

School accountability: mathematics teachers struggling with change

Samuel Obara
Texas State University

Abstract

In this period of accountability advocated by the No Child Left Behind Act of 2001, testing has been selected as a primary means of measuring the performance of schools. The State of Georgia is in the process of replacing its old curriculum - Georgia's Quality Core Curriculum (QCC) with a new curriculum - Georgia Performance Standards (GPS) to increase student achievement to meet NCLB initiatives. The accountability system which is part of the implementation of the GPS give student a pretest at the beginning of the nine weeks and posttest at the end of nine week to measure how much far they have grown. This article reports a qualitative case study of three teachers and their mathematics coach at a urban middle school in Georgia. The study investigated their views about the district accountability system during the first year of the implementation of the GPS by using standard-based curriculum materials. Data were gathered through interviews and observations. The analysis of data demonstrated that the three teachers and the mathematics coach were strongly opposed to the system because it did not consider a number of factors that affected student achievement. Nonetheless, the participants also identified advantages of the system. My analysis of the data also identified several areas of concern regarding the implementation of accountability systems.

Introduction

Teachers in the U.S. face new federal and state governmental policies that require them to be accountable for their students' performances. Reform seems to be part and parcel of an educational panorama in which assessment and accountability have played a pivotal role (Linn, 2000). As a matter of fact, there is general agreement within the American public that schools should be held accountable for student performance (Johnson & Immerwahr, 1995). Therefore, school districts and states have a right and responsibility to ensure students are doing what they are supposed to do. This responsibility implementation and how it is addressed at different local and state levels is an issue of contention. But there is a general consensus that the issue of delivering education is not to be shouldered by an individual school. Rather, a concerted effort of the whole system that supports the school's operation particularly with regard to funding should be involved with this responsibility (Newmann *et al.*, 1997).

The purpose of this study was to investigate the perceptions of teachers with respect to the accountability system in place in their district. Specifically, this study sought to answer the following questions: (a) how does a mathematics coach and the three teachers she works with perceive the accountability system in the school district where they work? (b) What does accountability mean in the eyes of these teachers? and (c) In what ways do teachers feel that accountability contributes to student achievement?

Accountability has been a force within education for some time. Issues such as the high dropout rate of students from disadvantaged backgrounds and low academic achievement among children (National Commission on Excellence in Education., 1983), have prompted external agents (districts, states, parents, etc.) to become concerned with students' performance in K-12 settings. In response, federal and state governments have established policies geared towards improving student performance. The main question that Newmann (1997) poses is: Will creating such policies of accountability lead to improvement in student performance?

Researchers (Newmann, King, & Rigdon, 1997) have identified at least four aspects of comprehensive school accountability systems:

- Information of performance (tests)
- How judging is done to measure/determine student achievement of within a school and between schools.
- The consequences attached to performance student
- The agency that gets the performance information, makes judgment, and distributes rewards and sanctions

One should ask, how do these aspects improve performance? One can expect "that teachers will try harder and become more effective in meeting goals for student performance when the goals are clear, when information on the degree of success is available, and when there are real incentive to meet the goals" (Newmann, King, & Rigdon, 1997, p. 43). This in theory seems to offer a very compelling explanation for better student performance. But, there are three issues that have not yet been resolved:

- Implementation issues: There is no question about the need to hold schools accountable. However, the means of implementing a strong accountability system is controversial. For instance, it is difficult to choose standards that try to reduce the drop-out rate and, at the same time, increase the graduation rate.

- Organizational capacity: Even though we assume external accountability is implemented—there is a need for schools to have technical resources available that support instruction.
- Internal accountability. In addition to strong external accountability, there is a need to promote internal accountability that creates a cohesive environment. Internal accountability promotes peer pressure that encourages standards to be implemented. The challenge is how to implement external accountability that promotes internal accountability.

Newmann, King and Rigdon (1997) conducted a study that investigated school accountability and school improvement by collecting data from observation, interviews and surveys of teachers and students from 24 schools. The authors found that:

Strong accountability was rare; that organizational capacity was not related to accountability; that schools with strong external accountability tended to have low organizational capacity; and that strong internal accountability tended to reinforce a school's organizational capacity (p. 41).

What this suggests, precisely, is that “efforts of external agencies to strengthen accountability should pay increased attention to stimulating the kind of internal accountability that we found linked to organizational capacity” (p. 63).

