, Population size=2,727,066.5. For model 3, Number of obs = 37,994, Population size = 2,709,439.3. For model 4, Number of obs = 34,109, Population size = 2,440,181.1. For all model, number of strata = 51.

*** means that the coefficient is statistically significant at the 0.01 level. ** means statistical significance at the 0.05 level. * means p-value < 0.10.

Across four ordered probit models, the accountability policy variable has a significant negative effect on the teachers' perception that teaching hard is not waste of time. This implies that teachers working under strong accountability policies are more likely to perceive that trying to teach as their best is waste of time or meaningless. In

addition, the size of coefficient is increasing when we include more teacher characteristic and school characteristic variables.

Table 3.6: Effects on the Teacher's Perception that Teaching Hard is not Waste of Time

	Linear		Ordered Probit		Ordered Logit	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Age	0.00009	0.006	0.003	0.009	0.0073	0.0156
Sqage	0.00003	0.00007	0.00002	0.0001	0.00002	0.0002
Salary	0.000004***	0.000001	0.00001***	0.000001	0.00001***	0.000002
Totexper	-0.012***	0.004	-0.018***	0.005	-0.0340***	0.0091
Sqtotexp	0.0002*	0.0001	0.0003**	0.0001	0.0005**	0.0002
Male	-0.026	0.017	-0.030	0.023	-0.0449	0.0391
Minority	0.071***	0.025	0.098***	0.035	0.1828***	0.0614
Unionmem	0.026	0.019	0.035	0.026	0.0625	0.0440
Mathscie	-0.068**	0.028	-0.087**	0.037	-0.1378**	0.0631
Mathalba	0.055	0.037	0.066	0.049	0.0963	0.0846
SciBA	0.013	0.034	0.018	0.045	0.0039	0.0778
MA	-0.024	0.016	-0.029	0.024	-0.0576	0.0403
Verycomp	-0.038**	0.018	-0.055**	0.025	-0.1045**	0.0418
Certrec	-0.025	0.031	-0.038	0.043	-0.0342	0.0747
Zcontrol	0.094***	0.009	0.146***	0.013	0.2776***	0.0223
Zinfluen	0.105***	0.009	0.147***	0.013	0.2460***	0.0218
Zstudent	0.220***	0.009	0.349***	0.016	0.6243***	0.0261
NewminPER	0.00045	0.00035	0.001**	0.001	0.0019**	0.0009
PerFRLkw	0.00045	0.00038	0.001	0.001	0.0013	0.0010
Totalenroll	0.00005***	0.00001	0.0001***	0.00002	0.0001***	0.00003
Stutearatio	-0.001	0.001	-0.002	0.002	-0.0043	0.0042
Suburban	0.024	0.017	0.038	0.024	0.0538	0.0402
PDdiscipline	0.015	0.016	0.027	0.023	0.0541	0.0391
PDindepth	0.037**	0.017	0.053**	0.023	0.0972**	0.0398
PDmethodte	0.039**	0.018	0.050**	0.025	0.0885**	0.0420
PDassessme	0.031*	0.017	0.056**	0.024	0.1074***	0.0400
PDstandards	0.064***	0.019	0.079***	0.026	0.1401***	0.0437
Strongacc	-0.064***	0.018	-0.097***	0.025	-0.1641***	0.0431
cons	3.147	0.126				

Note: Dependent variable is Notwasteoftime. Number of obs = 34,109, Number of strata = 51, Population size = 2,440,181 *** means that the coefficient is statistically significant at the 0.01 level. ** means statistical significance at the 0.05 level. * means p-value < 0.10.

Table 3.6 displays three full models: a linear OLS regression, ordered probit, and ordered logit model. Table 3.6 indicates that salary is positively correlated with the perception that teaching hard is not waste of time. Teachers receiving more salary are likely to think that doing their best as a teacher is worthwhile. Experience has a U-shaped relationship with the teachers' feeling that it is waste of time to do best. New teachers do not think trying hard is a waste of time, but as their experience increases further, they are more likely to think that it is a waste of time to try hard. However, as experience accumulates at some point further, their perception is evolving, so they think it is not waste of time to try hard.

