THE EFFECTS OF EDUCATION ACCOUNTABILITY ON TEACHERS: ARE POLICIES TOO-STRESS PROVOKING FOR THEIR OWN GOOD?

JOSEPH BERRYHILL
University of North Carolina at Asheville

JEAN ANN LINNEY
University of Notre Dame

JILL FROMEWICK
Harvard University

Education policies in the United States and other nations have established academic standards and made teachers accountable for improved standardized test scores. Because policies can have unintended effects, in this study we investigated U.S. elementary school teachers' perceptions of their state's accountability policy, particularly its effect on their job engagement. We found support for a path model relating lack of policy support to teacher burnout via two mediators: role conflict and reduced self-efficacy. Results of interviews with a subset of teachers were consistent with the model. We conclude with recommendations to reduce teacher stress in manners consistent with the goals of accountability policies.

In recent years, state and national leaders in the United States, backed by public opinion, have taken the view that public schools frequently fail to give students a good education, causing a myriad of bad outcomes and, in the view of some, even putting the national welfare in jeopardy (Madaus, Raczek, & Clarke, 1997). This perceived problem has resulted in policies to hold educators accountable for imparting higher academic standards, as measured by high-stakes, standardized tests (Elmore, Abelmann, & Fuhrman, 1996). These reforms reached a crescendo with the passage of the federal No Child Left Behind Act of 2001 that required each state to have standards on which students would be tested, with a demanding and intricate accountability system accompanying it (Anderson, 2005). Partly because nations seek to be competitive in an increasingly global marketplace, accountability also has been a frequent part of education reform worldwide (Levin, 1998; Tatto, 2006).

However, policies have the potential to induce costs as well as benefits (Darling-Hammond & Wise, 1985), and research has repeatedly shown that education reform policies may have unintended effects, or what Marshall (1998) terms "slippage between policy intent and outcomes" (p. 103). Linn (2000) argues that unintended negative effects often overshadow positive ones when accountability systems have high-stakes testing and calls for evaluation of such negative outcomes.

One hypothesized outcome of education accountability policies is pressure that contributes to teacher burnout (Farber, 1991). Schools with a strong drive toward measurable academic goals tend to have more burned-out teachers (Friedman, 1991). Studies also point to accountability as increasing stress for teachers (McNeil, 2000; Smith, 1991).

Increases in burnout, of course, could be related to teachers leaving the profession (Goddard & Goddard,

2006), a critical issue given that increasing enrollments and decreasing class sizes, among other factors, have made it difficult for school systems to attract and retain qualified teachers (Murnane & Steele, 2007). Furthermore, teachers are more likely to leave low-performing and minority districts (Guarino, Santibañez, & Daley, 2006), and there is evidence that accountability policies exacerbate the problem (Clotfelter, Ladd, Vigdor, & Diaz, 2004).

Thus in the study reported here, we explore relationships between burnout and teachers' perceptions of one state's accountability policy—principally its requirement that student academic achievement improve as demonstrated by results on a high-stakes test. Burnout is a construct with multiple definitions, but it is generally characterized by professionals feeling that their efforts are ineffective, the tasks never ending, and payoffs not forthcoming (Farber, 1991). Empirical evidence from the most widely used measure of burnout points to three dimensions: emotional exhaustion, or feelings of being emotionally drained; depersonalization, or callous treatment of those one is serving; and reduced personal accomplishment, or a decline in one's feelings of achievement (Maslach, Jackson, & Leiter, 1996). Stressors that teachers experience often lead to emotional exhaustion (Byrne, 1994; Leiter, 1991), which in turn can lead to depersonalization (Leiter, 1993). Personal accomplishment has weaker relationships to the other two burnout variables (Maslach et al.), but Byrne demonstrated that emotional exhaustion and depersonalization negatively relate to personal accomplishment for elementary teachers.

Potential Links Between Policy and Burnout

Numerous variables could potentially mediate the relationship between accountability and burnout. Two stand out conceptually and empirically: role conflict and teacher self-efficacy, which is often referred to as *teacher efficacy*. Role conflict is the occurrence of two or more requirements or pressures that are judged incompatible (Rizzo, House, & Lirtzman, 1970). Role conflicts include dilemmas such as trying to balance quantity of material to be covered with quality of the work (Byrne, 1994). Teacher efficacy is a cognitive construct that has developed out of Bandura's (1977) work (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). The concept is difficult to define and measure, but a general definition is "a teacher's belief or conviction that he or she can influence how well students learn, even those who may be difficult or unmotivated" (Guskey, 1987, p. 41).

Role conflict has demonstrated relationships with burnout (Byrne, 1994; Ginsberg, Cohn, Williams, Pritchett, & Smith, 1989; Schwab & Iwanicki, 1982). Results are inconsistent, with role conflict sometimes related to emotional exhaustion and depersonalization (e.g., Schwab & Iwanicki, 1982) and other times only to emotional exhaustion (e.g., Byrne). Likewise, role conflict also has emerged as a consequence of reforms increasing academic standards. Teachers in the United Kingdom required to follow a national curriculum and log children's progress felt they had to choose between what was to be tested and what they thought important (Black, 1994) and between the values of pursuing academic achievement and of social and emotional development (Dunham, 1992). In the United States, teachers modifying their reading and language arts programs were chronically nervous that standardized test scores might decrease as a result, although they did not (S. D. Miller, 1995).

Studies demonstrate that higher teacher efficacy is negatively related to job burnout (Brissie, Hoover-Dempsey, & Bassler, 1988; Chwalisz, Altmaier, & Russell, 1992). Higher levels of teacher efficacy also are strongly related to higher commitment to the teaching profession (Coladarci, 1992). Meanwhile, policy analysts have found that accountability policies put teachers in a position in which they do not feel efficacious. That is, many educators do not believe they can influence student learning as demanded by accountability systems (Abelmann, Elmore, Even, Kenyon, & Marshall, 1999; Conley & Goldman, 1995).