In the analysis by Hanushek and Raymond (2005) of the differences among the states in student performance on the National Assessment of Educational Progress (NAEP)—findings indicated that the accountability system had a positive impact on student achievement. The study showed that the achievement gap between White and Hispanic narrowed, but the same was not evidenced between Blacks and Whites. The writers concluded that there is no one answer for increasing student performance. Fuller and Johnson (2001) arrived at similar findings.

One older accountability system initiative with fascinating findings is that of Kalamazoo School System, Michigan in the early 1970s (Kalamazoo schools, 1974). This very complex system used “standardized achievement tests administered in both fall and spring, criterion-referenced tests developed by teachers, performance objectives, teacher peer ratings, student ratings of teachers, parent rating of teachers, and teacher self-ratings” (Patton, 2002, p. 17). This system drew a lot of attention across the nation. The American School Board Journal asserted: “...take it from Kalamazoo: a comprehensive, performance-based system of evaluation and accountability can work” (Kalamazoo schools, 1974, p. 32).

The teachers in the Kalamazoo system, however, did not share in this optimism. Responding to a confidential questionnaire, these teachers indicated that they were very opposed to the system, it did not help them be more effective, and it had the effect of demoralizing them. Such findings are very important to note, especially as school districts around the US are trying to comply with the No Child Left Behind Act (2001). Knowing what teachers are thinking about the system is important in designing policy initiatives focused on improving student's achievement.

On the other hand, a study by Powers, (2003) explored factors that contribute to school performance for two urban school districts using the California mandated Academic Performance Index (API). The findings in this study indicated that policy makers need to think very carefully in terms of what to use to measure school

performance. Factors such as socioeconomic status, teacher training, and experience played a pivotal role in school performance. These factors are often overlooked when districts are evaluating teachers in the light of an accountability system. In this study, I investigated what teachers thought about the district accountability system and how that might help district and state policy makers.

Method - Participants and Data collection

The state of Georgia is in the process of phasing out its old curriculum, The Quality Core Curriculum Standards (QCC) (Georgia Department of Education), and replacing it with the Georgia Performance Standards (GPS) (Georgia Department of Education). The main goal of the GPS initiative is to improve student achievement. To achieve this, the state, district, and schools have initiated this accountability system in order to comply with the No Child Left Behind act (2001).

Data collection took place at Tabaka middle school, an urban school, which was praised by the Bush administration as a model that has implemented the No Child, Left behind act. The school population is 21% White, 3% Asian, 49% Hispanic, 2% Multi-racial and 25% African American. Eighty percent of the student body qualify for free or reduced meals. The adoption of the GPS is being felt at district and school levels. To help teachers in this implementation process, Tabaka Middle School hired mathematics coaches as in-school professional development personnel. In addition, the school selected the Connected Mathematics Project (CMP) materials, which were judged to be a middle school mathematics curriculum consistent with the GPS.

The project follows a case study design (Stake, 1995; Yin, 2003a, 2003b), and the case presented in this article is part of a larger study examining the process and impact of implementing the Georgia Performance Standards (GPS) in Mathematics using the Connected Mathematics Projects (CMP) materials. This case study focused on observational and interview data from the first-year implementation of the GPS in the context of accountability systems. The observation focused on five in-school professional meetings run by the mathematics coach, Hellen, and attended by three mathematics teachers, Drusilla, Loice, and Alice. Hellen, who has a PhD in mathematics education, had taught at this school for two years and had several years of teaching experience at the middle and high school levels. Nyachoka had 20 years of teaching experience in various schools and had been at this school for 5 years. Loice joined the teaching profession 2 years ago after a 17-year career as a secretary. Alice, on the other hand, had worked at this school for 8 years, the longest among the four participants.

During these meetings, I observed comments teachers made in connection to the district accountability system. These comments were recorded in field notes which “contain the observer’s own feelings, reactions to the experience, and reflection about the personal meaning and significance of what has been observed” (Patton, 2002, p. 303). I also interviewed each of the four participants twice—at the beginning of the semester and at the end. Each interview was recorded and transcribed. These interviews investigated their opinion about the school accountability system, particularly during the implementation process of the Georgia Performance Standards (GPS). Whereas the same interview protocol was used for the three teachers, the protocol for the mathematics coach was modified so as to specifically address her role.

