SECOND FIELD TEST OF THE AEL MEASURE OF SCHOOL CAPACITY FOR IMPROVEMENT

Lisa D. Copley

Merrill L. Meehan

Caitlin W. Howley

Georgia K. Hughes

AEL

P. O. Box 1348 • Charleston WV 25325

AEL is a not-for-profit corporation that applies the latest education research and technology to spark innovation and promote strategies for student success. AEL delivers products and services that integrate research and development, evaluation, professional development, technology, and diverse perspectives. AEL operates several contracts funded by the U.S. Department of Education: a Regional Educational Laboratory, the Region IV Comprehensive Center, and the Eisenhower Regional Consortium for Mathematics and Science Education.

Information about AEL projects, programs, and services is available by contacting AEL.

Post Office Box 1348
Charleston, West Virginia 25325-1348
304-347-0400
800-624-9120
304-347-0487 (fax)
aelinfo@ael.org
http://www.ael.org

AEL is an Equal Opportunity/Affirmative Action Employer.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Since the 1960's, American schools have been under especial scrutiny for their capacity to educate youth effectively. Although school reform and improvement have always been national concerns (the Progressive era at the turn of the last century, for example), the launching of Sputnik in 1957, at a time when the Cold War shaped American fears, spurred alarm about the state of schooling in the country. If the Russians, who appeared to live under less prosperous conditions, were capable of such a scientific feat, citizens wondered, why had Americans not launched the first orbital satellite? One of the most frequently cited answers was that United States schools were not educating students sufficiently, particularly in subject areas of increasing prominence, such as math and science. The launch of Sputnik proved pivotal in our ongoing and contemporary concern with school improvement.

A number of school improvement trends have arisen since the 1960s in attempts to improve American education, each offering particular antidotes to educational troubles. Decentralization efforts in the 1960s and 1970s were approaches that sought to encourage local control of curriculum and finance, and to increase community participation in matters of education. Ultimately, however, many of these efforts became ineffective in terms of school improvement as involvement of community members was often token, or dominated only by the most influential community leaders (deMarrais & LeCompte, 1999).

Another wave of school improvement efforts, in response to the 1983 National Commission on Excellence in Education's report A nation at risk: The imperative for educational reform, focused on raising standards for students and teachers. This approach entailed establishing performance requirements for students and linking teacher accountability to student achievement on standardized tests. The standards movement continues to play a significant role in contemporary debate about how to improve education (Riordan, 1997).

In the 1990s, site-based management and shared decision-making were successors to the earlier decentralization efforts. These school improvement approaches sought again to render schools more responsive to community concerns. Nonetheless, participants with relatively little power continued to face obstacles to their full involvement, and research revealed little impact of site-based management or shared decision making on academic indicators (deMarrais & LeCompte, 1999; Riordan, 1997).

The Effective Schools movement was an attempt to discover what might make some schools better equipped than others to produce high performing students. According to this research (Levine & Lezotte, 1995), effective schools evidence specific characteristics, such as a clear mission, high academic expectations for all students, a safe school environment, and strong instructional leadership from administrators. However, this area of research failed to provide definitive insight into how schools developed such characteristics.

School improvement is increasingly viewed as an ongoing and comprehensive process. Recent legislation has encouraged the adoption of such a view, with the 1998 appropriation of $150 million by Congress to states for allocation to schools undertaking research-based schoolwide reform programs through the Comprehensive School Reform Demonstration

Program (CSRD). Earlier, in 1994, Congress altered regulations to allow schools receiving Title I funds, with free and reduced lunch 50% and above, to use such funds for whole school improvement (American Institutes for Research, 1999).

The reform models mentioned in the legislation instituting CSRD encompass a variety of approaches to reform, from skill-based, to comprehensive, to processual. In addition, the models vary in their degree of prescriptiveness. All claim to be based upon research and to have evidence of some positive impact. Yet investigations of and prototypes for school improvement extend far beyond the models forwarded in CSRD legislation: Contemporary literature on school improvement has roots in the school effectiveness literature of the 1970s and early 80s mentioned earlier (e.g., Levine & Lezotte, 1995).

Much current prescriptive education literature and some research suggest that the interplay between school cultural and structural conditions significantly affects how change at a particular school will be greeted (e.g., Newmann & Wehlage, 1996). They contend that if cultural characteristics, such as commitment to high expectations, support for inquiry, and caring relationships, intersect with structural factors, such as time for staff development and freedom from excessive organizational constraints, school reform will proceed more smoothly. These structural and cultural conditions can be seen as contributing to school capacity for improvement (Newmann, King, & Youngs, 2001).

Along with these intersections, school leadership must be an integral part of improvement efforts (van der Bogert, 1998), and collaboration among the many stakeholders in school communities must be pursued (Sarason & Lorentz, 1998). Fullan and Miles (1994) additionally suggest that those involved in improvement must recognize that it is a process, filled with ambiguity, uncertainty, and risk, rather than a scripted, easily implemented recipe. Moreover, Fullan's most important insight is that school reform will not proceed without the voluntary support of staff who view the reform as meaningful and in alignment with their own worldviews (Fullan, 1991).

Thus, efforts to improve schools are an ongoing and contemporary national concern. Research and policy in education are often devoted to imagining, mandating, defending, resisting, and assessing a wide variety of improvement strategies. Nonetheless, the majority of reforms have not resulted in significant change in practice (Cuban, 1993) or in student performance (American Institutes for Research, 1999; deMarrais & LeCompte, 1999; Riordan, 1997). As Brown, Halsey Lauder, and Wells (1997) imply, and as Anyon (1997) vividly demonstrates, other contextual factors play a pivotal role in how, and whether, school change is enacted. Newmann, King, and Youngs (2001) likewise suggest that school reform efforts interact with their context, part of which is school capacity for improvement. It is this important notion of school capacity that is the subject of the following section.

AEL's School Capacity Assessment—Pilot Version

A pilot version of AEL's School Capacity Assessment (SCA) was developed in the spring of 2002 by Caitlin Howley and Joy Riffle to assess the degree to which schools possess the potential to become high performing learning communities. This research and development

focus grows from the Department of Education's Institute of Education Sciences' (formerly Office of Educational Research and Improvement) concern with and commitment to investigating how low-performing schools may be transformed into learning communities for students, faculty, and community members. More specifically, the SCA was developed in response to AEL's School Capacity Development project, staff of which required an instrument to assess their efforts to enhance the capacity to improve in partner schools.

Based on a review of the education research on change, AEL research and evaluation staff defined school capacity as the presence of characteristics needed to support the development of a thriving learning community. These characteristics include certain teacher practices, perspectives, and school structures. School cultural and attitudinal factors were incorporated in this view of school capacity for improvement (Kruse, Louis, & Bryk, 1995). Structural components were also included in response to research showing the importance of school structures and policies to successful improvement initiatives (e.g., Fullan, 1991, 1994; Hord, Rutherford, Huling-Austin, & Hall, 1987; Howley & Brown, 2001; Kruse, Louis, & Bryk, 1995; Newmann, King, & Youngs, 2001). It is hypothesized that, lacking these structures, practices, and perspectives, school staff will be less likely to nurture and sustain significant school improvement.

Newmann and his colleagues (2001) contend that structural conditions, such as program coherence and alignment, the sufficiency of technical and professional resources, and the provision of adequate time for staff to plan collaboratively and/or implement change, are critical to the likelihood that school reform will be undertaken with commitment. Moreover, school improvement efforts cannot be sustained over time without sufficient support from district and school policies and structures (Howley & Brown, 2001). Structural conditions, though often invisible or taken for granted, significantly shape how people behave, of what they believe they (and their students) are capable, and to what they commit themselves (Bourdieu & Passeron, 1990; deMarrais & LeCompte, 1999; Fullan, 1991; Mills, 1959; Riordan, 1997).

In addition, teachers' practice also plays an important role in forecasting the success of school reform efforts. Louis, Marks, and Kruse (1996) illustrate how deprivatized practice, in which school staff regularly observe one another and provide constructive feedback, structures a conduit by which other change efforts may be brought to fruition. Meaningful collaboration becomes possible when staff are in the habit of crossing the thresholds of each other's classroom doors.

Equitable teaching practices and differentiated instruction together constitute a nuanced pedagogy that is at once attentive, equitable, and sensitive. As Darling-Hammond notes, "Successful education can occur only if teachers are prepared to meet rigorous learning demands and the different needs of students" (1997, p. 334). Teachers who are accustomed to applying themselves equitably to diverse students are better equipped to confront the challenges wrought by social, economic, and political devastation in low-performing schools and their communities (Anyon, 1997; Paley, 1979). However, it could also be argued that school staff are more likely to undertake serious change with commitment if they are already in the practice of differentiating instruction in ways intended to support their students fully and adequately.