Another variable, teaching in a low-achieving school, is a potential antecedent that might lead teachers to experience reduced self-efficacy and additional pressure from accountability policy, although the evidence is less direct. Educators tend to have lower efficacy with nonacademic track classes than with academic and honors classes (Raudenbush, Rowan, & Cheong, 1992) and to feel less able to influence how well low-ability students learn (Cooper, Burger, & Seymour, 1979). As for relationships with accountability policies, teachers who work with low-achieving students report having less faith that administrators would interpret test scores accurately (Urdan & Paris, 1994). Hence it may be that teaching in a low-achieving school drives the kind of pressure Lipsky (1980) describes for "street-level bureaucrats" to exonerate themselves of responsibility for outcomes.

Based on this empirical evidence, we developed a path model linking pressure from accountability policy to teacher burnout via the mediators of role conflict and

reduced self-efficacy. The model also had lower existing school test scores as an antecedent driving more policy pressure and reducing self-efficacy, although this was a tentative hypothesis due to more limited evidence. The model is depicted in Figure 1.

Given the exploratory nature of the research question, we wanted teachers to tell us about the effect of accountability policies in ways beyond Likert scale ratings. Thus the study included an interview so that teachers could explain the impact of policies on them.
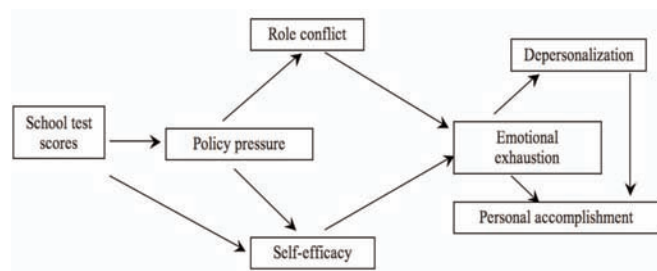

Figure 1. Model of proposed relationships among variables.

Context

Our study took place in South Carolina, where standardized test scores traditionally have been among the lowest in the United States. The focus of our study had its origins in the Education Accountability Act of 1998, which resulted in national praise for its high standards ("Standards and Accountability," 2001). It set academic standards for each grade and standards-based examinations in grades 3–8 and 10, the last of which serves as the exit exam. The test, called the Palmetto Achievement Challenge Test (PACT), included multiple-choice and open-response items asking students how they arrived at answers. For individual students, the law attempted to eliminate the practice of "social promotions," with parents of those not performing to grade levels required to come in for a conference to establish means to improve outcomes. Further poor performance could result in a student's being placed on academic probation or retained. At the aggregate level, the state department of education was to provide report cards to schools and districts on their performance—both by noting the percentage of students performing well or not in a given year and by measuring overall improvement from year to year by tracking students. Schools that exceeded expected results were to receive recognition and incentives. Schools doing poorly were to devise an improvement plan, with the state providing assistance to help with weaknesses.

Method

Setting and Participants

The study took place in spring 2000 in a part urban and part suburban school district (we use the pseudonym "Springfield") with teachers of grades 3–5 (students take the PACT beginning in 3rd grade). Slightly more than a third of the Springfield's students in grades 3, 4, and 5 received free or reduced-price lunch at the time of the study, whereas the state average for those grades was slightly above 50 percent. The district's 11 elementary schools had PACT scores in the initial year of testing ranging from average to well above average.

We approached principals at nine of the district's elementary schools with the highest and lowest PACT scores for permission to have their teachers participate, and each allowed teachers to take part. Most principals allowed the first author to meet with teachers and explain the study, although two principals told their teachers about the study themselves using an information sheet provided.

A total of 100 teachers—76 percent of all possible participants—completed the questionnaires, although one was dropped because she was a special education teacher, not a regular classroom teacher. The main variable separating participating teachers from nonparticipants appeared to be hearing about the study in person from the first author, as 61 percent of the nonparticipants were informed via their principal or electronic mail. We paid teachers $10 for completing the questionnaires, and we donated $5 to their schools for each participating teacher; all payments were from the first author.

The teachers had a mean of 13.1 years' experience in the classroom (median of 13.0), with a range of 1 to 32 years of teaching. They had a mean of 8.2 years' experience in the Springfield District and 6.1 years at their current school. Nearly two-thirds—64 percent—had master's degrees. The mean age was 39.3 years. Ninety were female. Participants were almost evenly split on grade taught: 35 3rd grade, 31 4th grade, and 33 5th grade teachers.

For the interview portion of the study, the goal was to select one of every five teachers completing questionnaires. Using a median split, we categorized teachers as high- or low-burnout based on scores from an emotional exhaustion subscale described below. We then grouped teachers in each category by school and ran-

domly selected as proportionately as possible from each school.

Of the 20 teachers selected, 14—70 percent—accepted. For each teacher who declined, we asked a randomly selected teacher from the same school (and the same side of the median split) to participate, resulting in nine teachers being contacted to fill the six open slots. We paid teachers $10 for participating in the interview phase of the study, with a donation of $5 again going to their school; all payments were from the first author.

The teachers interviewed were virtually the same demographically as the entire sample. One major difference was that 5th grade teachers were overrepresented (50 percent) and 4th grade ones underrepresented (20 percent). The interviewer was blind to the burnout score of teachers.

Measures

Teachers completed a series of questionnaires about their views of accountability policies and their work (e.g., self-efficacy and burnout). Several measures were not used for analyses in this study and thus are not reported here.

Policy Questionnaire

Because of the lack of a quantitative measure on teacher attitudes toward accountability policies, we devised a 24-item instrument, the Accountability Policies Stressors and Supports Questionnaire (see Appendix A) that asks teachers about perceived stressors and support resulting from policies. We pilot-tested the questions with teachers at a school in a different district than the one studied. We utilized data from the measure in testing the causal model; as will be reported below, we ultimately reduced it to a three-item measure, policy support, with an alpha internal consistency coefficient of .80.

Other measures

We also utilized other measures:

Teacher efficacy. We used a shortened version (see Tschannen-Moran et al., 1998) of the Teacher Efficacy Scale (Gibson & Dembo, 1984) as a measure of two factors: general teaching efficacy and personal teaching efficacy. Personal teaching efficacy reflects teachers' perceived ability to effect positive student change. General teaching efficacy has been interpreted in different manners, but it may be considered teachers' views of the impact of constraints outside the classroom (e.g., home conditions) that can negatively affect teaching (Tschannen-Moran et al.).