Data Analysis

The case study qualitative approach is a particular way of collecting, organizing and analyzing data as outlined by (Stake, 1995) and was chosen to enable the researcher to “maximize what we can learn” (p. 4). In this study, the analysis included: (a) transcribing and reading transcripts in the process of data collection; and (b) Coding data, grouping it into categories, and connecting them. “Further data collection... is based on emerging concepts. These concepts are developed through constant comparison with additional data” (Dey, 1999, p. 1). Glaser and Strauss (1967) describe four components of constant comparison: (a) comparing incidents applicable to each category, (b) integrating categories and their properties, (c) delimiting the theory, and (d) writing the theory. (p. 105). In this study, data analysis and data collection occurred simultaneously, helping the researcher to evaluate what was observed and informing him as to what aspect to focus on or make connections with in order to find themes in participants responses (Ezzy, 2002).

Results

The accountability system in the district is part of the statewide initiative, which is part of the implementation of the GPS and has the goal of improving student achievement. Teachers reported both positive and negative aspects of the accountability system. But before discussing that, the participants were first asked to describe the district accountability system. Hellen noted:

We implement a pre-post test system here in Tabaka City Schools that is recognized by the governor of our state as a model of what other schools should be doing. Other districts also use pre-post tests every nine weeks, the same as we do, and the difference, from what I understand it to be, is the way we use the information that we get. ... We use it here to support instruction and celebrate excellence or achievement or something.

Celebrate excellence. That is what we use it for. Not a “gotcha.”

Although that was what Hellen was told when she was hired, it was not the culture she has experienced while working at the school.

The teachers also described the pre-post test system of which they were not willing to discuss after several attempts if it was a good test of mathematics learning. Alice described how the system measured student achievement:

We have a pretest and a post test and we give the pretest on the objectives that we will be covering for the nine weeks, before they are taught at the beginning to see what the students know and then at end of the nine weeks they are give a very similar test over the same objectives to see how far they have grown.

Loice, however, focused on the public nature of system:

... the pretest and the post test. And it is posted in the school and on the web sites so parents can go up and look how their children are doing, they can look per teacher too so they can kind of get a feel for things.

Drusilla declined to describe the district accountability system. It was not clear whether she did not understand it or thought it was a sensitive issue to tackle.

When describing the negative aspects of accountability system, Hellen, the mathematics coach, noted that teachers are not comfortable with pre- and post-testing, especially when the test scores are posted in the hallways for all to see and make comparisons of teachers where graphs hang in full view of passing students and teachers. Each bears a teacher's name and shows a growth curve, indicating plainly whether students in a class are making progress. Furthermore, teachers whose classes had a 40% gain in student performance receive an e-mail from the Superintendent, while teachers whose students had only a 20% gain do not. In Hellen's view, this conveys a message that "We are comparing." In addition, posting results on the wall contradicts the notion of celebrating excellence. She also noted, however, that it is important for teachers to realize the accountability system is not going away. Because the teachers accepted jobs in the school district, choosing to work at the school this year in which they knew the pre- and posttests were used, they have no choice but to accept it and make the best of it.

In her interview, Drusilla did not view the testing and posting results in the hallway as a negative consequence but rather as one way to improve her teaching. She saw the pre and post testing as one a way of improving test scores. In her words:

I'm hoping test scores will improve. I am hoping to see test scores skyrocket is what I am hoping to see. I know I feel sure that is what the system wants to see; is that test scores will go up...that they will not go down, I think that they are going to hold us accountable for that—this program will boost test scores.

Drusilla noted that by being put under a microscope is a good thing. In this process, she believes, bad teachers will be weeded out. According to her, bad teachers pose a big challenge not only when new reforms are initiated, but also setting a bad example for the incoming teachers. Hopefully, the best teachers, who actually love teaching, will remain and therefore be a great plus in the implementation process of CMP materials.

However, the teachers' discussion during in-school professional development and the math coach's responses to inquiries from researcher showed that Drusilla's statements were contradictory. While she publicly praised the system, she was critical of it when not on camera or knowing her statements would not be documented. From my observation, Nyachoka was uncomfortable being judged. The posting of pre and post test scores sometimes caused her to contemplate quitting the job.

Maraa noted that it is not a bad idea for parents, teachers, and students to access pre and post test scores as long as it is not used a way of comparing performances among teachers. As Loice puts it, "It is good for the parents to know, it is good for the teachers to know, as long as they are not evaluating the teacher". On the other hand, Loice observed that the act of posting scores in the hallway and Internet could affect the self-esteem of those teachers with lower student test scores. This can lead to low motivation and even poorer grades. As Loice puts it, "if I look out there and see somebody's score, and my test scores are a lot lower them, my self-esteem will go down as a teacher. But I don't mind people seeing my scores at all".