Teachers' attitudes, perceptions, expectations, and assessments are also closely bound to the likelihood that their school is well positioned to undertake significant school improvement work. Faculty who believe that they are not capable as a group of teaching their students are not likely to have much faith in their attempts to effect any broader change in their school. Collective teacher efficacy is critical to the capacity schools possess for committing to and implementing improvement efforts (Goddard, Hoy, & Hoy, 2000).

Expectations for student performance, as with teacher efficacy, constitute an important gauge of school capacity. Depressed expectations indicate a professional fatalism not conducive to improvement or, obviously, enhanced student achievement (Tauber, 1998). In addition, schools with capacity are schools with a predisposition toward nurturing learning. If teachers do not expect much from their students, their school cannot possess much capacity for nurturing student achievement.

AEL's pilot version of the SCA was developed in response to the paucity of definition, operationalization, and assessment of school capacity in the education research and evaluation literature. It was intended for administration to K-12 school professional staff. Data from administration of the survey was to assist school staff in ascertaining how well positioned their schools are to begin the development of a high performing learning community. In addition, subscale data would allow staff to identify dimensions of school capacity in need of further development in their schools. The instrument was intended for diagnostic use, for instance at the beginning of school reform efforts. It also was intended for administration and analysis over the course of school improvement undertakings.

The SCA was a 99-item, four-page instrument. Response options to the items were forced-choice, using a scale of 1 to 4, in which 1 means *strongly disagree*, 2 means *disagree*, 3 means *agree*, and 4 means *strongly agree*. Subscale items were randomly distributed throughout the instrument so that subscales were not readily apparent to respondents. The instrument was in a machine scannable format.

Eight subscales constituted the survey: Collective Teacher Efficacy, Deprivatized Practice, Program Coherence, Technical Resources, Equitable Practice, Differentiated Instruction, Expectations for Student Performance, and Time for Planning. All eight subscales were either drawn directly from other research endeavors or were the result of syntheses of research efforts that did not necessarily produce assessment instruments.

The first two subscales had been validated previously. They are defined as follows:

- *Collective Teacher Efficacy:* a 12-item scale assessing "the extent to which a faculty believes in its conjoint capability to positively influence student learning" (Goddard, 2002, p. 97)

- *Deprivatized Practice:* a 7-item scale assessing "the frequency with which teachers observe each other's classes to critique colleagues' teaching and provide meaningful feedback; it also measures the frequency of constructive reviews from supervisors" (Louis et al., 1996, p. 769)

4

The remaining subscales were pilot tested in an effort to establish their validity and reliability. These scales were defined as follows:

- *Program Coherence*: a 12-item scale measuring "the extent to which the school's programs for student and staff learning are coordinated, focused on clear learning goals, and sustained over a period of time" (Newmann, King, & Youngs, 2001, p. 6)

- *Technical Resources:* a 7-item scale evaluating the availability to faculty of working equipment, technology, instructional materials, facilities, and professional resource materials, such as journals (Newmann, King, & Youngs, 2001)

- *Equitable Practice:* a 38-item scale measuring the degree to which faculty understand diversity and engage in classroom practices that equitably support the learning of all students (deMarrais & LeCompte, 1999; Pohan & Aguilar, 2001; Sadker & Sadker, 1994; University of Minnesota, Diversity Work Group, 2002)

- *Differentiated Instruction:* an 8-item scale assessing the extent to which faculty adapt their instructional strategies and grouping arrangements to meet the learning needs of diverse students (Baber, C.R., 2001; Tomlinson, 1995, 1999a-b, 2000; University of North Carolina, 2001)

- *Expectations for Student Performance:* a 10-item scale evaluating the degree to which faculty believe their students are capable of mastering material presented to them and the level at which teachers anticipate that their students will perform (Baber, 2001; Bourdieu & Passeron, 1990; deMarrais & LeCompte, 1999; McLeod, 1987; Ogbu, 1983; Paley, 1979; Riordan, 1997; University of North Carolina, 2001; Willis, 1981)

- *Time for Planning:* a 5-item scale assessing the extent to which school staff have sufficient dedicated time for planning and teaching (Abdal-Haqq, 1996; Lashway, 1998).

School Capacity for Improvement—Review of Literature

The importance of each subscale to a conceptualization of school capacity is explained below. It should be noted that three subscales were intended to assess various structural conditions under which teachers work; these are the Program Coherence, Technical Resources, and Time for Planning measures. The Deprivatized Practice, Equitable Practice, and Differentiated Instruction subscales were meant to ascertain teacher practices. The Expectations for Student Performance subscale was primarily attitudinal.

Collective Teacher Efficacy. Collective teacher efficacy extends the notion of individual teacher efficacy to a faculty's shared sense of capacity to effect positive student outcomes. Whereas an individual's assessment of his or her own efficacy as a teacher may vary according to specific contexts (such as class size, subject area, or student demographics), a measure of collective teacher efficacy provides a more global evaluation of the specific social

and organizational context in which a faculty works. Teachers' shared beliefs about their collective ability to teach students effectively is, according to Goddard, Hoy, and Hoy (2000), a better gauge of school capacity than measures of individual efficacy or internal locus of control. Collective teacher efficacy is "an emergent group-level attribute, the product of the interactive dynamics of the group members. As such, this emergent property is more than the sum of the individual attributes" (p. 482).

Further, collective teacher efficacy is "a way of conceptualizing the normative environment of a school and its influence on both personal and organizational behavior" (Goddard, 1998, p. 65). Teachers' perceptions of their faculty's ability to teach with efficacy shape their strivings and behaviors in the classroom. Thus, if teachers believe themselves to belong to a very efficacious faculty, "the normative environment will press teachers to persist in their educational efforts" (Goddard, 1998, p. 65). On the other hand, a faculty with little sense of collective efficacy will be less likely to exert normative pressure on each other to undertake rigorous pedagogy.

Because of its link to faculty behavior and its hypothesized (Goddard, 1998, 2002; Goddard, Hoy, & Hoy, 2000) and tentatively confirmed (Goddard, Hoy, & Hoy, 2002) impact on student achievement, collective teacher efficacy appears to constitute an important component of school capacity for improvement. A faculty that does not believe in its capabilities will not likely impel itself toward improvement. However, a faculty with a strong sense of its ability to effect change in student achievement will be better positioned to seek improvement.

Goddard's (2002) revision of his earlier measure of collective teacher efficacy was adopted for inclusion in AEL's pilot version of the SCA. The 12-item revision possesses adequate internal consistency reliability with a Cronbach's alpha coefficient of .94. Moreover, Goddard's analysis indicates that the new version is valid; the revised measure correlates highly with the earlier instrument, and multilevel tests of predictive validity showed that the new version is a good predictor of between-school variability in student mathematics achievement.

Deprivatized Practice. Louis et al. (1996) contend that, among other phenomena, deprivatized practice is pivotal in the development of school professional community. In this view, deprivatized practice is the degree to which faculty observe one another's work, provide feedback, and serve as mutual mentors or coaches. Schools in which practice is deprivatized tend to view teaching less as an autonomous individual project and more as a collaborative undertaking (Sarason & Lorentz, 1998). As a result, faculty in such schools experience less professional isolation and greater opportunity for learning from colleagues (Education Commission of the States, 1996). Deprivatized practice, then, provides faculty with a wider network of resources.

In terms of school capacity for improvement, serious change is not likely to take hold if faculty are not aided by norms or mechanisms that support collegial learning, critique, and cross-fertilization. As Cuban's (1993) historical analysis of school change reveals, professional isolation and conservative norms in schools have rendered most improvement efforts irrelevant, and ultimately teachers have made very few serious changes in their practice as a result. However, schools that provide the structural support for deprivatized practice invite

collaboration and collegiality, which in turn invite opportunities for sustainable improvement (Corallo & McDonald, 2002).

The 7-item Deprivatized Practice subscale is a closed-response option adaptation by Meehan and Cowley (1998) to the original open-ended questionnaire developed by Louis et al. (1996). Although the 1998 administration of the adaptation by Meehan and Cowley indicated that the subscale possessed less than ideal reliability, with Cronbach's alphas ranging between .65 and .69, a later administration by Nilsen (1999) revealed the scale to be more reliable, with an alpha of .84.

