

Middle School Reform Through Data and Dialogue: Collaborative Evaluation with 17 Leadership Teams

By: David Strahan, Jewell Cooper, Martha Ward

[Strahan, D.](#), [Cooper, J.](#) & Ward, M. (2001). Middle school reform through data and dialogue: Formative evaluation with seventeen leadership teams. *Evaluation Review*, 25(1), 72-99. DOI: 10.1177/0193841X0102500104

Made available courtesy of SAGE Publications: <http://dx.doi.org/10.1177/0193841X0102500104>

*****Reprinted with permission. No further reproduction is authorized without written permission from SAGE Publications. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document.*****

Abstract:

This report describes a 2-year, longitudinal study of one school district's effort to link site-based, collaborative evaluation with formal, centralized program evaluation. Participants formed a research team in partnership with a local university. Team members assisted leadership teams in identifying issues for informal, site-based assessments and then used the information to monitor progress toward established goals. Participants collaborated in designing and conducting in formal assessments of student achievement, school climate and safety, discipline, and parent involvement. Leadership teams used these data in developing their school improvement plans. Researchers and administrators used these data to revise the district wide survey. At the end of the 2-year cycle, analysis of school improvement plans showed that collaborative evaluation is creating a connection between dialogue and data. These 17 middle schools are approaching school improvement in a more integrated fashion by actively involving key stakeholders (students, parents, and teachers) in the evaluation process.

Article:

Good schools are symbolically rich places, where vivid and interesting conversations are taking place up and down the hierarchy. Adults are visibly engaged in inquiry, discovery, learning, collaborative problem solving, and critical thinking. Poor schools, by comparison, are symbolically impoverished; people are mum or secretive, isolated from one another or afraid to speak their minds.
—R. G. Brown (1991, 233)

As Brown (1991) found in his study of school reform, meaningful change often begins with language. "You have to help the people in the organization listen to themselves and raise questions about what they hear" (p. 235). His study and others have emphasized the importance of ongoing conversations about priorities and progress. Evaluation plays a key role in this process. Leaders in successful schools involve stakeholders in identifying shared goals and collecting information related to those goals. This integration of "data" and "dialogue" is essential to self-renewing schools (Fullan 1990; Joyce, Wolf, and Calhoun 1993; Elmore, Peterson, and McCarthy 1996; George and Shewey 1994). For the past 2 years, the 17 middle schools of Guilford County, North Carolina, have been developing a system of collaborative evaluation, linking data from formal and informal assessments with conversations among members of leadership teams to identify ways to improve schooling. This article reports the results of 2 years of data collection and discussion.

THEORETICAL PERSPECTIVE: THE DEVELOPMENT OF A COLLABORATIVE EVALUATION MODEL

Joyce, Wolf, and Calhoun (1993) characterize good schools as "self-renewing," that is, they feature continuous processes of collaboration, coordination, and setting priorities. These processes require consistent use of data to identify needs and monitor progress. Finding the most productive balance between the data needed to identify needs and the dialogue needed to make use of the data is a challenge in any program evaluation effort. On one hand, stakeholders need factual information to make judgments about the worth of programs and assess their

progress toward identified goals. On the other, stakeholders need to talk with each other about their perceptions of efforts and initiatives. Sometimes, decision makers want to rely on external evaluators and “objective” data to avoid the biases and personal interests inherent in reflective analyses. Other times, decision makers want to keep track of the personal impact of the issues under discussion. When carried to excess, an emphasis on either “just the facts, please” or “we know what’s best for us” can result in decisions that ignore the complexity of an educational endeavor.

As Stufflebeam (2000) has suggested, attempts to balance these conflicting tensions have resulted in the development of a number of different approaches to program evaluation over the past 40 years. Defining program evaluation as “a study designed and conducted to assist some audience to assess an object’s merit and worth” (p. 35), Stufflebeam describes 22 different approaches that have emerged in four categories that differ by purpose: “pseudoevaluations,” which do not meet expectations for research integrity; “questions/methods-oriented approaches,” which begin with prescribed expectations; “improvement/accountability-oriented” approaches, which assess merit and worth with both a priori assumptions and emergent issues; and “social agenda-directed/advocacy” approaches, which focus on the evolving needs and perceptions of stakeholders (p. 36). Although all of the useful approaches feature close links between the collection of data and the consideration of that data by participants, some approaches are more “participatory” than others in the extent to which decisions are shared.

Cousins and Earl (1992) trace the origins of “participatory” approaches to the “stakeholder-based” models of evaluation that emerged in the 1970s. They note that the essence of participatory approaches is the heightened involvement of primary stakeholders and a wider range of participation by major stakeholders (p. 399). They define participatory evaluation as “a partnership between trained evaluation personnel and practice-based decision makers” and argue that participatory evaluation combines the responsiveness of more conventional stakeholder evaluation with the technical rigor of more traditional external evaluation (p. 400). By reviewing evaluation reports from 26 participatory studies, they show that this approach is viable when it reflects organizational commitment, collaboration among participants, and training for evaluators (p. 414). Their conclusions echo Greene’s (1987) findings. In his analysis of two case studies of evaluations that featured high levels of stakeholder involvement, he found that to link participation and utilization, participants need to discuss evaluation information and reflect together on this information (p. 114).

More specifically, Cousins, Donohue, and Bloom (1996) have defined collaborative evaluation as “any evaluation in which there is a significant degree of collaboration or cooperation between evaluators and stakeholders in planning and/or conducting the evaluation” (p. 210). Recent work by O’Sullivan and O’Sullivan (1998) extends this concept to emphasize the role of stakeholders in the evaluation process. They demonstrate ways that stakeholders are often more than informants who provide information to the evaluation. Their “evaluation voices” approach encourages participants to “reconceptualize evaluation as a dynamic process that requires their active participation” (p. 22). True collaboration requires participants to share in decisions relating to the content and direction of both the program and the evaluation. A study of the impact of this approach with members of a partnership for improving early childhood education (O’Sullivan and D’Agostino 1998) documented ways that collaboration improved both the quality of information gathered and the sharing of this information with other stakeholders.

The purpose of the evaluation project reported here has been to promote the use of data through dialogue. In particular, we have tried to find ways to assist leadership teams in gathering data they wanted and in using the data to make decisions about school improvement. This, to us, is the essence of formative evaluation. As Scriven (1997) reminds us,

Formative evaluation is, to a large extent, best designed as summative evaluation of an early version, with particular attention to components of dimensions rather than a holistic account.... The role of formative evaluation is to provide feedback on midstream merit, as a service to assist program improvement, and given that the program itself is constantly evaluating midstream merit in an informal

way, what the professional independent evaluator brings to the party is a fresh eye and some technical skills. (Pp. 498-99)

Our notion of collaborative evaluation has been to provide fresh eyes and technical skills in ways that middle school teachers and administrators would find most useful.

THE NEED FOR COLLABORATIVE EVALUATION IN MIDDLE SCHOOL REFORM

Finding ways to connect data and dialogue is especially important in middle-level schools where the pace of change is often accelerated by the rapidly changing needs of students. In their survey of practices employed by the most successful middle schools, George and Shewey (1994) found that 71% of the schools reported “a regular and systematic process for evaluating the middle school program” (p. 88). They concluded that access to information is not enough; administrators and teachers need time to analyze information and discuss priorities.