Poor test scores, according to Loice, may affect the way teachers are teaching. “This can be a good or bad thing in that if teachers are trying new innovative ways and get poor results, the teacher may resort to the old ways that worked”. From my observation, the three participants resorted to old ways that worked for them such as drilling students and supplementing a lot from the traditional textbooks they had used the previous year—hence diluting the CMP materials. But when asked, the teachers said they had implemented the CMP as intended.

Teachers were under a lot of pressure. For instance, by the time the three teachers recommended to the coach to supplement materials to CMP because of some issuers for example reading; Drusilla noted “now I can do what I enjoy doing—me being the teacher and not the other way around and using materials that can improve my kids scores”. This statement was supported by observation of her classes where she used drilling method as a method of instruction.

During observation of the in-school professional development of teachers, Loice did not see the act of posting the results as a good thing. She resented it and wished that the system was different. However, she also noted that teachers with poor scores may want to seek advice from those teachers with high student scores, hence fostering teacher collaboration. From my observation and interaction with the participants, that did not happen. There was more of a competition among teachers for better grades instead of collaboration so that all teachers can become better.

Alice had a view similar to Hellen’s in that the administration made it clear that the test scores have nothing to do with their jobs and, therefore, they should not worry about it. But there were instances in the school that suggest otherwise. As Alice put it:

And they tell us that your scores have nothing to do with your job or anything. However, I know of two teachers personally who have been called to meet with school district offices and asked why their scores did not go up. But they say it has nothing to do with the scores, do not worry about this, do not worry about that, but they have called them in and said what’s going on here? Are you not teaching them? Why are they not learning this? Like I said, the administrator is not in the classroom. They don’t know what kind of class they have, they don’t know if the students can’t behave. You know they do not know anything about it but they just put all of this into one big nutshell and say; then you must not be doing your job.

Alice wondered why students were not called in also to explore what they think about why they are not learning. She thought that the accountability system was not put in place as a way to inform parents about how children are doing in school but something to show off the school’s accomplishments. This was because a good number of parents, who happen to be Mexican do not know how to read, never come to school to see their children’s teachers, and maybe have no access to the Internet to check the postings of the pre and post tests. In contrast, Hellen thought that a good number of parents have responded to the accountability system by coming to ask questions whenever they have concerns.

Another negative element that Hellen identified was that teachers were being judged without taking into consideration the kind of kids each teacher taught. In her words:

And I think people who teach kids who have a lot of support at home or have some natural aptitude to begin with are going to see more gain than the teacher whose kids are struggling with poverty and whose home life is not necessarily supportive. And I think a 20% gain in that classroom may say more about the instruction than a 40% gain in another class. It is a one dimensional look at a teacher's class room.

Whereas Drusilla did not identify that as negative element, Loice had a similar view as Hellen. She asserted, "demographically, 49% of students are Hispanic. Most of the Hispanic students are migrants to the United States, which means English is their second language". Loice emphasized that this demographic background of students is precisely what the accountability system does not take into account. Some students have limited access to English as a language, which definitely affects their test scores.

Loice gave an example of a class with 90% Hispanics who were second language learners—there is a great chance their scores will be lower than other students for whom English is their first language. In such instances, Loice observed that parents will not see that there were 90% Hispanics in that class but rather "see the different teachers". Loice stressed that this does not mean that Hispanics are non-performers but rather the issue of language is the factor that contributes to their performance on the tests even though they were taught the same way. She noted that, in here experience, students like this needed closer examination of their pre test and post test scores "but nobody could see that they started out at 4% and went to 70%. All they could see is 70; "Oh they didn't meet their goal, they only went up to 70." But from 4 to 70% was a big jump". To her, it is very discouraging because that increase from 4% to 70% may mean more that an increase from 70% to 90%.