Program Coherence. An important structural condition supporting school capacity for improvement is instructional program coherence. According to Newmann, King, and Youngs (2001), program coherence is a measure of the extent to which a school is sufficiently programmatically integrated. The continual and shifting presence of unrelated, unfocused, and multiple improvement programs weakens schools' organizational efficacy. Conversely, aligned initiatives that are implemented and monitored carefully for sustained periods, at the very minimum, do not detract from a school's efforts to educate students.

Program coherence also encompasses the alignment of curriculum and instruction within grade levels and between grade levels (Corallo & McDonald, 2002; Newmann, Smith, Allensworth, & Bryk, 2001). Adequate alignment and sequencing assists in the maintenance of an appropriate intellectual pace and rigor, and focuses attention on the primary purpose of education. It also reduces redundancy and fosters communication and collaboration among teachers.

Program coherence is viewed as critical to school capacity for improvement because schools struggling to implement many unrelated programs are not immediately equipped to undertake significant improvement work. Already burdened with other competing and shifting priorities, teachers in schools with little programmatic coherence are unlikely to accommodate additional serious change. Focus and carefully allocated resources to a committed, shared purpose prepare a more hospitable environment for improvement.

The Program Coherence subscale on AEL's SCA is an adaptation of items from a survey of professional development to build school capacity. In addition, AEL staff added several other items. Newmann, King, and Youngs (2001) provided no reliability or validity analyses, although their study seems to confirm that program coherence constitutes a critical component of school capacity for improvement.

Technical Resources. Newmann, King, and Youngs (2001) also found the presence of adequate technical and professional resources to be a useful indicator of school capacity for improvement. Instructional materials, functioning technical and computer equipment, and adequate workspace represent some of the material conditions under which teachers work. Improvement efforts, which depend on such tools, are likely to fail if teachers do not have access to them.

In addition, teachers who feel that they do not have the material resources with which to teach to their objectives in the classroom will feel additionally hampered if asked to institute significant change across their school. If teachers' fundamental resource needs are unmet, the likelihood that their school can effect and sustain improvement is small.

As with the Program Coherence subscale, the Technical Resources subscale is an adaptation of survey items developed by Newmann, King, and Youngs (2001). Some items were used verbatim, others were modified, and still others were developed by AEL staff to extend and elaborate on the concept assessed by the subscale. Reliability and validity information about the items was not available.

Equitable Practice. Schools are increasingly diverse organizations, with larger percentages of African American and Latino/a students. In addition, national attention is focused on increasing the academic achievement of racially/ethnically-defined youth and of low socioeconomic status (SES) students (Fortune, 2002; Schwartz, 2001a). Education Week, for example, covered the issue in 2000 with a four-part series (Johnston & Viadero, 2000; Viadero, 2000; Viadero & Johnston, 2000a, 2000b). Equitable education for all students is, however, both a national challenge and a legal imperative since the 1954 *Brown v. Board of Education* Supreme Court decision, which overturned the "separate but equal" doctrine justifying school segregation by racial category.

Equity must also be applied to gender, as much research indicates that curriculum and instruction tend to favor boys (deMarrais & LeCompte, 1999; Sadker & Sadker, 1994). For instance, boys may receive more attention, praise, and opportunities to elaborate or correct their answers to instructional questions (Mid-Atlantic Equity Center, 1993). Female figures appear less often in literary or historical accounts in curricula, and girls confront sexist language at school in which being called female is an insult (Thorne, 1995). In addition, girls enroll in fewer advanced math and science courses than do their male counterparts (Perez, 2000).

Equitable practice can be defined in numerous ways, along multiple dimensions. Rose (1999), for instance, identifies 10 indicators of fair teaching, ranging from equal distribution of response opportunities to courtesy and respect. The University of Minnesota Diversity Work Group (2002) cites a long list of practices identified by educators as conducive to the development of an equitable environment. Kahle (2002) explicates a variety of strategies to enhance the equity of science teaching, and Rickford (2001) illustrates how the use of culturally relevant texts and higher order questioning techniques are useful strategies for engaging low SES and ethnic minority students. Ensuring that curriculum and discipline practices honor students' backgrounds is another strategy suggested as important to creating an equitable classroom (Thompson & O'Quinn, 2001). Multicultural education research also points up a wealth of practices that ensure students receive equitable educational opportunities (c.f., Banks & Banks, 1995). Ultimately, equitable practice is a multiple concept: More than one strategy is required for the creation and sustenance of an academic environment that is fair and sensitive to all students (NWREL, 1997).

Schools equipped to teach their students equitably, fairly, yet also sensitively are likewise equipped to make improvement equitably. Improvement can hardly be considered full and meaningful unless it is salient to the experience and achievement of all students.

The Equitable Practice subscale of AEL's pilot version of the SCA was developed by AEL staff using the research literature cited above as a catalyst. Items were constructed to account for a variety of equitable practices, including racially/ethnically and socioeconomically sensitive pedagogy, relevant curriculum, active discouragement of stereotypical comments and behavior, equitable praise, multicultural content, and use of students' preferred speaking styles to enhance learning.

Differentiated Instruction. Classrooms are not homogenously populated; students hail from various communities, bring disparate skills and strengths, and have differing academic needs. Varying content, process, products, and learning environment to meet students' assorted needs is differentiating instruction (Tomlinson, 2000). The University of North Carolina's School of Education (2001) makes the teaching of differentiated instructional strategies to pre-service teachers one of its priorities because it is considered so essential to effective pedagogy.

The rationales for differentiating instruction are many. Instruction that honors the linguistic and literacy styles of young children augments their reading skills (Vernon-Feagans, Hammer, Miccio, & Manlove, 2001), and by extension, their learning of any subject that requires literacy skills. Moreover, differentiated instruction has been shown to improve student achievement (Dahl, Scharer, Lawson, & Grogan, 1999; although see Rowan & Miracle, 1983, for an alternative view). Differentiated instruction accommodates students of various cognitive abilities (Tomlinson, 1999a) and accounts for the myriad ways in which we all learn (Tomlinson, 1999b). Undifferentiated instruction and curriculum, conversely, may stifle student enthusiasm for learning and ultimately for achieving to the fullest (Kohn, as interviewed by O-Neil & Tell, 1999). Sizer (1999) similarly points out that a "rigid system" of schooling will ultimately fail those students whom it does not accommodate (p.1). "A one-size-fits-all approach to classroom teaching is ineffective for most students and harmful to some," suggest Tomlinson and Kalbfleisch (1998, p.1) in their analysis of brain research, because "to learn, students must experience appropriate levels of challenge" (p. 3). As Tomlinson put it earlier, "There simply is no single learning template" for all students (1995, p.1)

The Differentiated Instruction subscale developed for the SCA attempts to measure the degree to which school faculty adapt their classroom teaching, grouping, and assessment practices in order to meet the needs of their various students. AEL staff constructed items with close attention to the literature cited above.

Expectations for Student Performance. School staff's expectations for student academic performance play a powerful role in how students actually perform. Teachers' expectations for students inform how they treat students. For instance, teachers holding depressed expectations for certain students may then treat them differently than other students perceived to be more capable. Such differential treatment, very different than the differential instruction described above, results in fewer opportunities to learn challenging material, less time to answer questions or complete assignments, and less frequent encouragement and praise

9

(deMarrais & LeCompte, 1999; Lumsden, 1997; McLeod, 1987; Willis, 1981). Over time, students' performance conforms to the expectations of teachers (Tauber, 1998), thereby confirming teachers' original expectations. In addition, teachers are in positions of power relative to students, making their expectations even more influential.

Wilson and Martinussen (1999) show dramatically how teacher expectations based on students' socioeconomic status and prior achievement significantly shape the final grades study participants accorded their students. Ogbu (1983) likewise illustrates how important teacher expectations are to students' academic involvement and, ultimately, to their achievement.

Expectations for student performance are often shaped by stereotypical assessments based on race/ethnicity, socioeconomic status, gender, family structure, language, immigrant status, religion, transience, sexual orientation, and other contextually significant social characteristics (Bourdieu & Passeron, 1990; deMarrais & LeCompte, 1999; McLeod, 1987; Ogbu, 1983; Paley, 1979; Riordan, 1997; Willis, 1981). Hence, teachers sometimes may anticipate that, for instance, white middle-class boys will perform better academically than working-class Latinas (Schwartz, 2001b). This is not to blame teachers for holding differential expectations; rather, such expectations are endemic to our stratified society (c.f., Rose, 1990; Takaki, 1987). Nonetheless, American education also seeks to nurture meaningful democratic involvement through equal opportunity to all citizens, and in this regard, differential expectations based on social and economic characteristics run counter to such ideals.