The five-item general efficacy scale has an alpha internal consistency coefficient of .72, whereas the five-item personal efficacy scale has an alpha of .77 (Hoy & Woolfolk, 1993). For the current sample, the alpha coefficients were .75 for the general measure and .74 for the personal efficacy measure.

Role conflict. We used a shortened version of the Role Questionnaire of Rizzo and colleagues (1970) to measure "intrarole intersender" conflicts in which individuals perceive work tasks as beyond their resources or unnecessary (Schwab, Iwanicki, & Pierson, 1983). This four-item subscale was validated with a sample of teachers and had an alpha of .78. With the current sample, the alpha was .82.

Teacher burnout. We used the 22-item Maslach Burnout Inventory for Educators (Maslach et al., 1996) to measure the three subscales of burnout. The first subscale, emotional exhaustion, has an alpha internal consistency of .90; the second, depersonalization, has an alpha of .76; and the third, personal accomplishment, has an alpha of .76 (Iwanicki & Schwab, 1981). For the current sample, the alpha coefficients were .90 (emotional exhaustion), .70 (depersonalization), and .79 (personal accomplishment).

Semi-structured interview. Following principles delineated by Patton (1990), we developed an interview asking about the impact of accountability policies on teachers' work. After pilot tests for clarity and completeness, we added a few questions and reworded a few, resulting in a semistructured interview of 22 questions. We asked open-ended questions about what teachers thought legislators wanted them to do, followed by questions on the impact of accountability policies, including whether teachers viewed policies as supportive or stressful; whether they saw relationships between policies and their efficacy and role conflicts; and whether policies affected their relationships with administrators, colleagues, or students.

Procedure

We completed quantitative measures a few weeks before PACT testing began. Interviews, which ranged between 45 and 90 minutes, took place at school, either in the weeks before, during, or after PACT testing. We tape-recorded and transcribed them verbatim.

Results

We report quantitative results first, with interview results following.

Quantitative Results
We report quantitative results with (1) factor analysis of the policy stress and support questionnaire; (2) descriptive statistics; and (3) tests of the model.

Factor Analysis
We undertook a factor analysis of the policies stress and support questionnaire to create the policy pressure variable for the model. Maximum likelihood extraction methods, followed by an orthogonal rotation with Varimax procedures, resulted in four distinct factors of policy outcomes: support for efforts to educate students, general support from sources such as the state and parents, pressure from supervisors and parents, and a lack of fairness of teacher evaluations.

We undertook a confirmatory factor analysis to determine if the four factors represented manifest variables for the latent construct of policy pressure, but the model provided inadequate fit (e.g., Tucker-Lewis index = .84). Thus we selected one of the factors, the three-item help in educating students, for the model because its items reflected teachers' views of accountability policies helping them educate students, a construct closer to the core of the policies' impact on their daily work than the other factors. The factor consisted of items saying that policies to increase student academic performance "help me more effectively impart knowledge," "overall provide more support for me in my efforts to educate students," and "make me feel the state is helping me improve students' academic performance."

Descriptive Statistics
In measures used in the model (see Table 1), policy support was moderately low (*M* = 7.7 on a scale ranging from 3 to 18, with higher numbers indicating more support). Personal self-efficacy was relatively high, whereas general self-efficacy was somewhat low. Role conflict occurred at moderate levels. As for burnout, emotional exhaustion was relatively elevated (*M* = 26.5), whereas depersonalization was low (*M* = 5.6) and personal accomplishment relatively high (*M* = 40.7).

Tests of the Path Model
The depersonalization and personal accomplishment subscales had high degrees of skewness, and thus we undertook square root transformations, resulting in relatively normal distributions. Because the transformation reflected the personal accomplishment scale, resulting in a reversal of direction of interpretation, the variable is called *decreased personal accomplishment*.

The personal efficacy measure had a severely restricted range, and one item ("If one of my students couldn't do a class assignment, I would be able to accurately assess whether the assignment was at the correct level of difficulty") had very low correlations with the four others ("When I really try, I can get through to most difficult students;" "If a student did not remember information I gave in a previous lesson, I would know how to increase his or her retention in the next lesson;" "If a student in my class becomes disruptive and noisy, I feel assured that I know some techniques to redirect him or her quickly;"and "If I try really hard, I can get through to even the most difficult or unmotivated students"). Furthermore, in the interviews almost all teachers pointed to influences other than themselves (particularly home and family factors used in the general efficacy scale) as influences on student academic outcomes. Therefore, we utilized the general self-efficacy measure in the path model.

Table 1. Means, Standard Deviation, and Scale Ranges for Model Measures			
Measure	Mean	Standard Deviation	Range
Policy support	7.7	3.4	3–18
Personal self-efficacy	25.2	3.3	5–30
General self-efficacy	13.4	4.7	5–30
Role conflict	18.3	6.2	4–28
Emotional exhaustion	26.5	11.5	0–54
Depersonalization	5.6	5.2	0–30
Personal accomplishment	40.7	6.1	0–48

As a final step before testing the model, we used SAS (statistical analysis software) (version 8.0) to examine correlations between school PACT scores, within-school grade-level PACT scores, self-efficacy, and policy support. State PACT scores and grade PACT scores were highly correlated (*r* = .94, p < .0001). However, neither school nor grade PACT scores were significantly associated with self-efficacy (*r* = -.07, *r* = -.02, respectively) or policy pressure (*r* = -.16, *r* = -.15, respectively). We also tested these relationships using multilevel analytic techniques, with teachers (level 1) nested within grades (level 2) within schools (level 3). Using MLwiN (version 1.1), we modeled the impact of state and grade PACT scores