Alice also raised the issue of language barrier but stressed that on top of that, the students read far below their grade level. This issue has been complicated with the implementation of CMP materials, which have more word problems that pose a big challenge to second language learners. Information concerning the kinds of students these teachers have is concealed from the public, "it does not tell what level the student is on. It does not tell if they can't read or speak English. I personally don't like them being posted everywhere with your name on them". Some classes have a co-teacher, whereas others do not, and such information is not revealed. When observed, Loice had a co-teacher who could teach some lessons. Also, the number of students in the respective classes is not revealed in the displayed scores, "like last year I had like 30 students whereas other teachers will have like 15. You know there is a big difference going from 15 to 30 and the perception of it on the wall will be; this is a good teacher, this is a bad teacher". On the other hand, Alice said, "also the students may not test well. They may have test anxiety—also it is multiple choice and they can guess. Nothing else, they have 25% of getting it right. That is a weakness to me. Just guess and get it right".

These are some of the issues where Alice thought the accountability system fails to educate the public. Hellen was also concerned about the accuracy of measuring what students know and don't know.

Well there are 25 questions [during the first nine weeks' 25 objectives]. So, one question for one objective? I mean, a kid could guess and get it right and we would think the kid knows how to—knows the material in that objective. Or, they get it wrong and we assume they don't know it. So

I think that is a pretty shallow way of getting at what students know and don't know.

A sample of the pre and post test that I reviewed underscored the concerns of the teachers about the validity of the test. Hellen, the math coach, constructed them and were multiple-choice questions based on the 25 objectives.

Hellen also identified that teachers stop teaching three or four days before the post-test to get the students ready for the test. Teachers keep on guessing what might appear in the test and make students practice over and over again questions likely to appear in the test.

As a result of this, valuable instruction time is lost. I understand they are worried about the list of standards that were handed out and told to teach all these and I am going to come in here and test you at the end to make sure you did.

Drusilla did not cite this as a negative factor did spend one to two class periods going over what was expected to be on the post test. She showed the kids tricks on how to answer questions and gave several examples of what to expect. This practice was corroborated in a conversation during the in-school professional development, when Drusilla disagreed with Loice about what they expected on the post test. Hellen, whose responsibility was to design the pre and post test, was seeking feedback on what they wanted to be in the test. Drusilla wanted some aspect be left out because she had skipped that section whereas Loice wanted it to be included. The entire discussion was focused on getting kids to score well on the post test in order to fulfill the expectation set for them. In every class I observed on a day before the post test, each of the teachers was busy drilling their kids and giving them tricks on how to do better. But none of the teachers cited this as a negative aspect of the accountability system. Rather, it seemed to be a way to prepare kids for a test.

One issue that Loice raised was that teachers are tempted to tell their kids that the pre test does not count and, therefore, not to put forth much effort. With the kids scoring as low as possible on the pre test, and giving their best effort on the post test, students could maximize the percentage gain, making the teacher look good. In my observation, I noticed that during the pre test, there was no seriousness on the part of the teachers to create proper examination conditions. Teachers just came in and distributed the exam paper and, sometimes, left the room. Their behavior was totally opposite for the post test.

In spite of the negative factors discussed, the participants also talked about the positive aspects that the district accountability system brings. Hellen noted that one positive factor of the accountability system is that it informs teaching. The pre test and post test taken at the beginning and at the end of each nine weeks help teachers with their instructional decisions. This happens with the aid of Thinkgate[®] software that the school district uses to analyze the scores. The software breaks down the scores, showing how students did on each item. In Hellen's words:

You can see how many students missed a given item and you can see not only how many students missed it, but what the wrong answers that they put were. So you can know if there is a misconception. And then teachers

get those scores and use those scores to make instructional decisions. Like if 90% of your students already know how to find the greatest common factor in the pre test, then you wouldn't spend as much time, focus, on that idea in the classroom as you would on something like students adding fractions with non-common denominations, where only 20% got it right. So that is how we use that kind of data to inform our teaching.

Drusilla had a similar view that the pre test and post test would help her improve “my teaching and I can help my students be better students” but declined to explain when asked how. Loice noted that the pre test and post test will inform their instruction but stressed that, for that to happen, more emphasis needs to be placed on the way the pre test is administered, ensuring reliability. Alice also reiterated that she has found the accountability system helpful in that it helps teachers with their instructional planning. Teachers can use pre-tests to know which areas students are weak in and which ones they can ignore. Alice observed that:

I do like seeing wide objectives that are going to be taught exposing them with the pretest. Then we can spend more time on the ones that they don't know. They know something—if 98% of them get something right, we do not have to spend a lot of time on it. We can just touch on it and go onto what they are really having trouble with. I think having a pre test to build from; I think it is a really good idea.