Math and science teachers or teachers who graduated from very, highly, or mostly competitive colleges are more likely to perceive that it is waste of time to try to do best as a teacher. Interestingly, school enrollment size has a significant positive effect, and teachers' perception that trying to teach as their best is meaningful. In Table 3.4, the school size variable is also significant and shows that teachers working bigger schools are more likely to become a teacher again. Both results would imply that teachers working in bigger schools perceive that teaching is more meaningful. Z_{inflen} , $Z_{control}$, and $Z_{student}$ which capture some organizational circumstances of the schools where teacher are working have all expected and significant effects. Teachers regard trying hard as a teacher as useful when they have more influence on school policies, have more control in classroom instructional contents, and when students do not cause problems at schools. One thing we need to note is that $Z_{climate}$ is dropped in the equation since it includes the item used as the dependent variable.

Most importantly, teachers working in the states with strong accountability

policy are more likely to think that it is waste of time to do best as a teacher, which indicates that teachers lose their intrinsic motivation or self-determination for teaching under strong accountability policies. Thus, we find evidence that strong accountability policy diminishes teachers' intrinsic motivation significantly.

8.3 Do Teachers Become More Dissatisfied When Working Under Strong Accountability Policy?

Teacher satisfaction could be an indicator of teachers' current motivation. However, the SASS questionnaire on the teachers' satisfaction is somewhat site-specific. That is, rather than asking teachers about their general satisfaction with being a teacher, the question asks whether the teacher is satisfied with teaching in their school. Thus, this dependent variable could be a relatively weak proxy variable for teachers' intrinsic motivation.

Like other dependent variables examined in this paper, teacher dissatisfaction could be caused by many factors such as salary and working conditions. Controlling for such possible factors, whether teachers from strong accountability policy are more satisfied with working at their schools is examined. Table 3.7 illustrates four ordered probit models.

Table 3.7: Four Ordered Probit Models on the Effects of Accountability on Teachers' Dissatisfaction

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>		<u>Model 4</u>	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Age			0.006	0.008	0.003	0.009	-0.006	0.009
Sqage			-0.0001	0.0001	-0.0001	0.0001	-0.00001	0.00011
Salary			-0.000003***	0.000001	-0.000004***	0.000001	-0.000003**	0.000001
Totexper			0.002	0.005	0.006	0.005	0.007	0.005
Sqtotexp			-0.0001	0.0001	-0.0002	0.0001	-0.0002	0.0001
Male			0.089***	0.020	0.059***	0.021	0.034	0.023
Minority			0.086***	0.029	0.084***	0.029	-0.042	0.035
Unionmem			0.052**	0.023	0.047**	0.024	0.055**	0.026
Mathscie					0.104***	0.029	0.022	0.030
Mathalba					-0.051	0.040	-0.071	0.045
SciBA					0.002***	0.0004	0.0005	0.0004
MA					0.057***	0.022	0.062***	0.024
Verycomp					0.063***	0.023	0.074***	0.025
Certrec					-0.075*	0.040	-0.001	0.045
Zcontrol							-0.212***	0.013
Zinfluen							-0.294***	0.013
Zpercept							-0.369***	0.015
NewminPER							0.003***	0.001
PerFRLkw							-0.002***	0.001
Totalenroll							-0.0002***	0.00002
Stutearatio							0.005**	0.002
Suburban							-0.001	0.023
Strongacc	0.050**	0.021	0.0446**	0.0219	0.0423*	0.0221	-0.021	0.025

Dependent variable: Notsatisfaction

For model 1 and 2, Number of obs=38,375, Population size=2,727,066.

For model 3, Number of obs = 37,994, Population size = 2,709,439.

For model 4, Number of obs = 34,109, Population size = 2,440,181.

For all model, number of strata = 51. *** means that the coefficient is statistically significant at the 0.01 level. ** means statistical significance at the 0.05 level. * means p-value < 0.10.

Model 1, which is the simplest model, shows that teachers under strong accountability policy are significantly more likely to be dissatisfied. However, when we control more teacher and school level variables, the coefficient size becomes smaller and finally the likelihood that teachers are dissatisfied under accountability policy becomes insignificant.