(due to their high correlation, we included them in separate models) on self-efficacy and policy pressure, respectively. We found no significant relationships and, therefore, did not utilize test scores in the model.
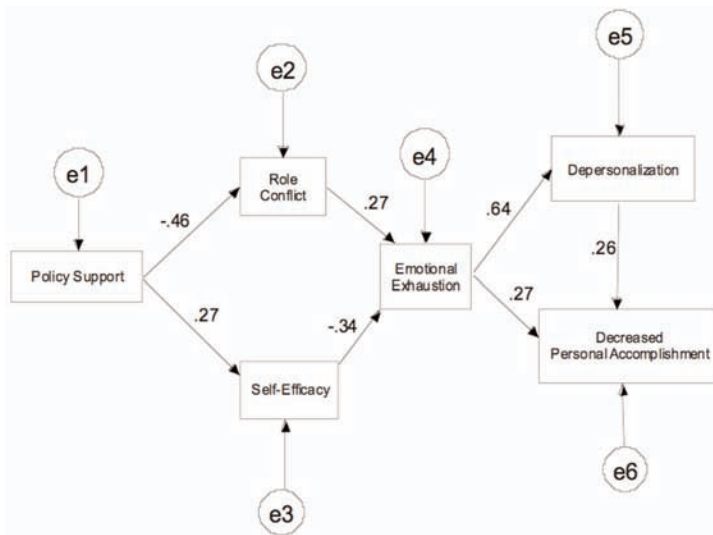
We analyzed the hypothesized path model using maximum likelihood estimation. Results with standardized regression coefficients are presented in Figure 2. All paths in the model were significant, indicating no need for model trimming or building. The model displayed excellent goodness of fit by both absolute and comparative measures (see Table 2). Path coefficients ranged from .26 to .64, indicating generally strong relationships among model variables.

Interview Results

The first author analyzed transcripts from the 20 interviews using anthropological methods centering on creat-

ing taxonomic analyses and searching for cultural themes (Spradley, 1980. Then the first author wrote codes for

Table 2. Goodness-of-Fit Indices for the Original and Modified Path-Analytic Model	
Fit Index	Model Statistic
Generalized Likelihood Ratio (chi-square)	4.80[a]
Normed Fit Index	.97
Adjusted Goodness of Fit	.96
Tucker-Lewis Index	1.0
Standardized Root Mean Squared Residual	.04
[a] df = 8 (*p* = .60)	

sections of text and subsequently refined them and checked them for accuracy with iterative passes through the data. The first author next created taxonomies for most questions, and reviewed them for similar themes, resulting in a final taxonomy outlining constructs, including multiple types of teacher stress, related to accountability policies. Finally, the first author created a separate taxonomy of teacher suggestions for changing policy to improve student academic outcomes.

Figure 2. Path-Analytic Model with Standardized Regression Coefficients (All Significant)

Stressors

Limits on teacher self-efficacy. All teachers interviewed made at least one reference to their influence on student academic performance being limited. Efficacy for improving academic achievement can be conceptualized as stemming from teachers' personal factors, such as having the ability to help students, as well as limited by situational factors, such as pupils' home life (see Gist & Mitchell, 1992). Using that rubric, the results are striking. Virtually no teachers noted any lack of ability on their part as limiting their efficacy. However, 80 percent of teachers made references to home or family factors as having an influence on student achievement, noting factors such as parental involvement, socioeconomic status, stressors children experience, and exposure to opportunities to learn outside of school. One 4th grade teacher said this about home factors:

… I feel in my experience that there's a direct correlation between children who are spoken with at home and children who can write. Those who are spoken to at home can write good stories. Those who are never spoken to can't write anything. Or it's much harder for them.… I think that that puts too much pressure on me, you know? I can only improve as much as, you know, the hopes are. I can only do so much.

Many teachers—40 percent—also felt pressured by student deficits or problems that could limit their academic abilities. These responses typically included references to learning disabilities, low intelligence quotients, or other deficits. One 3rd grade teacher said:

And once again, I've had certain students that have tested like borderline, they're at 71 in their IQ. Well, they're being tested. Now if you know you have a 71 IQ student, and I taught him all these standards, and I prepped him for PACT, do I think he's going to do well on PACT? No. And I don't think it's because I taught him. I think he's going to do what he's capable of.

Limited time and too many tasks. Eighty percent of teachers reported having either limited time to complete

tasks related to accountability policies or having so many accountability-related tasks that it was difficult to complete them. Half of the teachers discussed the lack of time as affecting their ability to integrate other goals. These included academic goals—such as working with all students sufficiently, teaching all subjects thoroughly, or making academic work engaging for students—as well as nonacademic ones—such as helping students develop prosocial behaviors. Thirty percent of teachers said education objectives and standards were so numerous that it was difficult to cover them all, particularly by the time standardized tests were administered. As one 5th grade teacher said:

Well, we're mandated to teach certain things by the district and then the state has standards. Umm, and then they're testing on those things, so and it's a lot to cover and cover it well, and it's just like hurry, hurry, hurry. And it's not like you're working in a factory here with little cookie-cutter things that just, you've got to have just some time to reach the children that maybe need a little bit more help and it's so fast-paced sometimes to get everything in, that some of the children get left behind and that's what bothers me.

Changes in teaching practices. In a related theme, two-thirds of teachers discussed accountability policies changing either the manner in which they taught or the emphasis on the subjects they taught, often as a result of the time pressures delineated above. The most common change in teaching, mentioned by 35 percent of teachers, was no longer emphasizing that all students master principles before moving to the next topic. Other changes included de-emphasizing group work (15 percent) and making learning more teacher-directed, as in having more lectures (20 percent). A 3rd grade teacher, noting a day when it had snowed, said the state's accountability policy:

makes us, for better or worse, umm, stick more to those plans that we have in those books, which before we always said the plans aren't written in stone, if you see a teachable moment, grab it. And we're not grabbing those teachable moments anymore. I mean, if something were to happen like when we had snow, instead of doing something fun and we could have really had a great lesson on snow and what makes it snow, that wasn't in our standards, that's not in our curriculum, we don't do that anymore.