Another factor that Hellen pointed out as positive is that teachers need to know that at the end of nine weeks they are held accountable for teaching the material to the students. She thinks there is nothing wrong with telling teachers that they will be held accountable, “they've a responsibility to our students to get them ready otherwise they decided to work this year by signing the contract”. Nyachoka noted that, when she accepted a job to work at Tabaka Middle School, she knew what she was getting into. Reinforcing this point, she states, “You have a job to do, you are told what that job is so you accept the responsibility of that job and you do it to the very best of your ability and I think that is what we have been told here”. Loice and Alice also agreed that they accepted and signed the year's contract, but that does not mean the school district can do anything that they want. They continued to say that it is important to consider teachers' views as well.

Alice alone said that the program of choice that existed in the school was a positive factor of the accountability system. The program perceived as low performing was doing better and therefore conveying the message to the rest of the students that all students could achieve academically, no matter the program they were in. After the school did away with tracking four years ago, the school introduced three programs of choice: Classical studies, Earthquest, and Humanities. Students joined these programs voluntarily but expectations of the three programs are presumed to vary. As Alice puts it:

I do like and this is a little contradictory to what I was saying before. They have put them up by programs and we have three programs of choices of classes: Earthquest, Classical, and Humanities. And I am in Earthquest and our students from day one when they have graded these programs — the Earthquest program is seen as the dumb program. For all the dumb students, that is the students' perception of it when they get here, but what I'd love to do is, this sounds bad, but I would like to show them

Earthquest scores compared to classical compared to humanities. Because we have always come out on top, so that gives them ego boost.

With all said about positive and negative factors concerning the accountability system, the four participants discussed what accountability means to the implementation of the GPS. Hellen noted it was going to have a great impact on teachers for the GPS in terms of tending to teach to the test, but the state so far has not supplied any test samples. As Hellen put it:

the big issue is that the state hasn't decided what the end of the year test is going to look like. We do not have any sample questions. We don't know what that is going to look like. They haven't told us yet. They just say it won't be criterion reference. It will be performance based. We don't really know how we are going to do that in multiple-choice questions. But the test will be multiple choice, and we will get back to you later on that. So we are, we are stuck. Maybe it is a good thing. We can't possibly teach to a test because we have no idea what the test is going to look like. I did the best I could when I made our pretest for these nine weeks to try to find questions that were more conceptual that seemed more task oriented or performance oriented but yet were multiple choice.

Loice did not see how accountability will affect the implementation of the GPS but felt that, by virtue of using CMP materials, the pre-test and post test scores will go up:

I think the post-tests are going to be a lot higher with this material than it would be without it. Only because it goes so in depth in with taking a whole bunch of things and culminating them together into one and they are getting it, and with all the manipulative that we are going to be using, they are just getting it. You know that I don't know how else to say it, but they are getting it and the light bulbs are coming on. I think this is going to make the post-test be a lot bigger than if we were not using it.

Alice had a similar view, saying she did not see how the accountability is going to affect the implementation process, "because we are already doing GPS, and CMP is still going to cover these standards that we must cover". She noted that although there is a lot to do with accountability system, she does not think it will affect the implementation process in any way. Drusilla also agreed with Alice saying that accountability will not play any role in the implementation of the GPS.

Discussion

The accountability system at Tabaka Middle School has received a great deal of attention. The school was recognized by the Bush administration as a model of the No Child Left Behind initiative. Despite this recognition, there are issues uncovered by this study that are worth noting. First, it seems that the issue of pretests and posttests made sense to the teachers as a way helping them make instructional decisions. Teachers use the pretest to know how much the children knows and don't know to help them decide what to cover or not and at what depth. Teachers have the notion that they are being evaluated and consequences are not loudly stated but they exist. Due to this, the efficacy of the pretest needs to be more scrutinized. First and foremost, the students know that pretest scores do not count towards their grade and, therefore, do not pay any attention to

the test. Therefore the pretest is not as useful as it should be. Because the goal of the system is to maximize student gain and teachers with higher student gain are rewarded, some teachers felt that some of their colleagues conveyed a message to their students that the pretest does not count with the aim of eliciting low scores on the pretest and very high on the posttest to maximize test gain. These findings are very troubling in that teachers were trying to maximize test gains in any way they could and making the pre-test results useless in supporting instruction.

There was no evidence that teachers used data generated from the pretest to inform their instruction. On the other hand, there were issues with the posttest and how accurately it measured what students had learned. The posttest covered 25 objectives and there were 25 questions from those objectives. On average, there is only one question per objective—how much could be measured here?