Van Tassel-Baska, Hall, and Bailey (1996) described how structural changes occurred faster than changes in curriculum and instruction. They selected three schools that had been involved in reform efforts for at least 5 years from nominations from major national groups involved in the change process. They interviewed administrators, teachers, student, and parents; observed practices; and reviewed available data related to the three schools. Results show that each school accomplished structural changes such as flexible scheduling, teacher teaming, theme-based curriculum, heterogeneous grouping, cooperative learning, and inclusion (p. 108). Each school clearly articulated a rationale for these changes and demonstrated strong support for them. Although one of the schools seemed to be successful with project-based education, “actual instruction lagged behind the articulation of what instruction should be” in all three schools (p. 110). Researchers concluded that the biggest problem at all three sites is the “gulf between the articulation of practice and actual practice itself” (p. 110). Teachers talked about changes, yet researchers found little evidence of change, especially in regard to student learning. Although each school had gathered “perceptual data” indicating positive feelings, “there is little evidence that any structural changes made have improved student achievement” (p. 110). Van Tassel-Baska, Hall, and Bailey concluded,

One lesson that emerged is that changing a school’s philosophy and/or mission is only one step toward systemic change. Having a coherent mission, even under the guidance of a visionary leader, does not complete the school reform process. The level of change necessary needs to trickle into each classroom. This will not happen until curriculum and instruction are reformed in the same manner that structural organization has been reformed. (P. 111)

Their findings suggest that, even in schools recognized as “reformed,” meaningful changes in the day-to-day lives of students in classrooms may not be as prevalent as planners had wished.

Elmore, Peterson, and McCarthy (1996) reached almost exactly the same conclusions in their case studies of reform at the elementary level. Schools in their study had also made structural changes. They changed grouping practices, reorganized teachers into teams, and gave teachers more power in making decisions regarding budget and staff development. Like Van Tassel-Baska, Hall, and Bailey (1996), these researchers analyzed what teachers did as well as what they said. They found that although teachers talked about active student learning, many of them depended on repetition and recitation. They concluded that transforming teaching practice is “fundamentally a problem of enhancing individual knowledge and skill, not a problem of organizational structure” (Elmore, Peterson, and McCarthy 1996, 240).

These studies have suggested that meaningful improvements in teaching and learning require a “reculturing” of teaching, the development of new notions of who we are and how we teach. Brown’s (1991) study of the dynamics of change at seven different schools demonstrated the importance of teachers’ perceptions of empowerment. He noted that “the educational structures that need ‘restructuring’ are not just inside schools or districts; they are inside people’s heads” (p. 226). In this regard, improving practices may mean changing the culture of the school. In practical terms, school culture may be defined as “who we are” and “how we do things

around here” (Strahan 1994, 7). The concept of collaborative evaluation and the need for teacher empowerment provided a foundation for the model developed in this particular investigation. The following questions guided this research:

1. When provided support for gathering data, what issues do middle school leadership teams choose to explore?
2. How do they use this information to set priorities for school improvement?
3. How do they monitor progress toward these priorities?
4. How does the collaborative evaluation process inform the revision of the districtwide Middle School Survey?

COLLABORATIVE EVALUATION PROCEDURES

The Guilford County School District began a comprehensive evaluation of its 17 middle schools in 1996. The research office developed a detailed survey and administered it to teachers, students, and parents in the spring of 1996 and again in 1997. More than 13,000 respondents replied each year. Leadership teams in the 17 middle schools used these results, as well as achievement data, to focus school improvement plans. The district requires each school to submit a formal school improvement plan each year, following guidelines established by the state department of education. Reports must include summaries of performance on the state-mandated End-of-Grade (EOG) achievement tests and plans to enhance school safety.

In the fall of 1997, a team of principals and administrators decided to administer the survey every third year and provide technical assistance to leadership teams in the intervening years. This group identified a set of systemwide issues that had emerged from the first two Comprehensive Middle School Surveys: differentiated instruction, adviser/advisee programs, block scheduling/flexible scheduling, team organization and interdisciplinary teaming, collaboration between core and encore teachers, intramural programs, school safety, discipline, and parent involvement. These issues provided a basis for each leadership team to review its school improvement plans and report its priorities for evaluation.

THE COLLABORATIVE EVALUATION MODEL

The lead team of principals and administrators invited a research team from University of North Carolina at Greensboro (UNCG) to assist them in designing and conducting the next phase of the evaluation. Dr. David Strahan recruited three recent doctoral graduates with specialization in middle-level education to work directly with the schools. Each member of the research team assumed responsibility for coordinating efforts with 5 or 6 of the 17 middle school leadership teams. The process began with face-to-face meetings with each of the 17 leadership teams. At these meetings, members of the research team encouraged teachers, parents, and administrators to identify priorities for data collection and analysis. Based on these discussions, the research team proposed the model for collaborative evaluation presented in Figure 1. This model defines collaborative evaluation as the process that connects discussions with stakeholders, analysis of achievement data, development of school improvement plans, and revisions to the Comprehensive Middle School Survey.

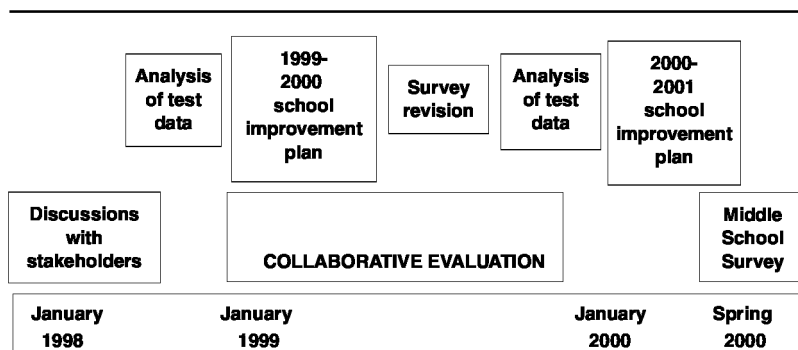


Figure 1: A Collaborative Evaluation Model for Guilford County Middle Schools

DATA COLLECTION AND ANALYSIS

To assist each team in connecting their analysis of achievement test data with issues they identified related to school improvement, the research team helped participants collect three major types of data:

1. **Small group interviews:** In many cases, the research team explored issues with students, teachers, and parents by meeting with small groups for semistructured interviews. Members of the research team asked teachers and administrators to recruit volunteers who were willing to discuss the issues to be addressed and who represented the demographic composition of the school. During the discussions, researchers rarely asked structured questions. Instead, they facilitated discussion with prompts such as “tell me about” or “is there anything else you would like to add.” Researchers recorded responses without using names or identifying indicators. They reported responses in summary fashion, using direct quotes only as illustrations.
2. **Informal surveys:** Another way that the research team gathered data was by developing informal surveys with members of the leadership team. They employed open-ended questions and simple Likert-type rating items to elicit perceptions from students, teachers, and parents regarding the issues to be explored. Researchers then tallied responses and provided general summaries of comments to share with the leadership team.
3. **Formal surveys:** During the second year of the study, a number of teams expressed a need to respond more systematically to a new state mandate to document perceptions of school safety. The research team identified a survey that could gauge students’ perceptions of school safety in regard to school climate and classroom practices. Developed by Purkey and Lehr (1996), this instrument assesses the extent to which students feel safe both physically and psychologically in their specific classes and at school in general. In their analysis of the psychometric properties of this instrument, Shoffner and Vacc (1999) found that 24 of the original 50 items fit a strong factor structure. These items yielded four factors (teachers and staff, places and policies, stressors and discomforts, and attitudes toward the school environment). These factors accounted for 41% of the variance (eigenvalues of 4.4, 2.1, 1.9, and 1.4) and showed adequate internal consistency (Cronbach alphas from .61 to .73). They recommended the use of these 24 items. The research team shared the revised 24-item instrument with principals and asked for feedback regarding readability from 50 sixth-grade students. Based on their feedback, the survey was revised to modify wording and include items that tapped perceptions of classroom learning. Leadership teams from three schools decided to use the resulting “Middle School Inviting School Safety.” The team also assisted administrators in revising the districtwide Middle School Survey, incorporating items from the school safety survey and offering suggestions based on responses from participants during the collaborative evaluation process. The final section of this article reports the survey development process for the Middle School Survey.