As for changes in teaching subjects, 20 percent of teachers discussed emphasizing or de-emphasizing subject material. One 5th grade teacher decided to focus almost exclusively on the PACT subjects—English and math—six weeks before the test after fellow teachers said they had virtually dropped non-PACT subjects: "It's almost if you talk to other people in the school, people will say, 'Oh, you're still teaching social studies? Are you still teaching science?' So, it's kind of, it's almost school-wide that people have quit.…"

Concerns about the fairness of standardized test scores. Eighty-five percent of teachers had concerns about the ways they were evaluated based on test scores. Their reasons varied widely and included concerns that the PACT only indirectly measured their work because students took the test. The use of the scores was a large concern, with 40 percent believing the results were overemphasized. As one 3rd grade teacher stated, "I think my problem is really not so much with standards as it is the use of a state-created test as the one determining factor of whether those standards have been met or not."

Teacher Suggestions

Teachers made a variety of suggestions for changes that would improve academic outcomes of students, usually given in response to a question about what the teacher would do if given the power to change policy. The suggestions are reported in four categories.

Changes to increase fairness. Nearly two-thirds—65 percent—of teachers made at least one suggestion to enhance the fairness of the assessment of teachers and students, including measuring student progress with pre- and post-tests and having a variety of assessments instead of one test.

More resources or fewer demands. Nearly two-thirds—65 percent—of teachers suggested reducing the demands on them or giving them more resources, including reducing class size, hiring teacher aides, or obtaining more updated materials or equipment.

Changes in teacher-legislator relations. Fifty-five percent of teachers suggested changes in interactions with legislators. Nearly half advocated for lawmakers come to their classrooms to gain better understanding of teachers' jobs and the education process. Others called for legislators to involve them in the policymaking process or praise them more for their efforts.

Other suggestions. More than a third—35 percent—of teachers made suggestions about increasing the professionalism of teaching. The most popular suggestion was an increase in salary. Fifteen percent of teachers made suggestions that might be considered broad-scale changes, such as a more even distribution of school funding statewide or smaller school sizes.

Discussion

This study set out to discover whether elementary school teachers perceived accountability policies as having unintended consequences for their well-being. The answer was yes: teachers reported a myriad of negative policy consequences, a finding consistent with other investigations (e.g., Abrams, Pedulla, & Madaus, 2003; Barksdale-Ladd & Thomas, 2000; M. G. Jones, Jones, Hardin, Chapman, Yarbrough, & Davis, 1999; B. D. Jones & Egley, 2004, 2007; Webb, 2006). A lack of policy support was associated with teachers' emotional exhaustion via the mediators of role conflict and low self-efficacy.

The interviews were consistent with the quantitative results. Teachers discussed numerous types of role conflicts, often driven by the lack of time to impart all education standards. Racing to teach all standards, teachers often altered their teaching practices and had less time to pursue other goals. Similar dynamics have been documented in other states with accountability policies (Barksdale-Ladd & Thomas, 2000; Lasky, 2005; McNeil, 2000; Smith, 1991; Valli & Buese, 2007).

In terms of self-efficacy, teachers had few concerns about personal factors, such as their own pedagogical capabilities, but often believed that situational, student-related factors (e.g., home conditions) played a role. In the context of accountability policies, efficacy has a new twist, as task attainment is evaluated not by teachers' direct performance but rather indirectly via their students' test scores. Teachers' lack of efficacy may represent their relieving stress by absolving themselves from responsibility for student outcomes, a typical outcome among those who implement policies (Lipsky, 1980). Teachers also may be at least partly on the mark in their assessment of having limited efficacy, as school outcomes on standardized tests are highly influenced by student socioeconomic status (Powers, 2003). However, teachers' pointing to student deficiencies raises the question of whether accountability policies lead teachers to focus on deficits rather than overcoming them.

One factor that might plausibly have been related to teachers' views of accountability policies and their burnout levels, the school PACT scores, showed no relationship. It is likely that there was not enough variance in school PACT scores for differences to emerge. Alternatively, teachers who instruct different types of students may be under relatively equal pressure. Grant (2000) found that teachers at varying types of schools all experienced uncertainty over whether test scores would increase and that, if anything, suburban teachers were more concerned than urban counterparts, perhaps to maintain an existing high status.

Limitations and Alternative Explanations

This study has limitations, including use of a new measure of policy support and reliance on teacher report alone. Alternative explanations for the results exist. For example, policy-related stress may have been less salient in this study had it taken place in the earlier stages of the school year.

It also is possible that the variables interact in ways other than that prescribed in the model. An alternative explanation to the model's relationships is that teachers already experiencing some burnout had negative reactions to South Carolina's new accountability system. Even if this is so, most teachers nonetheless find the policy unsupportive and therefore its chances for reaching the goal of enhanced student learning remain diminished.

The study also involves teachers from one locale and does not examine the multiple levels of policy, which in the United States can pass from the federal government to states to districts to schools and finally to teachers themselves. Context is likely to play a strong role in the impact of policies on teachers, because their responses to accountability are affected by factors such as principal support (Kelley & Finnigan, 2003) and district administrators (Louis, Febey, & Schroeder, 2005).

A final limitation is that the data examined in this study were collected in 2000, and teachers' views of accountability may have changed either with time, changes in policy, or both. However, data were collected after the passage of South Carolina's law that established the accountability framework lasting through 2008. Replications of this study would be helpful to determine whether teachers' views are changing.

Implications for Policy

Given the limitations of the study, we do not suggest that policy be reformulated solely based on its results. However, the possibility of accountability policies resulting in diminished job engagement has implications that policymakers should consider. Principally, if teachers become emotionally exhausted, the long-term outcome may be leaving the profession or decreased commitment and enthusiasm (Leiter & Maslach, 1998). Either result is unlikely to lead to what accountability policies seek: better-prepared students.

What can policymakers do to promote teachers being engaged to lead students to better academic outcomes without pressures that may push them toward burnout? One possibility is undertaking stress-reduction interventions with individual teachers. However, these strategies are unlikely to appeal to policymakers because they often encourage less work or lowered expectations— and may run counter to the hope of increasing student academic achievement. Furthermore, burnout results from both individual and contextual factors (Kelchtermans & Strittmatter, 1999). Making changes in individuals when the system is part of the problem leaves basic structures intact and is unlikely to affect the problem (Seidman, 1988). Therefore, policymakers should consider making changes *for* teachers rather than *in* teachers.