The school district intended the pre-test and post-test to be used to celebrate excellence. However, that seemed not to be the way the three teachers and the mathematics coach interpreted it. The teachers interpreted it as a way of comparing them, as evidenced by the superintendent sending an email to teachers with higher score gains and by the posting of pre- and posttest scores of each class with the teacher's name underneath on the front wall of the school. Posting teachers' names with their students' test scores was demoralizing to some teachers. It seemed that the teachers did not object to having their names in the hallway as long as it was not used as a way of evaluating their performance. Similar findings were arrived at by the Kalamazoo (1974) study.

One issue that this study cannot address is the impact the accountability system had on the achievement of students and to what extent the tests reflected genuine learning of mathematics. For that, more research needs to be done, but according to a study done by Newmann, King and Rigdon (1996),

external accountability alone offers no assurance that a school faculty will have adequate technical knowledge and skills, sufficient authority to deploy resources wisely, or shared commitment to a clear purpose for student learning. Theoretically, both accountability and organizational capacity are required for high performance. But it is also possible that strong accountability can be achieved within a school community, without prescriptive mandates from a district or state. (p. 62)

The assumption that is brought forth under the external accountability is that teachers will work hard by being held accountable of student performance, but “in reality there are a large number of factors over which they have very little control but which impact negatively on student achievement” (Watson & Supovitz, 2001). For this study at Tabaka middle school, a number of such factors were noted. One factor was language barrier and reading—49% of students in this school came from parents of recent immigrants from Mexico and South America for whom English was a second language. It was noted that some teachers ended up having a majority of those children with problems in reading. This class composition was compounded with the use of CMP materials that required a lot of reading. It was inferred by the participants that children—even the native speakers—were struggling with these new materials, but the situation was more serious for second language learners. This factor was overlooked when posting the results. As a result, individual teachers were blamed when the parents and school district did not have enough data to justify individual teacher's performance.

Secondly, some classes had co-teachers whereas others did not. It was indicated that having co-teacher may have had some effect on the outcome of the students' scores. When it came to posting results, however, only the main teacher was listed.

Third, class size may have an impact on teacher-student relationships in classroom instruction. It was noted that some classes were smaller than others depending on the program of choice. As a result, some teachers may have had more children than others, contributing to the overall gain that the school district is looking at when assessing the performance of teachers.

Fourth, the socioeconomic status of the students was not taken into account. For students who are struggling in poverty with no support at home, a small gain means a lot than for those who are not in a similar situation. As Powers, (2003) points out that factors such as socioeconomic status, teacher training, and experience played a pivotal role in school performance. All of these factors were not highlighted when the district held teachers accountable.

Elmore and Fuhrman (2001a) argue that, before teachers are held accountable for a given outcome, administrators and policy makers need to make sure teachers have all they need to meet the outcomes. The district had initiated a number of initiatives to help teachers meet these outcomes, but the efficacy of the initiatives needs to be examined further. The district employed the mathematics coach to be the in-school professional development personnel to help teachers with their instructional needs and issues related with instruction. Factors such as reading were discussed, but little could be done to help a problem that needed more resources and time than what the teachers could possibly provide.

The mathematics coach was very concerned about the accountability system and had even requested the school district to delay it for 2 years, but the district would not listen. She thought having the implementation of the GPS in the context of the accountability system was too much for teachers to bear.

Tabaka middle school basically emphasized external accountability and neglected internal accountability. There was no evidence in this school that indicated the existence of internal accountability. As some researchers note, internal accountability precedes external accountability—for instance, Elmore (2001b) argues that:

A school's ability to respond to any form of external performance-based accountability is determined by the degree to which individuals share common values and understandings about such matters as what they expect of students academically, what constitutes good instructional practice, who is responsible for student learning, and how individual students and teachers account for their work and learning. (p. 11)

More needs to be done in this school in terms of promoting internal accountability. If this can be accomplished, then perhaps the negative impact of external accountability will be minimized. Strengthening the school accountability system to reflect the GPS implementation at Tabaka Middle School should have been carefully thought out and should have incorporated input from the teachers. All four participants had similar concerns about the accountability system, but the case of Drusilla was unique. She never clearly stated her stand on the accountability system in that she kept changing her stand, more so when she was being interviewed and recorded.