Each school team began collecting data in the spring of 1998. That summer, members of the research team worked with members of each leadership team to analyze the information collected and drafted a first-year report to the Board of Education. During the fall of 1998, members of the research team began a new round of conversations with leadership teams, helped them refine and/or redefine their priorities for evaluation, and assisted them with collecting data related to their progress toward their goals. During the summer of 1999, the research team completed its second-year report and offered suggestions for revising the districtwide Comprehensive Middle School Survey based on findings from the collaborative evaluation. The revised survey was administered in the spring of 2000.

RESULTS

ISSUES IDENTIFIED DURING THE FIRST YEAR (1997-1998)

During the first year of collaborative evaluation, the 17 leadership teams identified priorities in four general areas of school improvement. Nine schools decided to examine their efforts to promote student achievement. Four schools selected schoolwide discipline as their top priority. Two schools selected school safety. Two

schools decided to focus on parental involvement. Table 1 presents a list of the topics selected and data sources employed for the first and second years of this study.

Of the nine leadership teams focusing on student achievement, four of these schools selected “differentiated instruction” as the specific focus of their study. Conversations with leadership teams at these schools suggested a concern about the extent to which students experienced a wide range of instructional activities. Members of these leadership teams wanted to know whether lessons offered students activities that engaged varied ways of understanding. Members of the research team developed an informal survey (the “Instructional Survey”) for teachers to use to describe the configuration of activities they offered. At all four schools, teachers reported that the most frequently used activities were cooperative learning, paper/pencil worksheets, reading assignments, and advanced organizers. Least frequently used activities included enactments, role-playing, spreadsheets, and databases.

At two of these four schools, the research team drafted an additional informal survey to explore students’ perceptions of instructional variety. Responses from 909 students from the two schools suggested two clusters of student ratings. “Traditional” activities included reading books and answering questions, taking notes while the teacher talks, having discussions, writing papers, and doing practice pages and worksheets. “Varied” activities included watching videos, working in small groups, making projects, working with computers, acting things out, working with a partner, doing experiments, listening to music, and playing games. Students at both schools rated the traditional activities as most frequently used (with the exception of “working with a partner”). At the four schools studying differentiated instruction, results from these surveys have encouraged leadership teams to examine ways to incorporate varied activities more effectively.

At the fifth school, a team of language arts teachers analyzed scores from the statewide assessment of writing performance over the past 3 years and developed an improvement plan to share with colleagues. The leadership team at the sixth school studied the impact of two staff development efforts by examining data from teachers’ journals and focus groups. Teachers and administrators at the seventh school identified the school’s lowest performing readers and worked with them individually for 3 months. Analysis of reading logs, achievement test scores (from the NC assessment system), and interviews demonstrated that students and teachers found this program to be very successful. The leadership team at the eighth school explored the perceptions of students who had previously done poorly in school and were now doing better. A member of the research team conducted informal group interviews with 15 students identified by their teachers as having improved their schoolwork dramatically over the past year (and with permission from parents for interviews). These students attributed much of their success to the caring of teachers and parents. Specific interventions that they found productive included encouraging student questioning, tutoring, feedback sheets on assignments, and conferences with students and parents. The leadership team at the ninth school decided to try to improve achievement by improving attendance. The research team gathered ideas from the professional literature and shared five major suggestions for improvement: monitoring chronic absentees from year to year, making early interventions with parents, tapping the resources of social services, proactive incentives/clubs, and counseling.

Four of the leadership teams focused their inquiry on issues related to schoolwide discipline. To explore these issues, the research team drafted an informal survey (“Classroom Management Survey”) to get a general description of how teachers approach problems. Researchers then met with small groups of teachers to discuss their specific concerns. Although the concerns they expressed varied by school and by grade level, teachers in all three schools wanted to find ways to clarify expectations for student behavior, especially in the hallways, lunchrooms, and bus areas. They felt that students would benefit from more consistent enforcement of policies. They particularly wanted to improve communication between administrators and teachers in regard to discipline. At one school, members of the leadership team recruited volunteers to work with the research team to explore options for improving schoolwide discipline. This study team developed a set of procedures and shared them with their colleagues in a work session.

TABLE 1: Guilford County Middle Schools by Evaluation Topics, 1998-1999

| <i>School</i> | <i>Spring 1998 Focus and Data Sources</i> | <i>Spring 1999 Focus and Data Sources</i> |
|---------------|--|---|
| Washington | Achievement/differentiated instruction Instructional Survey—teachers | Achievement/differentiated instruction Instructional Survey—teachers and students |
| Adams | Achievement/differentiated instruction Instructional Survey—teachers | <i>Achievement/literacy</i> <i>Test scores</i> |
| Madison | Achievement/differentiated instruction Instructional Survey—teachers and students | Achievement/differentiated instruction Small group interviews—teacher and students |
| Jefferson | Achievement/differentiated instruction Instructional Survey—teachers and students | <i>School climate and safety</i> <i>School Climate and Safety Survey—teacher and students</i> |
| Monroe | Achievement/writing Action research | Achievement/writing <i>Small group interviews</i> |
| Polk | Achievement/staff development Small group interviews and journals | Achievement/differentiated instruction Small group interviews—teachers and students |
| Van Buren | Achievement/literacy Test scores and interviews | Achievement/literacy Test scores and interviews |
| Tyler | Achievement/students at risk Small group interviews | <i>Teacher/student advisory program</i> <i>Teacher/Student Advisory Survey—students</i> |
| Johnson | Attendance Review of research | <i>School climate and safety</i> <i>School Climate and Safety Survey—teachers and students</i> |
| Grant | Schoolwide discipline Classroom Management Survey | <i>General school improvement</i> <i>General School Improvement Survey</i> |
| Hayes | Schoolwide discipline Classroom Management Survey | <i>General school improvement</i> <i>Small group interviews</i> |
| Garfield | Schoolwide discipline Classroom Management Survey | <i>Achievement/literacy</i> <i>Test scores</i> |
| Cleveland | Schoolwide discipline Study team | <i>School climate and discipline</i> <i>Small group interviews</i> |

TABLE 1 Continued

| <i>School</i> | <i>Spring 1998 Focus and Data Sources</i> | <i>Spring 1999 Focus and Data Sources</i> |
|---------------|---|---|
| KcKinley | School safety Focus groups | <i>Parent involvement</i> <i>Parent Involvement Survey</i> |
| Roosevelt | School safety Small group interviews | <i>Differentiated instruction</i> <i>Small group interviews—teachers and students</i> |
| Taft | Parent involvement Small group interviews | <i>General school improvement</i> <i>General School Improvement Survey—students, teachers, and parents</i> |
| Wilson | Parent Involvement Survey | <i>School Climate and Safety</i> <i>School Climate and Safety Survey—teachers and students</i> |

NOTE: Notations in italics indicate changes in emphasis from the first year to the second.