The Expectations for Student Performance subscale evaluates the degree to which teachers expect that their students are capable of mastering material presented to them this academic year. It also assesses the level at which teachers believe their students will perform vis-à-vis their peers. Items were developed by AEL staff following a review of the literature on the impact of teacher expectations on student performance described above.

Time for Planning. School improvement efforts may have little chance of success if faculty lack fundamental structural support for their implementation. Among the most important of such conditions is the provision of adequate time to allow staff to plan, implement, experiment with, and evaluate their improvement initiatives (Howley & Brown, 2001; Howley-Rowe, 1999; Raywid, 1993). "Insufficient time to plan for implementing [reform] is a common barrier to implementation and a frequent concern of teachers," reports Desimone (2000, p. 12) in her analysis of schools instituting comprehensive school reform. Teachers are better equipped to develop professionally if they have time during their workday to reflect, collaborate, and focus on their own learning. Such opportunities, moreover, are fundamental to the development of schools as professional learning communities (Abdal-Haqq, 1996; Lashway, 1998). Conversely, lack of time to plan and implement contributes to teacher turnover (Adelman, Haslem, & Pringle, 1996).

An adequate allotment of time for reform to be learned about and practiced, implemented, institutionalized, assessed, and reflected upon is crucial (Adelman & Walking-Eagle, 1997). Some researchers have even argued that time is so important to the success of any school improvement undertaking because change proceeds according to standard development

phases; without time, reform has no chance to develop (Hord, Rutherford, Huling-Austin, & Hall, 1987).

Sufficient time for planning is therefore an important structural resource to which teachers require access if reform is to have the opportunity to become institutionalized. For this reason, Time for Planning subscale items were developed by AEL staff to evaluate the extent to which faculty are provided enough time for within-grade and across-grade planning and for appropriate professional development.

In Sum. School capacity is an often-used phrase in discussions of educational reform and improvement. However, very few researchers have attempted to define and operationalize school capacity for improvement (although, see Newmann, King, & Youngs, 2001). Rather, school capacity is a vague, albeit appealing, reference to some ephemeral quality predisposing schools to successful change.

SCA Pilot Test Results

AEL staff have attempted to define and operationalize the concept of school capacity through the development of the SCA. Nonetheless, we were also interested in testing our definition empirically. Thus, a pilot test of the instrument was conducted during the summer of 2002 (Howley & Riffle, 2002).

The purpose of the pilot test of AEL's SCA was to begin an exploration of the instrument's subscales. AEL staff wanted to discover the correlations between items intended to constitute distinct subscales and assess discrete concepts, and to delete items not highly correlated with others in their respective subscales. In other words, AEL staff sought data reduction, as the 99-item instrument was cumbersome. Staff also were interested in the degree to which subscales were reliable. In sum, an exploratory analysis of the SCA's statistical properties was wanted.

The SCA was administered to 453 participants from one of two school districts with histories of social, economic, and political struggle, as well as depressed student achievement, in an effort to establish the psychometric properties of the instrument and its subscales. The piloted version of the SCA was a 99-item, four-page instrument. Response options to the items were forced choice, using a four-point Likert scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). Subscale items were randomly distributed throughout the instrument.

Pilot test results suggested that the SCA appeared to hold some promise for assessing school capacity for improvement. As would be expected given the nature of the sample of low-performing schools, item and subscale means were relatively low and negatively skewed. Overall, the instrument was internally consistent (alpha = .97) and most of the subscales possessed sufficient internal consistency reliability (range .69 to .97). Exploratory factor analyses confirmed most scales, but differentiated the Equitable Practice subscale further into the Anti-Discriminatory Teaching and Responsive Pedagogy subscales. Items within each were moderately to highly correlated. Moreover, correlations between the subscales were moderate to very strong with those assessing structural conditions highly correlated with one another, as were

those evaluating practice and attitudinal stances. These findings suggested that the overall instrument effectively assesses both structural and practice/attitudinal stances, and that, although the subscales are interrelated, they remain distinct measures. Moreover, the SCA appears to be able to identify struggling schools, although it is not yet clear that the instrument is also capable of identifying schools with a great degree of capacity for improvement.

Based on the pilot test, the SCA was revised to eliminate redundant and poorly worded items. The Equitable Practice subscale was also divided into the two subscales discerned by the exploratory factor analysis. Too, the instrument was renamed the "AEL Measure of School Capacity for Improvement."

AEL Measure of School Capacity for Improvement

A first field test of the revised and renamed AEL instrument was conducted in the spring of 2003. The results of the early field test were reported by Riffle, Howley, and Ermolov (2004). The purpose of the first field test was to assess the psychometric properties of the revised instrument with a larger number of respondents than were in the pilot test. This early field test was designed to assess: internal consistency reliabilities, test-retest (stability) reliabilities, concurrent validity with another instrument measuring similar constructs, and construct validity via factor analyses.

Professional staff in 35 schools (12 elementary, 10 middle, and 13 high schools) in six school districts in a southeastern state completed the AEL Measure of School Capacity for Improvement (AEL MSCI) in the spring of 2003. A total of 1,274 professional staff completed the survey. The majority (n=912) were regular classroom teachers, while the other role groups were: special education teacher (n=110), counselor (n=43), principal/assistant principal (n=39), librarian/media specialist (n-25), and other (n=107). Almost three fourths of the respondents were female (n=885) and more than half classified themselves as Black or African American. A total of 174 professional staff representing schools (3 elementary, 2 middle, and 2 high schools) from three districts completed the survey twice for test-retest purposes. The time between administrations was between two and three weeks.

Results from the first field test of the AEL MSCI were encouraging in terms of the refinement of the instrument (Riffle, Howley, & Ermolov, 2004). For example, the total instrument score was internally consistent for this administration with a Cronbach alpha of .97. Also, the subscales in the AEL MSCI were internally consistent with alphas ranging from .79 (on one subscale) up to .91 (on one subscale). The test-retest (stability) reliability for the total instrument score was .87, while the subscales ranged from .68 (one subscale), through the .70s (six subscales), to .86 (one subscale). Factor analysis with oblique rotation showed six fairly robust factors accounting for 45% of the total variance. Two other factors were rather weak. The six factors were named: Collective Professional Capacity, Peer Reviewed Practice, Equitable Practice, Technical Resources, Program Coherence, and Differentiated Instruction. Interestingly, all the items from the original Expectations for Student Performance subscale loaded on the Collective Professional Capacity subscale in this administration. Also, all of the items designed to assess responsive pedagogy and one other item loaded onto a separate factor. These items originally were part of a larger scale in the SCA instrument labeled Equitable

Practice. Therefore, that original scale name was applied to this new subscale in this administration of the AEL MSCI. In terms of concurrent validity, the correlation of the total AEL MSCI score with the total AEL CSIQ score was .68, with the former instrument accounting for 47% of the variance in the latter (Riffle, Howley, & Ermolov, 2004).

Results of the first field test of the AEL MSCI suggest that 58 of the original items comprise six subscales that have a high degree of internal consistency, are stable over time, and are correlated with a measure of successful engagement in continuous school improvement (Riffle, Howley, & Ermolov, 2004, pg. 26). However, the developers identified two areas of needed improvement for the AEL MSCI. First, all the development and testing of the instrument has been with low-performing schools. They suggest that it should be tested with schools other than those at the low end of the performance continuum. Second, they suggest that the four-point Likert type response options may not have generated enough variance to distinguish low and high performing schools. Their recommendation was to offer a wider range of response options—perhaps up to six points—in a subsequent test of the instrument.

Purpose of This Second Field Test

The major purpose of the second field test of the AEL MSCI instrument was to assess the psychometric properties of the refined version with a larger, more diverse group of respondents. The first objective of this field test was to expand the four-point Likert-type response scale to six points in order to yield more variance in responses. The second objective was to secure a larger and more diverse group of respondents to complete the six-response option version of the same 64 items as employed in the first field test. The third objective was to analyze the responses from this second field test to discover its psychometric properties. These analyses included: internal consistency reliability estimates, test-retest reliability estimates, construct validity assessment via factor analyses, and the correlations among the scales emerging from the factor analyses and with the total score. The fourth objective was to compare the results of the first field test with those of the second field test and the fifth objective was to make recommendations for the next steps with the AEL MSCI based on those field comparisons.