Table 3.8: Effect on Teachers' Job Dissatisfaction

	<u>Linear Regression</u>		<u>Ordered Probit</u>		<u>Ordered Logit</u>	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Age	-0.002	0.005	-0.005	0.009	-0.009	0.016
Sqage	-0.00001	0.0001	-0.00002	0.0001	-0.00003	0.0002
Salary	-0.000002**	0.000001	-0.000003**	0.000001	-0.00001**	0.000002
Totexper	0.002	0.003	0.007	0.005	0.014	0.009
Sqtotexp	-0.0001	0.0001	-0.0002	0.0001	-0.0003	0.0002
Male	0.016	0.013	0.025	0.023	0.045	0.039
Minority	-0.021	0.021	-0.035	0.035	-0.084	0.060
Unionmem	0.032**	0.015	0.061**	0.026	0.106**	0.044
Mathscie	0.028	0.022	0.033	0.037	0.064	0.065
Mathalba	-0.069**	0.029	-0.088*	0.050	-0.146*	0.088
SciBA	-0.028	0.027	-0.038	0.046	-0.047	0.079
MA	0.040***	0.013	0.067***	0.024	0.103**	0.040
Verycomp	0.044***	0.014	0.078***	0.025	0.123***	0.042
Certrec	0.001	0.027	0.005	0.045	0.007	0.076
Zcontrol	-0.110***	0.007	-0.216***	0.013	-0.395***	0.022
Zinfluen	-0.164***	0.008	-0.286***	0.013	-0.499***	0.023
Zstudent	-0.182***	0.007	-0.366***	0.015	-0.659***	0.026
NewminPER	0.002***	0.0003	0.003***	0.001	0.005***	0.001
PerFRLkw	-0.001***	0.0003	-0.002***	0.001	-0.004***	0.001
Totalenroll	-0.0001***	0.00001	-0.0002***	0.00002	0.000***	0.000
Stutearatio	0.003*	0.002	0.005**	0.003	0.010*	0.006
Suburban	-0.002	0.013	0.003	0.023	0.0001	0.039
PDdiscipline	-0.018	0.013	-0.039*	0.023	-0.077**	0.039
PDindepth	-0.037***	0.014	-0.076***	0.024	-0.134***	0.040
PDmethodte	-0.048***	0.014	-0.072***	0.025	-0.126***	0.042
PDassessme	-0.008	0.014	-0.018	0.024	-0.030	0.041
PDstandards	-0.028*	0.015	-0.049*	0.026	-0.085*	0.044
Strongacc	-0.009	0.015	-0.010	0.025	-0.017	0.043
_cons	1.786***	0.109				

Note: Dependent Variable: Notsatisfaction

Number of obs = 34,109., Number of strata = 51, Population size = 2,440,181.

*** means that the coefficient is statistically significant at the 0.01 level. ** means statistical significance at the 0.05 level. * means p-value < 0.10.

Table 3.8 displays three models which further add professional development variables in the equations. Salary has a negative and significant coefficient, which means that more salary increases teachers' job satisfaction. Age and experience variables are not significant. Teachers who graduated from more selective college or had master degree are less likely to be satisfied. School size variable is significant and has negative sign, which means that bigger school size has a positive effect on teachers' satisfaction.

The coefficients of accountability policy in these three full models are not significant. Thus, accountability policy does not reduce teachers' satisfaction and in contrast to the results of the previous analyses on willingness to become a teacher again, and perception that teaching hard is waste of time. However, as mentioned earlier, the item on satisfaction asks a site-specific question, whether the teacher is generally satisfied with being a teacher at his/her school. Teachers who read this question may respond that whether they like teaching and are satisfied with being a teacher at their schools. This means that their response reflects their preference of their schools rather than overall job satisfaction as a teacher compared to other jobs.

9. Conclusion

We hypothesized that teachers under strong accountability policy receive pressure and such pressure works in a negative way. When teachers perceive the accountability policy and rewards as a controlling mechanism, they will lose their intrinsic motivation. This assumption was tested in this paper using various dependent variables which reflect teachers' intrinsic motivation.

Teachers pressed by strong accountability policy are more likely to report that they would not become a teacher again if they were to start over in college. If teachers are motivated and have obtained sufficient psychic reward from teaching, they will answer that they would become a teacher again. The fact that teachers under strong accountability policy are less likely to become a teacher again compared to those under weak or no accountability policy suggests that accountability policy weakens teachers' intrinsic reward and motivation. This means that at least for some teachers accountability policies take away the pleasure of being a teacher.