At the two schools investigating issues related to school safety, researchers conducted informal group interviews with sixth-, seventh-, and eighth-grade students. Participants defined school safety as protection from violence and threats of violence, especially in terms of fights. Both leadership teams then began examining ways to respond to students' concerns on an ongoing basis.

At one of the two schools focusing on parental involvement, an informal group interview with parents on the leadership teams explored their perceptions of the middle school concept. These particular parents were very supportive of the "student-centered" emphasis of the middle school concept in general and suggested ways to improve communication. The leadership team at the other school drafted an informal survey to assess the school's efforts toward parental involvement. Parents responding to the survey highlighted several initiatives as successful (open house, PTA meetings, newsletters, parent conferences), whereas responses to other endeavors varied.

ISSUES IDENTIFIED DURING THE SECOND YEAR (1998-1999)

Members of the research team shared their reports with leadership teams during the summer of 1998. During the fall semester, researchers visited with each team to discuss responses to the 1998 reports and plans for data collection in the spring of 1999. Table 1 reports the topics each team selected for collaborative evaluation and the data sources employed.

These listings show that 6 of the 17 middle schools enriched or extended their focus for the second year of this project. One of these teams (Van Buren) decided to maintain the emphasis on literacy and continue with the same sources of data. Another team (Washington) decided to repeat the instructional survey with teachers and add the student survey to their plan. Four other teams (Madison, Monroe, Polk, and Cleveland) continued to focus on achievement with more emphasis on data from small group interviews.

Table 1 also shows that 11 of the 17 middle school leadership teams shifted their emphasis for the second year. Given a new state mandate to gather information regarding students' perceptions of school safety, three of the schools (Jefferson, Johnson, and Wilson) decided to examine students' perceptions of school climate and school safety in a connected fashion. A total of 1,957 students from the three schools completed the Middle School Inviting School Safety. Three other teams (Grant, Hayes, and Taft) decided to focus on school improvement in a more holistic sense. The research team helped the leadership team at Grant draft an open-ended questionnaire for teachers regarding middle school practices. At Hayes, the team conducted informal interviews with small groups of students and teachers regarding middle school practices. Students, parents, and teachers at Taft completed open-ended surveys focusing on the school as a whole. The other three teams (Tyler, McKinley, and Roosevelt) selected different topics for emphasis. At Tyler, the research team helped the leadership team conduct an open-ended questionnaire with teachers regarding their classroom-based guidance practices. Parents at McKinley responded to open-ended questions regarding their involvement with the school. The research team conducted small group interviews with students and teachers at Roosevelt to discuss perceptions of differentiated instruction. Two schools (Adams and Garfield) decided to focus entirely on the collections of the EOG achievement data provided by the state.

In general, all of the 17 teams used the information from their first year reports as one of several considerations in refining their school improvement plans for the second year. At the end of the second year, the research team conducted a detailed analysis of all 17 formal school improvement plans submitted to the district by each of the leadership teams. Table 2 provides a summary of the priorities for school improvement identified by each of the school improvement teams in the 1999-2001 plans they submitted to the district.

As Table 2 indicates, all 17 teams identified priority goals for instructional improvement in math and writing, 16 of the teams in reading. All 17 teams identified school safety as another priority for improvement. The other areas for improvement that the teams targeted most frequently were advisory programs/character education (15), schoolwide discipline (15), and parental involvement (15). These results show that the teams continued to

focus on the major topics of concern they identified in their collaborative evaluations: achievement, discipline, safety, and parental involvement.

Analysis of the formal school improvement plans also documented the types of data each team planned to use to monitor progress toward these priorities. Table 3 provides a summary of the data sources listed.

TABLE 2: Priority Goals Documented in School Improvement Plans for 1999-2000

| School | Specific instructional improvements in the areas of | | | Middle school development in | | | |
|------------|---|------|---------|------------------------------|------------------------------|-----------------------|----------------------|
| | Reading | Math | Writing | School Safety | Advisory/Character Education | Schoolwide Discipline | Parental Involvement |
| Washington | √ | √ | √ | √ | √ | √ | √ |
| Adams | √ | √ | √ | √ | √ | √ | √ |
| Madison | √ | √ | √ | √ | | √ | √ |
| Jefferson | √ | √ | √ | √ | √ | √ | √ |
| Monroe | √ | √ | √ | √ | √ | √ | √ |
| Polk | √ | √ | √ | √ | | √ | √ |
| Van Buren | √ | √ | √ | √ | √ | √ | √ |
| Tyler | √ | √ | √ | √ | √ | √ | √ |
| Johnson | | √ | √ | √ | √ | | √ |
| Grant | √ | √ | √ | √ | √ | | √ |
| Hayes | √ | √ | √ | √ | √ | √ | √ |
| Garfield | √ | √ | √ | √ | √ | √ | √ |
| Cleveland | √ | √ | √ | √ | √ | √ | |
| KcKinley | √ | √ | √ | √ | √ | √ | √ |
| Roosevelt | √ | √ | √ | √ | √ | √ | √ |
| Taft | √ | √ | √ | √ | √ | √ | |
| Wilson | √ | √ | √ | √ | √ | √ | √ |

TABLE 3: Data Sources Documented in School Improvement Plans

| School | NC End-of-Grade Test Scores | | | School Safety Survey | Site-Based Sources of Data | | |
|------------|-----------------------------|------|---------|----------------------|----------------------------|------------------|--------------------------|
| | Reading | Math | Writing | | Needs Assessment | Group Interviews | Reading Logs and Records |
| Washington | √ | √ | √ | √ | √ | | |
| Adams | √ | √ | √ | | | | √ |
| Madison | √ | √ | √ | √ | | | |
| Jefferson | √ | √ | √ | √ | | | √ |
| Monroe | √ | √ | √ | √ | | | √ |
| Polk | √ | √ | √ | √ | | | √ |
| Van Buren | √ | √ | √ | | √ | | √ |
| Tyler | √ | √ | √ | √ | | | √ |
| Johnson | √ | √ | √ | | | | |
| Grant | √ | √ | √ | √ | | | √ |
| Hayes | √ | √ | √ | √ | | | √ |
| Garfield | √ | √ | √ | √ | | | |
| Cleveland | √ | √ | √ | √ | | √ | |
| KcKinley | √ | √ | √ | √ | √ | √ | |
| Roosevelt | √ | √ | √ | √ | | | |
| Taft | √ | √ | √ | | √ | | |
| Wilson | √ | √ | √ | √ | √ | | |

Table 3 shows that all 17 teams placed a high premium on the use of EOG test score data to assess progress in academic areas. Thirteen of the teams planned to use some type of school safety survey to evaluate perceptions of safety. In addition to these data sources provided by the district, 13 of the teams also planned to use site-

based sources of data. Five planned to conduct their own needs assessments, 2 planned to conduct group interviews with students and parents, and 8 described ways that they would use data collected from students in reading logs or reading records to monitor students' reading performances.