METHOD

Participants

A total of 2,357 professional staff representing 59 schools (22 elementary, 21 middle, 12 high, and 4 middle/high schools) from 19 districts completed the survey. Seven hundred six respondents worked in an elementary school, 827 were from a middle school, 620 were from a high school, and 204 were from a middle/high school. The majority of respondents (n=1670) were regular classroom teachers, with the remaining respondents fitting into the categories of special education teacher (n=251), principal/assistant principal (n=75), counselor (n=65), librarian/media specialist (n=41), and other (n=166). Half of the respondents held a master's, master's + 15, or master's + 30 or more (n=1142), while slightly less than half held a bachelor's, bachelor's + 15, or bachelor's + 30 or more (n=1010). The remaining respondents (n=126) categorized themselves as education specialist, had a doctorate, responded other, or chose not to respond (n=76).

Almost three-quarters of the respondents were female (n=1680), while slightly more than a quarter were male (n=593). More than half of the respondents classified themselves as White (n=1324) with slightly less classifying themselves as Black or African American (n=833). The remaining respondents (n=80) categorized themselves as American Indian or Alaska Native, Hispanic or Latino/a, Asian, other, or chose not to respond (n=120).

About one-quarter of participants (n=495) had taught or worked in any school for 25 years or more, while slightly less had taught or worked in any school for four to six years (n=379) and one to three years (n=379). In contrast, more than one-third of the respondents had taught or worked in the school in which they now teach one to three years (n=962) with slightly less reporting that they had taught in their current school for four to six years (n=480). In relation to how long participants had worked in a particular district, more than one-quarter (n=606) had worked in the district between one and three years, while somewhat less had worked in the district between four and six years (n=399) and more than 25 years (n=333).

Respondents noted that they had taught their current subject between 0 and 40 years and their current grade from 0 and 42 years. Please refer to the tables presented on the next page for the average number of years a subject and grade were taught, as well as the number of respondents reporting certification for the subject and grade taught.

Table 1

Average Number of Years Taught and Number of Teachers Certified by Subject for All Participants*

Subject	N	Mean	Number Certified
All subjects	465	10.22	518
Title I	75	6.91	70
Art	52	11.89	59
English	359	9.76	361
Geography	103	7.79	123
History	172	8.18	213
Math	405	10.05	375
Music	74	10.65	79
Physical Education/Health	127	11.09	131
Reading/Language Arts	387	9.03	336
Science	367	8.79	349
Social Studies	303	8.05	312
Other	385	10.86	408

* Note: Participants were asked to report the number of years taught and certification for all subjects.

Table 2

Average Number of Years Taught and Number of Teachers Certified by Grade for All Participants*

Subject	N	Mean	Number Certified
Pre Kindergarten	48	5.42	97
Kindergarten	206	7.11	378
First Grade	258	6.79	526
Second Grade	259	6.67	516
Third Grade	254	6.28	512
Fourth Grade	236	6.99	498
Fifth Grade	297	6.58	544
Sixth Grade	462	6.69	649
Seventh Grade	534	7.39	793
Eighth Grade	521	7.51	793
Ninth Grade	569	9.55	425
Tenth Grade	548	10.04	692
Eleventh Grade	552	10.42	698
Twelfth Grade	523	10.46	688

* Note: Participants were asked to report the number of years taught and certification for all grades.

Test-retest Participants

A total of 284 professional staff representing schools (3 elementary, 3 middle, and 2 high schools) from six districts completed the survey for test-retest purposes. One hundred one respondents worked in an elementary school, 107 were from a middle school, and 76 were from a high school. The majority of respondents (n=194) were regular classroom teachers, with the remaining respondents fitting into the categories of special education teacher (n=33), principal/assistant principal (n=9), librarian/media specialist (n=7), counselor (n=7), and other (n=30). Half of the respondents held a master's, master's + 15, or master's + 30 or more (n=137), while slightly less held a bachelor's, bachelor's + 15, or bachelor's + 30 or more (n=125). The remaining respondents (n=22) categorized themselves as education specialist, had a doctorate, responded other, or chose not to respond.

Almost three-quarters of the respondents were female (n=205), while slightly more than one-quarter were male (n=73). More than half of the respondents classified themselves as White (n=185) with slightly more than one-quarter classifying themselves as Black or African American (n=73). The remaining respondents (n=26) categorized themselves as American Indian or Alaska Native, Hispanic or Latino/a, other, or chose not to respond.

More than one-quarter of participants (n=75) had taught or worked in any school for 25 years or more, while slightly less had taught or worked in any school for one to three years (n=66) and four to six years (n=39). In contrast, approximately three-quarters of the respondents had taught or worked in the school in which they now teach one to three years (n=152) or four to six years (n=54). In relation to how long participants had worked in a particular district, more than one-third (n=94) had worked in the district between one and three years, while somewhat less had worked in the district more than 25 years (n=46) and between four and six years (n=42).

Respondents noted that they had taught their current subject between 1 and 40 years and their current grade from 0 and 40 years. Please refer to the tables on the next page for the average number of years a subject and grade were taught, as well as the number of respondents reporting certification for the subject and grade taught.

Table 3

Average Number of Years Taught and Number of Teachers Certified by Subject for Test-retest Participants*

Subject	N	Mean	Number Certified
All subjects	65	11.71	71
Title I	17	4.76	14
Art	6	23.33	9
English	58	9.36	53
Geography	20	11.05	21
History	29	7.48	33
Math	65	10.8	53
Music	12	11.08	12
Physical Education/Health	20	9.75	19
Reading/Language Arts	56	10.04	45
Science	59	7.44	551
Social Studies	43	8.95	41
Other	46	9.15	52

* Note: Participants were asked to report the number of years taught and certification for all subjects.

Table 4

Average Number of Years Taught and Number of Teachers Certified by Grade for Test-retest Participants*

Subject	N	Mean	Number Certified
Pre Kindergarten	5	6.00	15
Kindergarten	25	8.40	54
First Grade	35	8.24	64
Second Grade	38	8.12	66
Third Grade	37	8.46	70
Fourth Grade	33	9.83	61
Fifth Grade	39	9.32	70
Sixth Grade	69	5.89	85
Seventh Grade	67	6.49	96
Eighth Grade	70	6.99	99
Ninth Grade	57	9.67	79
Tenth Grade	59	9.29	74
Eleventh Grade	58	10.16	75
Twelfth Grade	55	10.24	71

* Note: Participants were asked to report the number of years taught and certification for all grades.

Instrumentation

The AEL Measure of School Capacity for Improvement (AEL MSCI) is a 64-item instrument designed to assess the degree to which schools possess the potential to become high performing learning communities. The AEL MSCI was developed in response to the paucity of definition, operationalization, and assessment of school capacity in the education research and evaluation literature. It is intended for administration to K-12 school professional staff to assist in ascertaining how well positioned schools are to undertake school reform efforts. It is also intended for administration and analysis over the course of school improvement undertakings. In addition, the survey may be used to assess professional staff's perceptions generally, or to explore other differences based on gender, socioeconomic status (SES), or ethnicity.

The AEL MSCI takes up to 25 minutes for participants to complete and is easily administered by school personnel, researchers, and others, with no advance preparation of participants required. For 31 items, professional staff are asked to rate the extent to which each item is true for their school, using a six-point Likert-type scale ranging from 1 indicating (*not at all true*) to 6 indicating (*completely true*). For the remaining items, professional staff are asked to rate how often each item is true for their school using a similar six-point Likert-type scale ranging from 1 indicating (*never true*) to 6 indicating (*always true*). Participants are also asked to respond to additional demographic items. The survey is formatted for machine scoring.

Data Collection

Several thousand copies of the instrument were shipped to AEL staff in Tennessee. The appropriate number of surveys, along with brown, sealable envelopes, were packaged and distributed to a Tennessee Exemplary Educator (TN EE) assigned to the participating school. Each TN EE distributed the surveys to school staff, who completed their surveys either in a group setting or on an individual basis. Each participant was provided with a brown envelope in which to place their completed survey to assure them of the confidentiality and anonymity of their responses. The completed surveys in their sealed envelopes were returned to the TN EEs, who then returned them to AEL. A letter to the TN EEs as well as an instruction sheet were prepared in January 2004 and sent with the copies of the instrument and envelopes.

For test-retest data collection purposes the appropriate number of surveys, along with brown, sealable envelopes (large and small), were packaged and distributed to a Tennessee Exemplary Educator (TN EE) assigned to the participating school. Each TN EE distributed the surveys to school staff, who completed their surveys either in a group setting or on an individual basis. Each participant was provided with a brown envelope in which to place their completed survey and was asked to sign their name across the seal in an effort to assure them of the confidentiality and anonymity of their responses. The completed surveys in their sealed envelopes were returned to the TN EEs, who held them until the survey was administered a second time.