Teachers in this study suggested such changes, and we believe policymakers would be wise to pay attention to them. One suggestion was having more opportunities to interact with legislators about education and policy. Teachers may take the initiative to contact legislators or have their opinions heard through their union or other means, but their voices are often not strongly present in debates over policies or reforms (M. G. Jones et al., 1999; S. D. Miller, Hayes, & Atkinson, 1997; Smith & Fey, 2000). If state officials don't provide an invitation and process in which teachers can provide input, they risk mistrusting the people they most need to make their policies work (for an example of this process in action, see S. D. Miller et al., 1997). Giving teachers a means to have more policy input also makes policies more consistent. The policy establishing new standards and tests in South Carolina strongly emphasizes that students employ problem solving and critical thinking. However, the state has not given teachers the opportunity to engage in those types of processes to formulate or improve policies (similar inconsistencies are discussed in Noble & Smith, 1994).

Teachers' other suggestions in this study included more professionalism in teaching. We believe an important step in that direction would be more widespread adoption of a professional development model, a step recommended for accountability policies (Sirotnik & Kimball, 1999). We define professional development as Lieberman and L. Miller (1990) do: a dramatic alteration of time and resources that creates schools of professional practice in which there are opportunities for activities such as teacher study groups to discuss areas of professional interest and peer observations to critique strengths and weaknesses. This represents an investment in improving the quality of teaching, a variable that has been more effective in the improvement of academic achievement than implementation of student testing (Darling-Hammond, 2000, 2004). Furthermore, there is evidence that professional development schools can serve as buffers against burnout (L. Miller, 1999) and that schools incorporating aspects of professional development result in teachers having increased efficacy (Chester & Beaudin, 1996; Henson, 2001).

A related trend toward professionalism in teaching has been the practice of teacher leadership (Katzenmeyer & Moller, 2001). A broad construct, teacher leadership involves teachers taking lead roles in tasks traditionally given to administrators, giving them influence and decision-making power in areas such as school management, curriculum, school improvement, and community involvement (York-Barr & Duke, 2004). There are numerous examples of teachers taking leadership roles in the United States and other nations (e.g., Crowther, Kaagan, Ferguson, & Hann, 2002; Frost, Durrant, Head, & Holden, 2000; Katzenmeyer & Moller; Lieberman & L. Miller, 2004; Muijs & Harris, 2006). While there is a need to know more about the effects of teacher leadership (York-Barr & Duke, 2004), evidence indicates that teachers sharing leadership with principals can have a positive impact on student achievement (Marks & Printy, 2003; Printy & Marks, 2006). Teacher leadership thus offers a potential means of increasing efficacy and lowering burnout.

Another suggestion teachers made in this study was to enhance the fairness of assessments of teachers and students. This issue is receiving increasing attention, as there is empirical evidence that, for example, the No Child Left Behind's measurement of academic progress lacks validity (Balfanz, Legters, West, & Weber, 2007). The federal government is granting some flexibility to states to change measurement rubrics, including using learning gains (Peterson & West, 2006). South Carolina is not among those states, although the state's accountability law tracks student progress. Another means to increase fairness would be to replace a single high-stakes test with a variety of measures, a step that Linn (2000) views as essential. Because more evaluations could increase stress on teachers, we advocate following Porter and Chester's (2002) recommendation that multiple measures be employed with the philosophy of promoting more opportunities for students to succeed—a plan that has been undertaken in schools districts such as Philadelphia's (Porter, Chester, & Schlesinger, 2004). Creating what Porter and Chester call "a bias in favor of

success" (p. 290) could ease pressure and increase support for teachers and students. If multiple measures are impractical, districts could invest in software that almost instantaneously assesses student progress in areas like math. Yeh (2006) found that teachers using such software were able to efficiently assess student weaknesses and counteract them, potentially reducing the stress of high-stakes testing.

As a final consideration, teachers in this study often focused on the outside forces in students' lives that may limit their academic outcomes. It is our hope that teachers not overemphasize these forces, but they are correct in seeing that factors beyond individual students have an impact in school.

One means for schools to have an effect on "outside forces" is through fostering greater family and community involvement in schools, which the No Child Left Behind Act requires (Epstein, 2005). Although a variety of factors affect parental involvement, it is influenced by schools, with teachers often playing important roles (Hoover-Dempsey et al., 2005). Strongly implemented partnerships of schools, families, and communities can lead to better outcomes for students, such as better attendance (Sheldon, 2007). Hence, teachers who are supported in involving parents may move away from the assumption that there are factors influencing student achievement that they cannot change.

However, we hope that policymakers who want improved academic outcomes will not stop at promoting school-community relations but also will consider addressing social and economic disparities to help all students come to school without disadvantages (Taylor, 1994). This may be idealistic, but without such changes, what teachers are asked to do under accountability policies is akin to asking physicians to return all patients, regardless of their initial condition, to optimal health.

References

Abelmann, C., Elmore, R., Even, J., Kenyon, S., & Marshall, J. (1999). *When accountability knocks, will anyone answer?* (Research Report No. RR-042). Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education. Retrieved June 20, 2006, from http://www.cpre.org/images/stories/cpre_pdfs/rr42.pdf

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory Into Practice*, 42, 18–29.

Anderson, L. W. (2005, April 4). The No Child Left Behind Act and the legacy of federal aid to education. *Education Policy Analysis Archives*, 13(24). Retrieved July 1, 2005, from http://epaa.asu.edu/epaa/v13n24/

Balfanz, R., Legters, N., West, T. C., & Weber, L. M. (2007). Are NCLB's measures, incentives, and improvement strategies the right ones for the nation's low-performing high schools? *American Educational Research Journal*, 44, 559–593.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.

Barksdale-Ladd, M. A., & Thomas, K. F. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education*, 51, 384–397.

Black, P. J. (1994). Performance assessment and accountability: The experience in England and Wales. *Educational Evaluation and Policy Analysis*, 16, 191–203.

Brissie, J. S., Hoover-Dempsey, K. V., & Bassler, O. (1988). Individual, situational contributors to teacher burnout. *Journal of Educational Research*, 82, 106–112.

Byrne, B. M. (1994). Burnout: Testing for the validity, replication, and invariance of causal structure across elementary, intermediate, and secondary teachers. *American Educational Research Journal*, 31, 645–673.