Although summarizing the school improvement plans provides one perspective on the collaborative evaluation process, the rest of the story lies in the specific ways that school leadership teams are using data. In the section that follows, illustrations from one of these teams will provide a more detailed picture of the uses of information in setting priorities for school improvement and monitoring progress.

USES OF INFORMATION: VAN BUREN MIDDLE SCHOOL

The leadership team at Van Buren Middle School began its collaborative evaluation by analyzing responses from students, parents, and teachers to the districtwide surveys of middle schools administered in 1996 and 1997. Issues that emerged from this analysis included concerns regarding student performance and discipline. Based on their analysis of these issues, the leadership team decided that their major goal for 1997-1998 would be to improve the academic performance of students who had not yet been successful. In particular, they hoped to "shrink the 43%," that is, reduce the number of students who were not performing at statewide proficiency levels.

As part of their partnership with UNCG, a group of 17 teachers and administrators began meeting with university faculty to design and implement a tutorial program for nonreaders. They worked with 17 individual students three times per week throughout the spring of 1998. In the fall of 1998, preservice teachers completed this same training and began working with cooperating teachers who were now experienced reading tutors. Organized in collaborative teams, preservice teachers, cooperating teachers, and university faculty began tutoring a new group of students and exploring ways to integrate reading instruction across the curriculum.

In all phases of data collection, researchers kept field notes that documented planning meetings and personal journals that recorded observations and impressions. The team met periodically to generate themes. These themes became the organizers for a series of small group interviews with teachers and students. The research team synthesized data from observations with results from achievement testing. They shared these preliminary reports with administrators and then prepared final reports.

Before they began this tutoring project, participating teachers expressed four shared perceptions of teaching basic reading at the middle level. With little variation, they believed that (a) the teacher plays a critical role in the process, (b) students face serious obstacles (lacking basic skills, lacking motivation), (c) they would need to establish trusting relationships with students, and (d) teaching basic reading would be very complicated. By the end of the first semester, they suggested that personal relationships with students were even more important than they thought and that students were more eager to learn than they had realized. The biggest surprise reported was that teaching reading was not as technical as they had assumed. As suggested by one respondent, "This was always 'do-able.' It felt like we were taking a step at a time. That's the real beauty of the program ... it worked for me and it worked for the children, too." Participants looked forward to sharing these insights with preservice teachers in the fall.

EOG reading scores administered by the state of North Carolina showed that 15 students had pre- and posttests; 11 made progress (scores rose), three scores declined slightly, and one remained the same. The group's average rose almost four points (from 136.0 to 139.8).

When asked to list two things they liked about the program, students listed "improved reading skills" (7), their "tutor" (5), and "enjoying opportunities to read" (4) as their top choices. They noted that they liked these things because they "felt comfortable" (6) and "improved reading skills" (4). Fourteen of the 15 students interviewed felt that they were reading better now than they were when they started the program. They attributed this to "learning how to figure out and pronounce words" (12). Eleven of the 15 suggested they would change "nothing" about the program. Two wanted "harder books," and 2 wanted to "increase tutoring time." Fourteen

of the 15 would like to participate in the program next year because they “improved reading/vocabulary” (11) and “enjoy reading” (2).

Teachers cited many ways that they had observed student progress: reading more difficult books (higher readability numbers), reading more words per minute, more patience figuring out words, more involvement and fewer disruptions in class, more awareness of the reasons for schooling, and more self-confidence. Teachers attributed these gains to one-to-one teaching, individualized attention, books students can read successfully/books that get harder, and an inviting after-school setting.

As the collaborative evaluation process continued at Van Buren during the 1998-1999 and 1999-2000 school years, participants continued to collect data about the progress of students in the tutoring program and expanded their focus to embrace an investigation of instructional strategies. During summer and fall of 1999, members of the research team assisted teachers in developing a set of instructional strategies to encourage more independent learning among students. Participants aligned these with the newly revised Guilford County Prioritized Curriculum. In October 1999, the team distributed the strategies to teachers at Van Buren Middle School and met with them in grade-level planning meetings to discuss implementation. Participants decided to field-test some of the strategies by observing lessons and interviewing students in the classrooms of teachers who volunteered.

During the months of November and December, research assistants conducted a total of 10 classroom visits. Prior to the observation, team members asked teachers to select 3 students for interviews who often had the most difficulty with reading lessons (being sure to include both boys and girls). The 30 students interviewed provided a representative group of those most likely to struggle with strategy lessons. During each of these visits, the research assistants gathered the following data:

1. Lesson observation—Using a three-column observation guide, research assistants recorded descriptions of “instructional practices” (how the teacher introduced the lesson, how the teacher gave directions given to students, which activities the teacher employed, and what assignments students completed for practice), “student responses” (what students did in response to these practices), and an estimate of the number of students on task during that portion of the lesson.
2. Student interviews—At the end of the class period, research assistants met with selected students individually, asking them the following questions:
 - What have you been studying in this class for the past week or two?
 - Tell me what you did during this lesson.
 - Tell me more about the lesson.
 - What helped you understand the lesson?
 - What do you think your teacher wanted you to learn today?
 - Was the instruction today typical of what happens in the class? How?
 - How did you feel while doing the lesson? Did you like it?
3. Teacher interview—Later that same day, research assistants met with the teacher and asked the following questions:
 - Tell me what you did during the lesson.
 - How do you think your students responded to the lesson?
 - How did you feel while doing the lesson? Did you like it?
 - Did you modify the lesson in any way? If so, how?

Observers took notes during the interviews to record the general responses.

TABLE 4: Frequencies of Student Responses to Interviews Following Strategy Lessons

| | <i>Lessons That Followed the Basic Sequence</i> | <i>Lessons That Did Not</i> |
|--|---|-----------------------------|
| Students who could describe in specific terms what they did during the lesson | 20 of 21 (95%) | 7 of 9 (78%) |
| Students who could articulate their understanding of the lesson and how they understood it | 21 of 21 (100%) | 5 of 9 (56%) |
| Students who expressed positive responses toward the lesson | 19 of 21 (90%) | 7 of 9 (78%) |

The 10 lessons observed featured five different strategies (six with graphic organizers, one each with cubing, key words, figurative language, and inference). Observation reports indicated that 7 of the lessons followed the basic strategy sequence (explain and demonstrate the strategy, give students guided practice, reflect on the effectiveness of the strategy). Three of the lessons omitted two of the steps. Average percentages of time on task ranged from 83 % to 94%. Although the average percentage of time on task was similar for the lessons that followed the sequence and those that did not (89%, 90%), there were noticeable differences in student responses in interviews. In the lessons that followed the basic sequence, 95% of the students interviewed could describe in specific terms what they did during the lesson (Questions 1, 2, 5). In the lessons that did not completely follow the sequence, only 78% of the students could do so. In the lessons that followed the basic sequence, 100% of the students could articulate their understanding of the lesson and tell how they understood it, in contrast to 56% articulation in the lessons that did not completely follow the sequence. In the lessons that followed the basic sequence, 90% of the students responded positively to the lesson format. In the remaining lessons, 78% of the students responded positively (see Table 4).