At the time of the second administration, each participant was given his or her signed envelope and asked to open the envelope and place the completed survey in a new, small brown envelope. After completing the survey a second time, each participant was asked to place the

sealed envelope containing the survey from the first administration as well as the second survey into a large brown envelope. The large brown envelopes were sealed and returned to the TN EEs, who then returned them to AEL.

Data Analysis

AEL staff scanned the returned and completed surveys using Remark optical scanning software. During and after scanning, they cleaned the data files; subsequently exporting them to a standard software program (Statistical Package for the Social Sciences, now known as SPSS) for statistical analyses. These analyses included the computation of descriptive statistics, including means and standard deviations, for the entire sample. To explore the validity of the MSCI, factor analysis using principal component analysis with oblimin rotation was conducted. Correlation matrices were likewise generated to examine validity. Several statistical techniques were employed to investigate reliability. Test-retest reliability was examined via the computation of correlations between two administrations of the MSCI.

Findings

This section presents the findings from the 2004 field test of the AEL MSCI instrument. Included are the internal consistency reliabilities, the test-retest reliabilities, and the factor analyses results.

Internal Consistency Reliabilities

Internal consistency of the total AEL MSCI and its original eight subscales in this administration was estimated with the Cronbach's alpha coefficient. The alphas showed that the total AEL MSCI (alpha = .97) and its subscales were very reliable, with alphas ranging from .80 for the Technical Resources scale to .93 for Differentiated Instruction scale. See below for the remaining alphas.

Total AEL MSCI	0.97
Technical Resources	0.80
Program Coherence	0.83
Peer Reviewed Practice	0.85
Anti-Discriminatory Teaching	0.85
Collective Professional Capacity	0.86
Responsive Pedagogy	0.87
Expectations for Student Performance	0.92
Differentiated Instruction	0.93

Test-Retest Reliabilities

The correlation between total AEL MSCI scores on the two administrations of the survey was .88 (p = .000) based on 197 respondents who completed all items. Accordingly, participants' responses on the two tests appear to have remained stable over time. Correlations by subscale mean scores from the two administrations are presented in the table below and range from .74 for Anti-Discriminatory Teaching to .83 for Program Coherence and Technical Resources. Therefore, original subscale scores appear to have adequate reliability over time.

Table 5

Descriptive Information and Stability of the Original Eight AEL MSCI Subscales Across Administrations

Subscales	1st Administration			2nd Administration			Correlation Coefficient
	N	Mean	SD	N	Mean	SD	
Collective Professional Capacity	284	4.42	.76	284	4.44	.74	.77*
Expectations for Student Performance	282	4.30	.85	284	4.30	.85	.78*
Peer Reviewed Practice	283	4.07	.97	284	4.04	1.02	.83*
Responsive Pedagogy	284	4.53	.72	284	4.51	.82	.83*

Subscales	1st Administration			2nd Administration			Correlation Coefficient
	N	Mean	SD	N	Mean	SD	
Anti-Discriminatory Teaching	284	4.88	.67	284	4.91	.73	.74*
Technical Resources	284	4.00	.87	284	4.05	.85	.79*
Program Coherence	284	4.39	.78	284	4.37	.84	.78*
Differentiated Instruction	282	4.48	.83	284	4.52	.81	.75*

* Correlation is significant at the 0.01 level (two-tailed).

Results of a similar analysis based on 60 items loading on seven factors revealed by the factor analysis show improved stability in subscale scores over time ranging from .75 to .84. Overall, the changes had no apparent effect on the stability of total scores on the AEL MSCI (r=.88, p=.000).

Table 6

Descriptive Information and Stability of the Seven Revised AEL MSCI Subscales Across Administrations

| Subscales	1st Administration			2nd Administration			Correlation Coefficient
	N	Mean	SD	N	Mean	SD	
Collective Professional Capacity	284	4.57	.78	284	4.59	.75	.77*
Peer Reviewed Practice	282	3.53	1.28	284	3.57	1.30	.79*
Equitable Practice	284	4.77	.66	284	4.77	.75	.78*
Time for Planning	284	4.12	1.12	284	4.17	1.12	.75*
Technical Resources	284	4.31	1.03	284	4.36	.98	.83*
Program Coherence	284	4.43	.79	284	4.34	.84	.84*
Expectations for Student Performance	284	4.10	.83	284	4.14	.83	.75*

* Correlation is significant at the 0.01 level (two-tailed).

Construct Validity—Factor Analyses

Factor analysis using principal component analysis with oblique rotation was conducted. Oblique rotation was selected because factors were expected to be closely related to one another as they were all related to school capacity for improvement. Initial results revealed 17 factors with eigenvalues greater than one and the total variance explained after rotation was 67.1%. Only 12 of these factors accounted for the variance with three of the 12 factors containing two items or less, each accounting for less than ten percent of the variance. Therefore, a secondary factor analysis was conducted using the same method, but forcing eight factors. In this analysis, 60.2% of the variance was explained after rotation. As can be seen in the Scree plot below, all of the factors appear to be fairly robust. This is further evidenced by factor loadings ranging from .30 to .92.

Figure 1

Scree Plot for Constrained Factor Analysis

Scree Plot
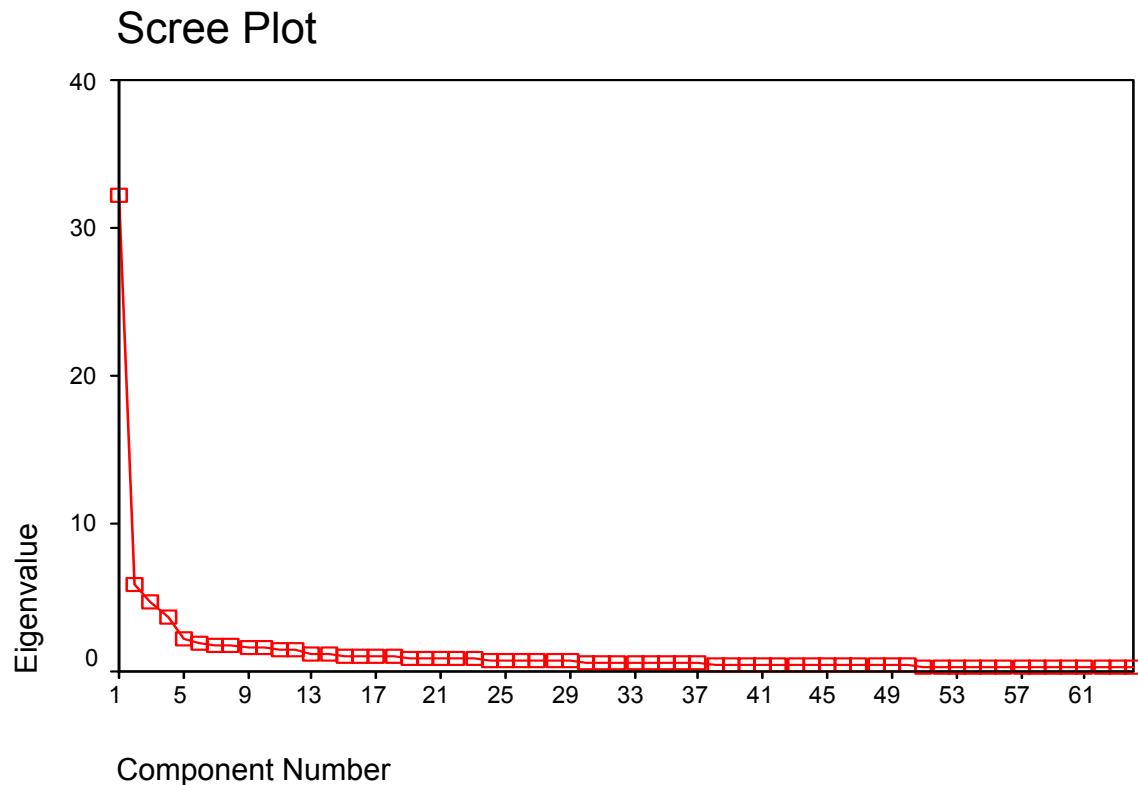
Upon further review of the factor analysis, staff noted that one of the factors consisted of only three items: two from Peer Reviewed Practice and one from Technical Resources. The loadings ranged from .35 to .69. Although this factor had an eigenvalue of 1.92, it only accounted for 2% of the total variance and had a Cronbach's alpha of .25. Therefore, this factor was excluded and the items were excluded. In addition, one survey item did not load on any of the factors and was subsequently excluded from further analysis as well.