Chester, M. D., & Beaudin, B. Q. (1996). Efficacy beliefs of newly hired teachers in urban schools. *American Educational Research Journal*, 33, 233–257.

Chwalisz, K., Altmaier, E. M., & Russell, D. W. (1992). Causal attributions, self-efficacy cognitions, and coping with stress. *Journal of Social and Clinical Psychology*, 11, 377–400.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Diaz, R. A. (2004). Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, 23, 251–271.

Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *Journal of Experimental Education*, 60, 323–337.

Conley, D. T., & Goldman, P. (1995). Reactions from the field to state restructuring legislation. *Educational Administration Quarterly*, 31, 512–538.

Cooper, H. M., Burger, J. M., & Seymour, G. E. (1979). Classroom context and student ability as influences on teacher perceptions of classroom control. *American Educational Research Journal*, 16, 189–196.

Crowther, F., Kaagan, S. S., Ferguson, M., & Hann, L. (2002). *Developing teacher leaders: How teacher leadership enhances school success*. Thousand Oaks, CA: Corwin.

Darling-Hammond, L. (2000, January 1). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1). Retrieved June 10, 2004, from http://epaa.asu.edu/epaa/v8n1/

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106, 1047–1085.

Darling-Hammond, L., & Wise, A. E. (1985). Beyond standardization: State standards and school improvement. *Elementary School Journal*, 85, 315–336.

Dunham, J. (1992). *Stress in teaching* (2nd ed.). London: Routledge.

Elmore, R. F., Abelmann, 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: Performance-based reform in education* (pp. 65–98). Washington, DC: Brookings Institution.

Epstein, J. L. (2005). Attainable goals? The spirit and letter of the No Child Left Behind Act on parental involvement. *Sociology of Education*, 78, 179–182.

Farber, B. A. (1991). *Crisis in education: Stress and burnout in the American teacher*. San Francisco: Jossey-Bass.

Friedman, I. A. (1991). High- and low-burnout schools: School culture aspects of burnout. *Journal of Educational Research*, 84, 325–333.

Frost, D., Durrant, J., Head, M., & Holden, G. (2000). *Teacher-led school improvement*. New York: RoutledgeFalmer.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76, 569–582.

Ginsberg, R., Cohn, E., Williams, C. G., Pritchett, S. T., & Smith, T. (1989). *Teaching in South Carolina: A retirement initiative*. Columbia, SC: South Carolina Education Policy Center.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review*, 17, 183–211.

Goddard, R., & Goddard, M. (2006). Beginning teacher burnout in Queensland schools: Associations with serious intentions to leave. *Australian Educational Researcher*, 33(2), 61–75. Retrieved February 1, 2008, from http://www.aare.edu.au/aer/online/0602e.pdf

Grant, S. G. (2000, February 24). Teachers and tests: Exploring teachers' perceptions of changes in the New York State testing program. *Education Policy Analysis Archives*, 8(14). Retrieved November 4, 2006, from http://epaa.asu.edu/epaa/v8n14.html

Guarino, C. M., Santibañez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, 76, 173–208.

Guskey, T. R. (1987). Context variables that affect measures of teacher efficacy. *Journal of Educational Research*, 81, 41–47.

Henson, R. K. (2001). The effects of participation in teacher research on teacher efficacy. *Teaching and Teacher Education*, 17, 819–836.

Hoover-Dempsey, K. V., Walker, J. M. T., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S., & Closson, K. (2005). Why do parents become involved? Research findings and implications. *The Elementary School Journal*, 106, 105–130.

Hoy, W. K., & Woolfolk, A. E. (1993). Teachers' sense of efficacy and the organizational health of schools. *Elementary School Journal*, 93, 356–372.

Iwanicki, E. F., & Schwab, R. L. (1981). A cross-validational study of the Maslach Burnout Inventory. *Educational and Psychological Measurement*, 41, 1167–1174.

Jones, B. D., & Egley, R. J. (2004, August 9). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives*, 12(39). Retrieved May 20, 2005, from http://epaa.asu.edu/epaa/v12n39/

Jones, B. D., & Egley, R. J. (2007). Learning to take tests or learning for understanding? Teachers' beliefs about test-based accountability. *The Educational Forum*, 71, 232–248.

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81, 199–203.

Katzenmeyer, M., & Moller, G. (2001). *Awakening the sleeping giant: Helping teachers develop as leaders* (2nd ed.). Thousand Oaks, CA: Corwin.

Kelchtermans, G., & Strittmatter, A. (1999). Beyond individual burnout: A perspective for improved schools. Guides for the prevention of burnout. In R. Vandenberghe & A. M. Huberman (Eds.), *Understanding and preventing teacher burnout: A sourcebook of international research and practice* (pp. 304–314). Cambridge: Cambridge University Press.

Kelley, C. J., & Finnigan, K. (2003). The effects of organizational context on teacher expectancy. *Educational Administration Quarterly*, 39, 603–634.

Lasky, S. (2005). A sociolcultural approach to understanding teacher identity, agency and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education*, 21, 899–916.

Leiter, M. P. (1991). Coping patterns as predictors of burnout: The function of control and escapist coping patterns. *Journal of Organizational Behavior*, 12, 123–144.

Leiter, M. P. (1993). Burnout as a developmental process: Consideration of models. In W. B. Schaufeli, C. Maslach, & T. Marek (Eds.), *Professional burnout: Recent developments in theory and research* (pp. 237–250). Philadelphia: Taylor & Francis.

Leiter, M. P., & Maslach, C. (1998). Burnout. In H. S. Friedman (Ed.), *Encyclopedia of mental health* (Vol. 1) (pp. 347–357). San Diego: Academic Press.

Levin, B. (1998). An epidemic of education policy: (What) can we learn from each other? *Comparative Education*, 34, 131–141.

Lieberman, A., & Miller, L. (1990). Teacher development in professional development schools. *Teachers College Record*, 92, 105–122.

Lieberman, A., & Miller, L. (2004). *Teacher leadership*. San Francisco: Jossey-Bass.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York: Russell Sage Foundation.