These data suggested that almost all of the targeted students in these 10 classes can employ and articulate the strategies developed when they are taught using the basic strategy sequence (explain and demonstrate the strategy, give students guided practice, reflect on the effectiveness of the strategy). The potential for strategic learning seemed strong enough to begin developing assessments of strategic learning to serve as “benchmarks.”

A team of teachers began collaborating with the research team to accomplish this task in January. The fact that three of the lessons did not employ the full sequence (even when being observed) suggests that participants need to offer better staff development in how to use the strategies. Plans are under way to meet with teachers in small groups during the spring to facilitate strategy instruction.

COLLABORATIVE EVALUATION AS A PROCESS TO GUIDE SURVEY DEVELOPMENT

As reported earlier, the Guilford County School District began a comprehensive evaluation of its 17 middle schools in 1996. Three years earlier, three districts had merged to form the Guilford County District. Having approved a middle school plan to unify programs across all 17 middle schools, the Board of Education wanted to assess the extent to which students, teachers, and parents perceived the program as effective. The first step in the evaluation was the development of the Middle School Survey. A team of administrators from the Office of Assessment and Evaluation and the Curriculum Division generated items for surveys for students, parents, and teachers based on surveys used previously by one of the three districts. Items on these surveys addressed perceptions of the basic elements of middle schooling: instruction, teaming, advisory programs, intramural programs, and parental involvement. The survey development team solicited input from middle school consultants working with the district, middle school principals, and members of the community to be sure that items addressed the goals of the program and concerns of the school board. The district administered this survey to teachers, students, and parents in the spring of 1996 and again in 1997. Results assisted leadership teams in the 17 middle schools in developing their school improvement plans. As it evolved, the collaborative evaluation process provided information that helped guide revisions to this survey.

Revision of the survey began in the fall of 1997 when the Office of Assessment and Evaluation convened a development team composed of middle school principals and consultants from UNCG to map out a 3-year evaluation cycle. This group identified a set of systemwide issues that had emerged from the first two Comprehensive Middle School Surveys: differentiated instruction, adviser/advisee programs, block scheduling/flexible scheduling, team organization and interdisciplinary teaming, collaboration between core and encore teachers, intramural programs, school safety, discipline, and parent involvement. These issues provided a basis for examining the original survey and suggesting revisions. The development team completed the revisions in two phases.

Phase 1: content analysis. The next step in the revision process was to review a structural equation modeling (SEM) study conducted by the Office of Assessment and Evaluation. In a report titled “Structural Models for Responses to the Middle School Survey” (Guilford County Schools 1999), staff members reported results from a statistical examination of the items from the 1997 Comprehensive Middle School Survey that identified patterns of responses and clusters of connected items. The purpose of this analysis was to address two issues that surfaced in discussions with leadership teams. First, members of the leadership teams reported that teachers, students, and parents had commented that the survey was too long. Second, the reporting format made it difficult to analyze trends over time. Each team received a printout that reported results by item, and the only way to assess changes from year was to go item by item. Members of the leadership teams wanted to be able to analyze results by clusters of items.

To determine whether items could be eliminated from the survey and results reported by cluster, staff members employed a combination of factor analysis and SEM procedures to analyze all of the responses from the 1996 and 1997 surveys. In 1996, 11,281 students, 6,938 parents, and 857 teachers completed their respective forms of the survey. In 1997, 11,774 students, 6,429 parents, and 878 teachers responded. Figures 2, 3, and 4 report the results of the SEM for all three forms of the survey.

Factor analyses of responses indicated 10 factors for the teacher survey, 9 factors for the student survey, and 7 factors for the parent survey. These “Level 1” factors were subsequently used to build structural models, which related the Level 1 factors to Level 2 factors. These 10 Level 1 factors were related hierarchically to 3 Level 2 factors: (a) The Middle School Concept, (b) School Climate, and (c) Instruction and Learning. All 3 of the Level 2 factors were related to the central construct, “Middle Schools.” Similarly, for the student form of the survey, the 9 Level 1 factors were related hierarchically to 2 Level 2 factors: (a) School Climate and (b) Instruction and Learning. Finally, for the parent survey, the 7 Level 1 factors were related to 2 Level 2 factors: (a) Instruction and Learning and (b) Middle School. The SEM analyses indicated that eliminating items that did not contribute statistically to precision in measuring the factors could strengthen some of the relationships among these factors. The authors of the 1999 report suggested that a total of 28 items should be omitted from the teacher survey, 22 from the student survey, and 33 from the parent survey.

The development team analyzed the results of this study in relationship to the issues originally identified for evaluation and the issues that emerged from 2 years of formative evaluation. Table 5 shows the relationships among these topics and the resulting framework for the development of the survey. As indicated, the topics recommended as organizers for the surveys reflected the organizers suggested by the structural model as well as the topics identified in the 1997 survey and the formative evaluations that have followed it. With these topics in mind, the team then analyzed each item remaining in the survey and each item recommended for deletion. This analysis indicated a need to replace several items to strengthen the revised topic organizers. This analysis also indicated a need to add several items that had not been included in the 1997 survey. Several items were drawn from surveys that proved effective in the formative evaluation, and others were written based on issues addressed in school safety surveys administered by some of the middle schools.