First Factor

All of the items designed for Differentiated Instruction, five of eight designed for Collective Professional Capacity, one item designed for Expectations of Student Performance, and one item designed for Program Coherence loaded on the first factor. The loadings ranged from .30 to .71. This suggests that collective professional capacity is closely related to differentiated instruction. Thus, the methods teachers employ to incite learning appear to overlap with their perceptions of their ability as a whole. Therefore, the factor retained the name Collective Professional Capacity.

Second Factor

The second factor consists of four items designed to assess Peer Reviewed Practice. The items' loadings ranged from .63 to .89. No other items from different subscales loaded on this factor. Therefore, this factor retained the name of Peer Reviewed Practice.

Third Factor

The third factor consists of all eight items designed for Anti-Discriminatory teaching and all but one designed for Responsive Pedagogy. The factor loadings ranged from .49 to .77. From these results, it appears that culturally sensitive teaching and student responsive teaching are related constructs. The subscale name of Equitable Practice from the early development of the AEL SCA (Howley and Riffle, 2002) will be reinstated to describe this factor.

Fourth Factor

The fourth factor consists of three items designed for Technical Resources and one item from the Program Coherence scale. The loadings ranged from .43 to .92. The three items from the Technical Resources scale related to time available for teaching and planning. The Program Coherence item related to professional development. The subscale named Time for Planning from the early development of the AEL SCA (Howley and Riffle, 2002) attempted to ascertain the extent to which faculty were given ample time for within- and across-grade planning, as well as for professional development. Therefore, this subscale name will be reinstated to describe this factor.

Fifth Factor

The fifth factor consists of four items designed for the Technical Resources subscale. No items from another factor loaded on this factor. The total variance explained by this factor was 11.5% and factor loadings range from .59 to .87. Each item deals with having the materials and/or equipment necessary to teach a subject properly. This factor retains the name Technical Resources.

Sixth Factor

The sixth factor contained seven items: two from the Peer Reviewed Practice scale and five from the Program Coherence scale. The loadings ranged from .32 to .51. All of these items relate to program coherence. Thus, the subscale was named Program Coherence.

Seventh Factor

The seventh and final factor contained 11 items: seven from the Expectations for Student Performance scale, three from the Collective Professional Capacity scale, and one from the Responsive Pedagogy scale. The loadings ranged from .32 to .78. From these results, it could be interpreted that collective professional capacity and expectations for student performance share underlying constructs. Since the items encompass student performance the subscale retained the name Expectations for Student Performance.

Table 7

Factor Loadings for 2004 Revised AEL MSCI Subscales

Item Numbers	Collective Professional Capacity	Peer Reviewed Practice	Equitable Practice	Time for Planning	Technical Resources	Program Coherence	Expectations for Student Performance
1	.71	.89	.77	.92	.87	.51	.78
2	.65	.84	.75	.86	.79	.50	.77
3	.61	.73	.75	.49	.66	.50	.69
4	.61	.63	.73	.43	.59	.39	.60
5	.60		.69			.35	.60
6	.59		.68			.34	.56
7	.56		.67			.32	.54
8	.55		.66				.49
9	.54		.64				.47
10	.51		.62				.44
11	.48		.54				.32
12	.42		.54				
13	.40		.53				
14	.35		.53				
15	.30		.49				

Correlations Among Scales and Total AEL MSCI Score

The correlations between each of the newly created subscales ranged from .28 for the Technical Resources and Peer Reviewed Practice subscales, to .80 for the Expectations for Student Performance and Collective Professional Capacity subscales. As expected, each of the subscales significantly correlated with the total AEL MSCI score ranging from .61 (Peer Reviewed Practice) to .91 for Collective Professional Capacity. See Table 8 for complete results.

Table 8

Intercorrelations Among the Subscales and Total AEL MSCI Score

Subscales	Collective Professional Capacity	Peer Reviewed Practice	Equitable Practice	Time for Planning	Technical Resources	Program Coherence	Expectations for Student Performance
Collective Professional Capacity	---						
Peer Reviewed Practice	.44	---					
Equitable Practice	.68	.39	---				
Time for Planning	.57	.55	.51	---			
Technical Resources	.46	.28	.34	.52	---		
Program Coherence	.62	.51	.54	.70	.55	---	
Expectations For Student Performance	.80	.40	.55	.52	.47	.58	---
Total AEL MSCI	.91	.61	.82	.77	.62	.81	.86

The reliability estimate based on Cronbach's alpha for the revised AEL MSCI total instrument remained at .97. The revised Collective Professional Capacity (15 items) and Equitable Practice (15 items) subscales each had a new alpha of .93. Expectations for Student Performance (11 items) with an alpha of .92 also appeared highly reliable. The Peer Reviewed Practice (4 items), Time for Planning (4 items), Program Coherence (7 items), and Technical Resources (4 items) subscales were slightly less reliable with alphas of .84, .82, .81, and .80 respectively.

DISCUSSION/CONCLUSIONS

This section presents a discussion of the findings by comparing the 2003 and 2004 field test reliability results and, also, by comparing the construct validity results across the two years. Next, some general conclusions are drawn from the results.

Comparison of 2003 and 2004 AEL MSCI Findings

The internal consistency reliability between the 2003 and 2004 total AEL MSCI administrations was quite stable, as each year the Cronbach's alpha coefficient was .97. When 2003 (reported in 2004 by Riffle, Howley, & Ermolov) and 2004 scores are correlated using Cronbach's alpha; the coefficient was .86. Cronbach's alpha was used to determine the internal consistency reliabilities of the eight subscales from the 2003 and 2004 administrations. There was a slight increase in reliability in the eight subscales. In 2003, alphas ranged from .79 to .91. In 2004, alphas ranged from .80 to .93. In 2004, the alpha for Collective Professional Capacity was .86. The alpha for Peer Reviewed Practice was .85. For Program Coherence the alpha was .83. For Technical Resources the alpha was .80 and the alpha for Anti-Discriminatory Teaching was .85. The alpha for Responsive Pedagogy was .87 and Differentiated Instruction had an alpha of .93. Expectations for Student Performance had an alpha of .92. These are good reliability scores –and not unexpected– as the AEL MSCI went from a four-point Likert Scale to a six-point Likert scale and the N increased by 1,083 surveys from the 2003 to 2004 administration.

Comparison of Construct Validity

The construct validity was very similar for the 2003 and 2004 administration of the AEL MSCI. In 2003, a forced eight factor analysis revealed six robust subscales, while seven robust factors were revealed by the forced eight factor analysis in 2004. Five of the subscales retained the same name; only Differentiated Instruction was not a named subscale in 2004. The new subscales in 2004 are Time for Planning and Expectations for Student Performance. However, many of the factors were quite different. Factor loadings in 2003 were from .34 to .86 for all items. In 2004, loadings ranged from .30 to .92. Six items were dropped from further analysis in 2003 due to not loading strongly with one of the six remaining factors. In 2004, only four items were dropped from further analysis.

In 2003, the first factor had 16 items whose loadings ranged from .34 to .86 and consisted of the items designed for Student Expectations and Collective Professional Capacity. In 2004, the first factor had 15 items whose loadings ranged from .30 to .71 and consisted of the items designed for Differentiated Instruction and Collective Professional Capacity with one item designed for Expectations for Student Performance and one item designed for Program Coherence. In both years the first factor was named Collective Professional Capacity. About half of this subscale remained the same from 2003 to 2004, but the other half changed completely. The possible reasons for this change are the larger N in 2004, possibly a more heterogeneous sample in 2004, and that the Likert scale increased from a four-point to a six-point. The Likert scale is the only thing that changed in the survey, the items were exactly the same and in the exact same order from 2003 to 2004. These are possible reasons to account for the variance in the loadings of the items.

In 2003, the second factor had four items whose loadings ranged from .70 to .81 and consisted of Peer Reviewed Practice items. In 2004, the second factor had four items whose loadings ranged from .63 to .89 with only items designed for Peer Reviewed Practice. The factor retained the name of Peer Reviewed Practice.

In 2003, the third factor had 16 items whose loadings ranged from .52 to .77 and consisted of items designed for Anti-Discriminatory Teaching and Responsive Pedagogy with one item from Peer Reviewed Practice. In 2004, the third factor consisted of 15 items whose loadings ranged from .49 to .77 and consisted of items designed for Anti-Discriminatory and Responsive Pedagogy. Again in 2004, there was more variance in the loading of items, possibly due to the greater number of surveys, the change in the Likert scale, and possible increased heterogeneity. Each year the factor was named the Equitable Practice subscale.