Louis, K. S., Febey, K., & Schroeder, R. (2005). State-mandated accountability in high schools: Teachers' interpretations of a new era. *Educational Evaluation and Policy Analysis*, 27, 177–204.

Madaus, G. F., Raczek, A. E., & Clarke, M. M. (1997). The historical and policy foundations of the assessment movement. In A. L. Goodwin (Ed.), *Assessment for equity and inclusion: Embracing all our children* (pp. 1–33). New York: Routledge.

Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, 39, 370–397.

Marshall, C. (1998). Bridging the chasm between policymakers and educators. *Theory Into Practice*, 27(2), 98–105.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach burnout inventory manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

McNeil, L. M. (2000). *Contradictions of reform: The educational costs of standardized testing*. New York: Routledge.

Miller, L. (1999). Reframing teacher burnout in the context of school reform and teacher development in the United States. In R. Vandenberghe & A. M. Huberman (Eds.), *Understanding and preventing teacher burnout: A sourcebook of international research and practice* (pp. 139–156). Cambridge: Cambridge University Press.

Miller, S. D. (1995). Teachers' responses to test-driven accountability: If I change, will my test scores drop? *Reading Research and Instruction*, 34, 332–351.

Miller, S. D., Hayes, C. T., & Atkinson, T. S. (1997). State efforts to improve students' reading and language arts achievement: Does the left hand know what the right is doing? *Reading research and instruction*, 36, 267–286.

Muijs, D., & Harris, A. (2006). Teacher led school improvement: Teacher leadership in the UK. *Teaching and Teacher Education*, 22, 961–972.

Murnane, R. J., & Steele, J. L. (2007). What is the problem? The challenge of providing effective teachers for all children. *The Future of Children*, 17(1), 15–42. Retrieved February 10, 2008, from http://www.futureofchildren.org/usr_doc/7_02.pdf

Noble, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy*, 8, 111–136.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Peterson, P. E., & West, M. R. (2006). Is your child's school effective? Don't rely on NCLB to tell you. *Education Next*, 6(4), 76–80. Retrieved February 4, 2008, from http://media.hoover.org/documents/ednext20064_76.pdf

Porter, A., & Chester, M. (2002). Building a high-quality assessment and accountability program: The Philadelphia example. In D. Ravitch (Ed.), *Brookings Papers on Education Policy* (pp. 285–337). Washington, DC: Brookings Institution Press.

Porter, A. C., Chester, M. D., & Schlesinger, M. D. (2004). Framework for an effective assessment and accountability program: The Philadelphia example. *Teachers College Record*, 106, 1358–1400.

Powers, J. M. (2003). An analysis of performance-based accountability: Factors shaping school performance in two urban school districts. *Educational Policy*, 17, 558–585.

Printy, S. M., & Marks, H. M. (2006). Shared leadership for teacher and student learning. *Theory Into Practice*, 45, 125–132.

Raudenbush, S., Rowan, B., & Cheong, Y. (1992). Contextual effects on the self-perceived efficacy of high school teachers. *Sociology of Education*, 65, 150–167.

Rizzo, J. R., House, R. J., & Lirtzman, S. I. (1970). Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly*, 15, 150–163.

Schwab, R. L., & Iwanicki, E. F. (1982). Perceived role conflict, role ambiguity, and teacher burnout. *Educational Administration Quarterly*, 18, 60–74.

Schwab, R. L., Iwanicki, E. F., & Pierson, D. A. (1983). Assessing role conflict and role ambiguity: A cross validation study. *Educational and Psychological Measurement*, 43, 587–593.

Seidman, E. (1988). Back to the future, community psychology: Unfolding a theory of social intervention. *American Journal of Community Psychology*, 16, 3–24.

Sheldon, S. B. (2007). Improving student attendance with school, family and community partnerships. *Journal of Educational Research*, 100, 267–275.

Sirotnik, K. A., & Kimball, K. (1999). Standards for standards-based accountability systems. *Phi Delta Kappan*, 81, 209–214.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8–11.

Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education*, 51, 334–344.

Spradley, J. P. (1980). *Participant observation*. New York: Holt, Rinehart and Winston.

Standards and accountability (2001, January 11). *Education Week*, 20(17), 94–95.

Tatto, M. T. (2006). Education reform and the global regulation of teachers' education, development and work: A cross-cultural analysis. *International Journal of Education Research*, 45, 231–241.

Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Educational Research Journal*, 31, 231–262.

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

Urdan, T. C., & Paris, S. G. (1994). Teachers" perceptions of standardized achievement tests. *Educational Policy*, 8, 137–156.

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44, 519–558.

Webb, P. T. (2006). The choreography of accountability. *Journal of Educational Policy*, 21, 201–214.

Yeh, S. S. (2006). High-stakes testing: Can rapid assessment reduce the pressure? *Teachers College Record*, 108, 621–661.

York-Barr, J., & Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research*, 74, 255–316.

Appendix A: Accountability Policies Stressors and Supports Questionnaire

For me, policies to increase student academic performance:

1.   Result in my getting more support from the state
2.   Result in my getting more support from parents
3.   Result in my getting more support from the school district
4.   Result in my getting more support from my supervisor
5.   Result in my getting more support from the public
6.   Result in more pressure on me from the public
7.   Result in more pressure on me from my supervisor
8.   Result in more pressure on me from the state
9.   Result in more pressure on me from parents
10.   Result in more pressure on me from my school district
11.   Make me feel like my job performance is being fairly evaluated by the public
12.   Make me feel like my job performance is being fairly evaluated by my supervisor
13.   Make me feel like my job performance is being fairly evaluated by lawmakers
14.   Make me feel like my job performance is being fairly evaluated by my school district
15.   Help me focus more on what I should be doing
16.   Make me choose between what I think children need and what the state says I should do
17.   Help me more effectively impart knowledge
18.   Result in standards that make it difficult to teach students who have different abilities
19.   Overall put more stress in my job
20.   Overall ease the tasks I have in teaching
21.   Overall provide more support for me in my efforts to educate students
22.   Overall add a lot of burdens to my job
23.   Make me feel the state is helping me improve students' academic performance
24.   Set standards that just are not reachable for all students