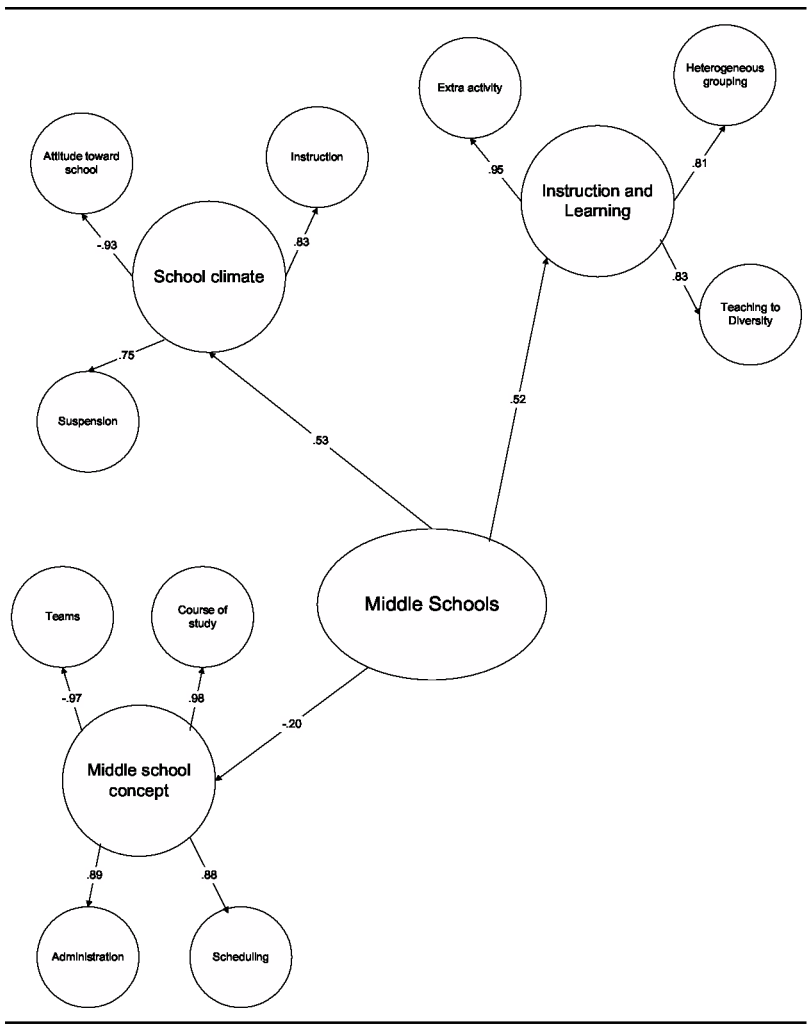


Figure 2: Structural Model for Teachers' Responses to the Middle School Survey

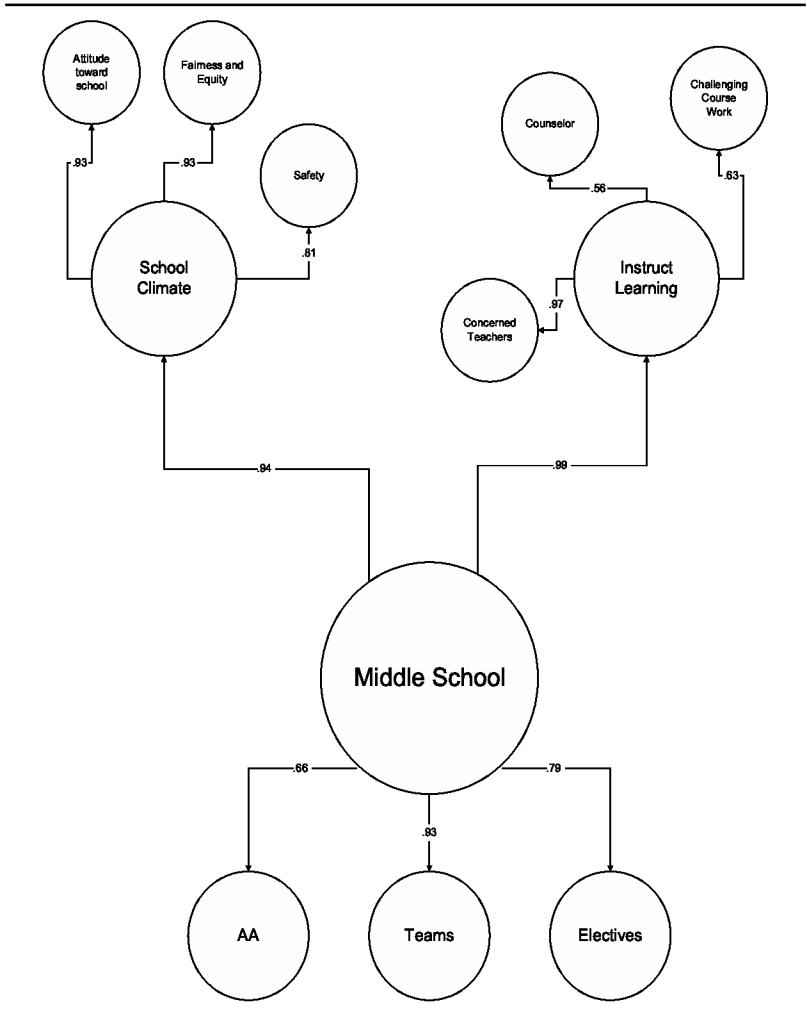


Figure 3: Structural Model for Students' Responses to the Middle School Survey

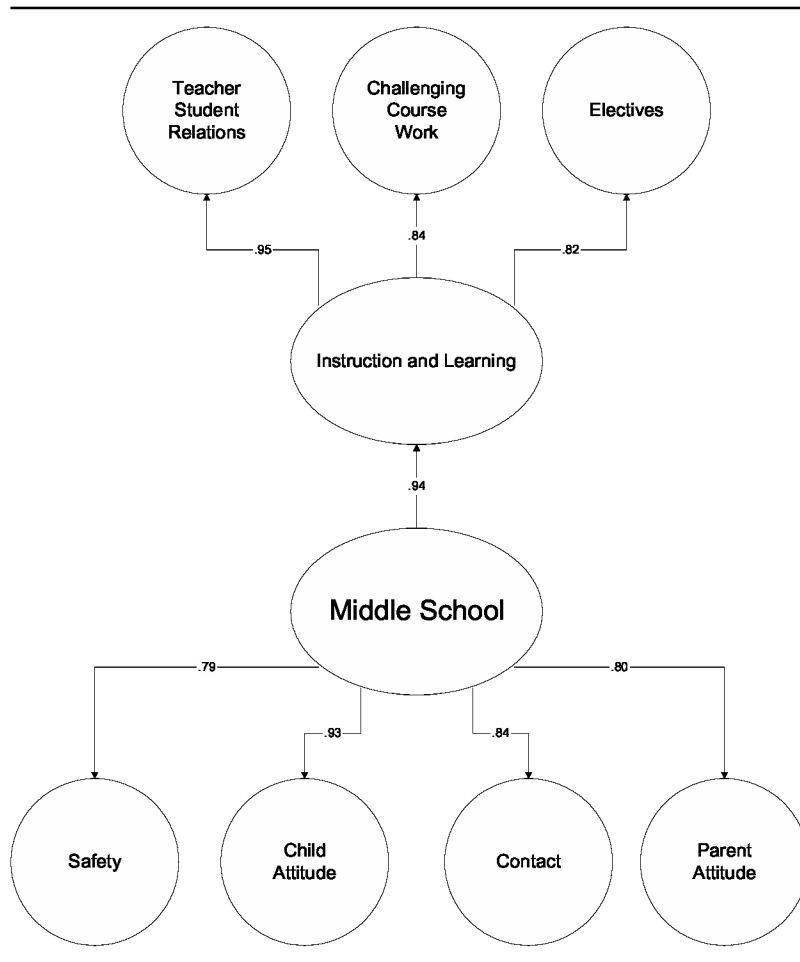


Figure 4: Structural Model for Parents' Responses to the Middle School Survey

Phase 2: review by stakeholders. The UNCG research team solicited input from students, parents, teachers, and administrators at three middle schools selected to represent an approximate mix of the urban, suburban, and rural schools in the district. During a 2-week period in September 1999, members of the research team conducted focus groups and interviews with 20 students, 10 teachers, and 8 parents. Interviewers asked participants to read each cluster of items and answer these questions: (a) Are these items clear to you? (b) Are there any confusing words in these items? (c) Are these items meaningful and useful to you? Interviewers asked each group to review the items suggested for omission and identify any item they thought the team should reconsider. Interviewers also asked them to identify any topic or concern that was not addressed by any of the items. The research team then made a final set of revisions to the surveys, replacing a total of 10 items across the three surveys and adding 4 new items suggested by participants. One change in wording was noted. Five items recommended for replacement or addition were not added because they seemed specific to issues at one of the schools and not as relevant across the entire district.

In summary, the 1997 Comprehensive Middle School Survey was taken apart item by item, reconstructed in a structural model, and revised to meet the needs of stakeholders. First, the framework from statistical modeling was revised to align with issues identified in the original model and the formative evaluations conducted over the past 2 years. Then, the research team recommended items for replacement and addition to strengthen this framework. Finally, students, parents, teachers, and administrators reviewed the surveys and recommended changes. Consequently, the survey was improved in the following ways:

1. Restructured to create clusters of items that more clearly address the needs of stakeholders and that will simplify interpretation.