So the question becomes: How can the school personnel promote and improve their internal accountability? Some promising ideas:-

- Carefully diagnose external standards and benchmarks and shorten its list and keep those that count most. For instance, all schools should be held accountable for their graduation rate, mathematics and reading at each grade level.
- Advocate for small schools in terms of the number of students. This will make each teachers to know and be known by colleagues and every students at the school and make it difficult for teachers to hind their performance. In a big school, teachers and students are too many to know each other in that respect. Although this claim is not data supported, it is a promising idea.
- Encourage peer reviewing/evaluation of teaching. There is need for teachers more so more experienced teachers to sit in the classroom of less experienced teachers and give feedback and advise. The experienced teachers should also be expected to be evaluated in similar ways simply because not always that experience means “good”. Experienced and less experienced can learn from each other with the goal of making their students learn. This has been tried in colleges when young faculty goes up for tenure and promotion. Although there is no hard data to support the claim, it seems a promising idea.
- Schools should encourage planning time for teachers where they can discuss their students, what they are doing with them and what they are not doing. In this study, there was a bit of this element of planning time with the guidance of the mathematics coach but the efficacy of this meeting in terms of internal accountability need to be investigated more. There is need for a mentorship program to help those teachers who are struggling.
- The issue of reducing class size can reduce student teacher ratio. This issue was raised in this study as one aspect that external accountability does not take into consideration. The assumption here is that the low student teacher ration, the more attention each student will get and may lead to providing opportunity of student to learn.
- Finally, students need also be given a chance to evaluate their teachers, be part of choosing the curriculum, being part of firing and hiring of teachers. Having student to evaluate the teachers can help a great deal because students are the best people who know their teachers best. This has been used in college where teachers are evaluated by students and is being considered for tenure and promotions. (Levine, 2005)

Finally, even if my assertion that there is a need to assess the impact of the accountability system, this study does not give evidence on how accountability affected or did not affect student achievement. There were a many factors that were in play such as the CMP materials and the presence of the mathematics coach. More research is needed specifically on this school district.

References

- Dey, I. (1999). *Grounding grounded theory: Guidelines for qualitative inquiry*. San Diego: Academic Press.
- Elmore, R. F., & Fuhrman, S. H. (2001a). Holding schools accountable: Is it working? *Phi Delta Kappan*, 83(1), 67-72.
- Elmore, R. F., & Fuhrman, S. H. (2001b). Research finds the false assumption of accountability. *Education Digest*, 67(4), 9-14.
- Ezzy, D. (2002). *Qualitative analysis: Practice and innovation*. London: Routledge.
- Fuller, J. E., & Johnson, F. J. (2001). Can state accountability systems drive improvement in school performance for children of color and children from low-income homes? *Education and Urban Society*, 33(3), 260-283.
- Georgia Department of Education. Georgia performance standards. *Curriculum frequently asked questions* Retrieved October 22, 2005, from <http://www.georgiastandards.org/faq.asp#faq1>
- Georgia Department of Education. Qcc standards and resources. *Quality Basic Education Act Overview* Retrieved October 20, 2005, from <http://www.glc.k12.ga.us/qcc/qbe.htm>
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing.
- Hanushek, A. E., & Raymond, E. M. (2005). Does school accountability leads to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Johnson, J., & Immerwahr, J. (1995). First things first: What Americans expect from the public schools. A report from public agenda. *American Educator*, 18(4), 4.
- Kalamazoo schools. (1974). *American School Board Journal*, 32-40.
- Levine, P. (2005). "internal accountability" in education. Retrieved July, 03, 2007, from <http://www.peterlevine.ws/mt/archives/000644.html>
- Linn, L. R. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Department of Education.
- Newmann, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review*, 67(1), 41-74.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Powers, M. J. (2003). An analysis of performance-based accountability: Factors shaping school performance in two urban school districts. *Educational Policy*, 17(5), 558-585.
- Stake, R. E. (1995). *The art of case study research: Perspectives on practice*. Thousand Oaks, CA: Sage Publications.
- U.S. Congress. (2001). *No child left behind act of 2001*. Washington, DC: Author.
- Watson, S., & Supovitz, J. A. (2001). Autonomy and accountability in the context of standards-based reform, *Education Policy Analysis Archives* (Vol. 9).
- Yin, R. K. (2003a). *Applications of case study research* (2nd ed.). Thousand Oaks: Sage Publications.
- Yin, R. K. (2003b). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.