The analysis also shows that teachers working in the states with strong accountability policies are more likely to perceive that it is a waste of time to try to do best as a teacher. This result provides strong evidence that accountability policy diminishes teachers' intrinsic motivation to teach. Teachers' satisfaction is not affected by accountability policy significantly. Since the questionnaire asks whether teachers are satisfied with being a teacher at their schools, it would measure whether teachers are satisfied with their school working condition rather than satisfaction as being a teacher in terms of the general sense of job satisfaction.

In sum, we find that bureaucratic accountability policy which does not focus on intrinsic or psychic rewards and only provides pressure and external reward for the desired teaching outcomes may not contribute to the improvement of public education. The empirical evidence reported here supports Sheldon and Biddle's (1998) argument. Teaching becomes a less enjoyable job under strong accountability policy; teachers would not become a teacher again if they were to start over, and they feel dissipated about trying to do their best as a teacher.

Since the sample represents almost the whole national public school teacher population, the results provide very strong evidence of the effect of current accountability policy on teachers. The SASS survey includes many questions related to organizational conditions and school climates, while it includes few teachers' motivation or other psychological measures. Policy analysts have paid extensive attention to organizational factors and less attention to factors affecting teachers' psychological disposition and motivation. In the future, more research and survey on the psychological effect of education policy on teachers, including the design of policy which can boost teachers' self-determination or intrinsic motivation, would be needed.

APPENDIX

1. Questionnaires Used as Dependents Measures

- I sometimes feel it is waste of time to try to do my best as a teacher

Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
[1]	[2]	[3]	[4]

- I generally satisfied with being a teacher at this school.

Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
[1]	[2]	[3]	[4]

- If you could go back to your college days and start over again, would you become a teacher or not?

- [1] Certainly would become a teacher
- [2] Probably would become a teacher
- [3] Chances about even for and against
- [4] Probably would not become a teacher
- [5] Certainly would not become a teacher

2. Table: Two Groups by the Intensity of Accountability in 1999-2000

<u>States with Weak Accountability</u>		<u>States with Strong Accountability</u>	
State	Index	State	Index
Alaska	1	Alabama	4
Arizona	2	California	4
Arkansas	1	Florida	5
Colorado	1	Kentucky	4
Connecticut	1	Maryland	4
Delaware	1	New Jersey	5
Georgia	2	New Mexico	5
Hawaii	1	New York	5
Idaho	1	North Carolina	5
Illinois	2.5	Texas	5
Indiana	3	Average Index Score	4.6
Iowa	0		
Kansas	1		

Louisiana	3
Maine	1
Massachusetts	2
Michigan	1
Minnesota	2
Mississippi	3
Missouri	1.5
Montana	1
Nebraska	0
Nevada	1.5
New Hampshire	1
North Dakota	1
Ohio	3
Oklahoma	1
Oregon	2.5
Pennsylvania	1
Rhode Island	1
South Carolina	3
South Dakota	1
Tennessee	1.5
Utah	1
Vermont	1
Virginia	2
Washington	1
West Virginia	3.5
Wisconsin	2
Wyoming	1
Average of Index Score	1.5

*Accountability Index was obtained from Carnoy and Loeb (2002)

REFERENCES

- Abelmann, C. & Elmore, R. with Even, J, Kenyon, S, & Marshall, J. (1999), When accountability knocks, will anyone answer? Consortium for Policy Research in Education *Research Report Series, RR-42*. Graduate School of Education, University of Pennsylvania.
- Amabile, T.M. (1979). Effects of external evaluations on artistic creativity. *Journal of Personality and Social Psychology*, 37, 221-233.
- Amabile, T.M., Dejong, W., & Lepper, M.R. (1976). Effects of externally imposed deadlines on subsequent intrinsic motivation. *Journal of Personality and Social Psychology*, 34, 92-98.
- Barns, C.A. (2002). *Standards reform in high-poverty schools : managing conflict and building capacity*. New York, Teachers College Press.
- Benware, C., & Deci, E.L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, 21, 755-765
- Bess, J.L. (1997). *Teaching well and liking it : motivating faculty to teach effectively*. Baltimore : Johns Hopkins University Press.
- Carnoy, M. and Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24, 305-331.
- Cohen, D.K. (1990). A revolution in one classroom: the case of Mrs. Oublier, *Educational Evaluation and Policy Analysis*, Fall 1999, v.12,n3, pp.311-329.
- Deci, E.L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105-115.
- Deci, E.L. (1995). *Why We Do What We Do: the Dynamics of Personal Autonomy*. New York : Putnam's Sons.
- Deci, E.L. & Ryan R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, Plenum.
- Deci, E.L., Schwartz, A.J., Sheinman, L., & Ryan, R.M. (1981). An instrument to assess adults' orientations toward control versus autonomy with children: Reflections on intrinsic motivation and perceived competence. *Journal of Educational Psychology*, 73, 642-650.
- Deci, E.L., Spiegel, N.H., Ryan, R.M., Koestner, R., & Kauffman, M. (1982). The effects of performance standards on teaching styles: The behavior of controlling teachers. *Journal of Educational Psychology*, 74, 852-859
- Flink, C., Boggiano, A. K., and Barrett, M. (1990). Controlling teaching strategies:

Undermining children's self-determination and performance. *Journal of Personality and Social Psychology*, 59, 916-924.

Feiman-Nemser and Floden. (1986). Wittrock, M.C. (Ed.) *Handbook of research on teaching*. New York : Macmillan ; London : Collier Macmillan

Garbarino, J. (1975). The impact of anticipated reward upon cross-aged tutoring. *Journal of Personality and Social Psychology*, 32, 421-428.

Goddard, R.D., Hoy, W.K., and Hoy, A.W. (2000). Collective teacher efficacy: It's meaning, measure, and impact on student achievement. *American Educational Research Journal*, 37(2), 479-507.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). Improving student achievement: What NAEP test scores tell us. Santa Monica, CA: RAND Corporation. Available at <http://www.rand.org/publications/MR/MR924>

Grolnick, W. S. & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*. 52, 890-898.

Harris, D.C. (2002). Lowering the bar or moving the target: A wage decomposition of Michigan's charter and traditional public school teacher. Working Paper. The Education Policy Center, Michigan State University. Available at www.epc.msu.edu

Igersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*. Vol.38, no. 3, pp.499-534.

Kreps, David. (1997). Intrinsic motivation and extrinsic incentives. *American Economic Review*, 87, 359-364

Ladd, H.F. (Ed.) (1996). *Holding Schools Accountable*. Brookings.

Lepper, M.R. & Greene, D. (1975). Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology*, 31, 479-486.

Lotie, D. (2002). *Schoolteacher*. Second Edition. The University of Chicago Press.

Miserandino, M. (1996). Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. *Journal of Educational Psychology*, 88, 203-214.

Murnane, R., Singer, J.D., Willett, J. B., and Kemple, J.J. (1991). *Who will teach?: Policies that matter*. Harvard University Press.

Newmann, F.M., King, M.B., & Rigdon, M., (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review*, V67, N1. pp 41-69.

Pintrich, D. & Schunk, H. (2002). *Motivation in Education: Theory, Research, and Applications*. Upper Saddle River, N.J.: Merrill.

Reeve, J., Bolt, E., and Cai, Y.(1999). Autonomy-Supportive teachers: How they teach and motivate students. *Journal of Educational Psychology*, 91, 537-548.

Roderick, M., Jacob, B.A., & Bryk, A.S. (2002), The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis* (2002), V24, N4. pp.333-357

Ryan, R.M. & Deci, E.L. (2002). *Handbook of Self-Determination Research*. Rochester, NY : University of Rochester Press.

Sheldon, K.M. and Biddle, B.J. (1998). Standards, accountability, and school reform: Perils and pitfalls. *Teachers College Record*, 100(1), 164-180.

Tschannen-Moran, M., Hoy, A.W. and How, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Review*, 68 (2), 202-248.

Tyack, D. & Cuban, L. (1995) *Tinkering Toward Utopia: A Century of Public School Reform*. Cambridge, Mass. : Harvard University Press.

Vallerand, R.J. and Fortier, M.S. (1997). Self-Determination and persistence in a real-life setting: Toward a motivational model of high school dropout. *Journal of Personality and Social Psychology*, 72, 1161-1176.

Wolfe, E. W., Ray, L. M., & Harris, D. C. (in press). A Rasch analysis of three measures of teacher perception. *Educational and Psychological Measurement*.

Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02504 0407