2. Streamlined to reduce unnecessary items. The teachers' survey is 13 items shorter, the students' survey is 9 items shorter, and the parents' survey is 18 items shorter.
3. Organized in a way that will facilitate connections with outcomes on achievement tests.

TABLE 5: The Identification of Topics to Be Addressed in the 2000 Survey

| <i>Topics Identified in the 1997 Survey</i> | <i>Topics Identified in Formative Evaluation</i> | <i>Topics Identified in Structural Models</i> | <i>Recommended Structure for Teacher Survey</i> | <i>Recommended Structure for Student Survey</i> | <i>Recommended Structure for Parent Survey</i> |
|---|--|---|---|---|--|
| Differentiated instruction Core/encore | Instructional improvement | Instruction and learning Teachers/teaching Extra activities/challenging courses Heterogeneous grouping counselor | Instruction and learning Teaching to diversity Electives and extra activities Heterogeneous grouping Instruction | Instruction and learning Supportive teachers Electives and extra activities Challenging coursework Counselor | Instruction and learning Teacher/student relations Electives and extra activities Challenging coursework |
| School safety | School safety | School climate Safety Instruction Fairness | School climate School safety and attitude | School climate Safety | School climate Safety |
| Discipline | Schoolwide discipline | In-school suspension Student's attitude Parental involvement | Discipline and in-school suspension | Fairness and equity | Child's attitude toward school |
| Parental involvement | Parental involvement | | | Attitude toward school | Parent involvement |
| Teaming Scheduling | Teaming Comprehensive middle school development | Middle school concept Teams Scheduling | Middle school concept Teams Scheduling | Teams | Parent attitude |
| Adviser/advisee Intramurals | | Course of study/ adviser/advisee Administration | Course of study Implementation | Adviser/advisee | Contact |

Finally, the research team recommended that results be reported by category to help school improvement teams interpret responses. Because the items that comprise the categories were not identical to those of the 1997 survey, exact comparisons would not be possible. Revised surveys composed of 78 items for teachers, 70 items for students, and 42 items for parents were administered in the spring of 2000. The Office of Assessment and Evaluation is currently studying the results from that survey.

CONCLUSIONS

Results from this study suggest that leadership teams in Guilford County middle schools are beginning to operationalize the concept of collaborative evaluation. Leadership teams have identified issues to investigate and developed ways to gather data regarding these issues. During the first year of the process, the issues they identified reflect four major concerns: improving achievement, enhancing discipline, maintaining safety, and involving parents. Although teams identified similar issues in the second year, some of the teams addressed these issues in more integrated ways, linking school safety with school climate, addressing school improvement more holistically, and placing more emphasis on perceptions of stakeholders in small group interviews. From the beginning, all of the teams have involved key stakeholders (students, parents, and/or teachers) in discussing and clarifying these issues. As a result, dialogue regarding data is beginning at each of these schools. This dialogue informed the revision of the Comprehensive Middle School Survey and should be enhanced by results from that survey as the leadership teams review them this year.

From the perspective of central administration, this process is linking districtwide evaluation of middle school programs with ongoing, site-based, formative evaluations of the aspects of the program selected by each leadership team. Middle school principals have reported that teams at their schools are using survey results and collaborative evaluation data in developing their school improvement plans. Annual reviews of these plans

indicate that the leadership teams are establishing clearer priorities and monitoring progress more systematically than they were several years ago.

Although exploratory in nature, this study suggests implications for the process of collaborative evaluation. As Greene (1987), Cousins and Earl (1992), and Cousins, Donohue, and Bloom (1996) discovered in their investigations, opportunities for participants to discuss evaluation information and reflect together on this information are essential to collaboration. Participants in this project seem to be finding ways to complement the “participatory” involvement of primary stakeholders with perspectives from “external evaluators.” Leadership teams are using the information gathered to guide their plans for school improvement. In those plans, they are specifying ways they will use data to monitor progress toward their goals. They are identifying ways to gather data that fit their needs, using formal surveys and structured interviews when they wish to assess changes over time, and developing informal measures to explore issues when they wish to understand contextual dynamics. Although it is too early to assess the extent to which these practices will result in changes in schooling that are substantive and sustained, it would seem that dialogue regarding data is becoming a more powerful dimension of the school improvement process at these schools.

References

- Brown, R. G. 1991. *Schools of thought*. San Francisco: Jossey-Bass.
- Cousins, J. B., J. J. Donohue, and G. A. Bloom. 1996. Collaborative evaluation in North America: Evaluators self-reported opinions, practices, and consequences. *Evaluation Practice* 17 (3): 207-26.
- Cousins, J. B., and L. M. Earl. 1992. The case for participatory evaluation. *Educational Evaluation and Policy Analysis* 14 (4): 397-418.
- Elmore, R., P. Peterson, and S. McCarthy. 1996. *Restructuring in the classroom: Teaching, learning, and school organization*. San Francisco: Jossey-Bass.
- Fullan, M. 1990. Staff development, innovation, and institutional development. In *Changing school culture through staff development*, edited by B. Joyce, 3-25. Alexandria, VA: ASCD.
- George, P., and K. Shewey. 1994. *New evidence for the middle school*. Columbus, OH: NMSA.
- Greene, J. G. 1987. Stakeholder participation and utilization program evaluation. *Evaluation Review* 12 (2): 91-116.
- Guilford County Schools, Office of Assessment and Evaluation. 1999. Structural models for responses to the Middle School Survey. Unpublished document.
- Joyce, B., J. Wolf, and E. Calhoun. 1993. *The self-renewing school*. Alexandria, VA: ASCD.
- O’Sullivan, R., and A. D’Agostino. 1998. How collaborative approaches promote evaluation with community-based programs for young children and their families. Paper presented at the annual meeting of the American Evaluation Association, Chicago, November.
- O’Sullivan, R. G., and J. O’Sullivan. 1998. Evaluation voices: Promoting evaluation from within programs through collaboration. *Evaluation and Program Planning* 21:21-29.
- Purkey, W., and J. Lehr. 1996. Creating safe schools through invitational education. Unpublished manuscript, University of North Carolina at Greensboro, Alliance for Invitational Education.
- Scriven, M. 1997. Truth and objectivity in evaluation. In *Evaluation for the 21st century: A handbook*, edited by E. Chalimsky and W. Shaich, 477-501. Thousand Oaks, CA: Sage.
- Shoffner, M., and N. Vacc. 1999. Psychometric analysis of the Inviting School Safety Survey. *Measurement and Evaluation in Counseling and Development* 32:66-74.
- Strahan, D. 1994. Putting middle school perspectives into practice: Creating school cultures that promote caring. *Midpoints* 4 (1): 1-12.
- Stufflebeam, D. 2000. Foundation models for 21st century program evaluation. In *Evaluation models: Viewpoints on educational and human services evaluation*, edited by D. Stufflebeam, G. Madaus, and T. Kellaghan, 33-84. Boston: Kluwer.
- Van Tassel-Baska, J., K. Hall, and J. Bailey. 1996. Case studies of promising change schools. *Research in Middle Level Education Quarterly* 19 (2): 89-116.