The fourth factor in 2003 was Technical Resources, and it was the fifth factor in 2004. Both years there were four items, but the loadings were higher in 2004 (.59 to .87) compared to 2003 (.54 to .80). It is not surprising that the items loaded higher in 2004 considering all those surveyed were from identified low-performing schools. From 2003 and 2004, this is a robust subscale.

The fourth factor in 2004 was not identified in 2003. It had four items with loadings ranging from .43 to .92 and consisted of three items designed for Technical Resources and one item from the Program Coherence scale. The items related to time available for teaching and planning, as well as professional development; therefore, the subscale name Time for Planning was reinstated from the AEL SCA. Possible reasons for the emergence of this factor include the greater number of surveys, the change in the Likert scale, and possible increased heterogeneity.

The fifth factor in 2003 was Program Coherence, and in 2004, it was the sixth factor. In 2003, there were nine items whose loadings ranged from .37 to .80 and consisted of items designed for Program Coherence, with two items from Technical Resources, and one from Peer Reviewed Practice subscales. In 2004, there were seven items whose loadings ranged from .32 to .51 and consisted of five items designed for the Program Coherence subscale and two from the Peer-Reviewed Practice subscale.

The sixth factor in 2003 was the Differentiated Instruction subscale. It had nine items whose loadings ranged from .38 to .64 and consisted of items designed for Differentiated Instruction with one item from Collective Professional Capacity. These item loadings do not appear to be outstanding. In 2004, the seventh factor was the Expectations for Student Performance subscale. It contained 11 items whose loadings ranged from .32 to .78 with seven items designed for Expectations for Student Performance, three from Collective Professional Capacity, and one from Responsive Pedagogy subscales. This factor is completely different from 2003 to 2004, possibly due to an increased Likert scale from four-point to a six-point, greater number of surveys completed, and greater heterogeneity of respondents.

Conclusions

In summary, the AEL MSCI is a highly reliable survey to measure a school staff's capacity for improvement. For the administrations of the AEL MSCI in the past two years, the Cronbach's alpha for the total score has been .97. In 2003, a factor analysis that forced eight factors revealed only six viable factors, while seven viable factors were revealed in 2004. The survey was designed for eight subscales with eight items each, but each year that is not the case after factor analysis is completed to examine the construct validity. The survey was much improved in 2004 by changing the Likert scale from a four-point to a six-point scale, having 1,083 more surveys completed, and by the increased heterogeneity of respondents.

In 2003, only 45% of the total variance was accounted for. In 2004, 60.2% of the total variance was accounted for. This increase most likely is due to the change in the Likert scale, the larger N, and the greater heterogeneity. Two of the six factors in the 2004 factor analysis changed dramatically. The first factor, Collective Professional Capacity, was comprised of items designed for collective professional capacity and expectations for student performance in 2003. In 2004, the same factor retained the name Collective Professional Capacity, but was comprised of items designed for collective professional capacity and differentiated instruction. In 2003, Differentiated Instruction was the sixth factor, while in 2004 it was Expectations for Student Performance. It is likely that items designed for differentiated instruction and expectations for student performance loaded oppositely from 2003 to 2004 due to the change in the Likert scale, the greater N, and the greater heterogeneity. In addition, the fourth factor revealed in 2004 was not identified in 2003. Again, this is likely the result of the greater number of surveys, changed in Likert scale, and greater heterogeneity. The remaining four factors remained stable from 2003 to 2004. The results of the factor analysis examining construct validity are much stronger in 2004 due to the much larger N, which probably added greater heterogeneity.

RECOMMENDATIONS

Several recommendations are derived from this second large field test of the AEL MSCI instrument. These recommendations follow.

First, the second field test (2004) with the expanded, 6-option response scale and the larger sample size proved very worthwhile, as detailed in prior sections. However, one glaring constraint on the viability of the results of both the 2003 and 2004 field tests is that both were administered to schools that were identified by their state department of education as being "low performing." The AEL MSCI has yet to be administered to a large group of "regular" or "normal" schools. Although the AEL MSCI shows much promise as an instrument to aid schools in their improvement efforts, the developers need to test it with schools that have not been identified as being low-performing. Therefore, the first recommendation is to test the AEL MSCI one more time with a large group of "regular" or "normal" schools. The most important part of this next test with regular schools would be to conduct the factor analyses over with the data from staff in those schools.

Second, with the internal consistency reliabilities, test-retest reliabilities, and concurrent validity established in the first two field tests of the AEL MSCI and with the latest factor analyses to be completed as part of the recommendation above, then the next logical step in the development of the instrument is to establish the norms for the final set of subscales and the total scale score. Of course, the success of the norming step is contingent upon securing both a large enough sample of schools in this third field test but, more importantly, securing these schools in the sample from a more typical pool of "regular" or "normal" schools and not those that have been identified, on the basis of any criteria, as belonging to just one or two categories, such as "medium" or "low" performing.

Third, once the third field test, the finalization of the subscales, and the norming of the AEL MSCI is completed, the development of a user manual and technical report would be appropriate. This user manual and technical report of the AEL MSCI should contain all the information that subsequent users and other researchers would need to make decisions about employing the AEL MSCI in their schools or in their research studies. These types of information would include: background, literature review, development of the instrument, various tests of the instrument, definition of the subscales, technical qualities of it, and the norms for the subscales and total score. Of course, references and appropriate appendices, including the final version of the AEL MSCI, should be included in the user manual and technical report.

REFERENCES

Abdal-Haqq, I. (1996). Making time for teacher professional development. *ERIC digest*. Retrieved from the World Wide Web August 1, 2002: http://www.ericsp.org/pages/digests/making_time_teacher_pro_dev_95-4.html

Adelman, N. E., Haslem, M. B., & Pringle, B. A. (1996). The uses of time for teaching and learning. Washington, DC: U.S. Department of Education, Office of Education Research and Improvement. Retrieved from the World Wide Web on August 19, 2002: http://www.ed.gov/pubs/SER/UsesofTime

Adelman, N. E., & Walking-Eagle, K. P. (1997). Teachers, time, and school reform. In A. Hargreaves, Ed., *Rethinking educational change with heart and mind, 1997 ASCD yearbook*. Alexandria, VA: Association for Supervision and Curriculum Development.

American Institutes for Research. 1999. *An educators' guide to schoolwide reform.* Retrieved from the World Wide Web July 16, 1999: http://www.aasa.org/Reform

Anyon, J. (1997). Ghetto *schooling: A political economy of urban educational reform.* New York: Teachers College Press.

Baber, C. R. (2001). UNCG conceptual framework for developing caring, collaborative, and competent professionals. Greensboro, NC: University of North Carolina. Retrieved from the World Wide Web: http://soe.uncg.edu/affiliates/teachers_academy/documents/uncg_cf_ovr.pdf

Banks, J., & Banks, C. M., Eds. (1995). *Handbook of research on multicultural education*. New York: Simon & Schuster.

Bourdieu, P., & Passeron, J. (1990) Reproduction in education, society and culture, 2nd ed. Sage.

Brown, P., Halsey, A. H., Lauder, H., & Wells, A. S. (1997). The transformation of education and society: An introduction. In A. H. Halsey, H. Lauder, P. Brown, & A. S. Wells (Eds.), *Education: Culture, economy, and society.* New York: Oxford University Press.

Corallo, C. & McDonald, D.H. (2002). *What works with low-performing schools: A review of research.* Lanham, MD: Scarecrow Education-AEL.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1880-1990*, 2nd ed. New York: Teachers College Press.

Dahl, K. L., Scharer, P. L., Lawson, L. L., & Grogan, P. R. (1999). Phonics instruction and student achievement in whole language first-grade classrooms. *Reading Research Quarterly, 34:* 3, pp. 312-341.

Darling-Hammond, L. (1997). Restructuring schools for student success. In A. H. Halsey, H. Lauder, P. Brown, & A. S. Wells (Eds.), *Education: Culture, economy, and society.* New York: Oxford University Press.

deMarrais, K. B., & LeCompte, M. D. (1999). *The way schools work: A sociological analysis of education.* New York: Longman.

Desimone, L. (2000). Making comprehensive school reform work. New York, NY: ERIC Clearinghouse on Urban Education. Retrieved from the World Wide Web on August 20, 2002: http://eric-web.tc.columbia.edu/monographs/uds112/index.html

Education Commission of the States. (1996). *Lessons learned: How collaboration contributes to school improvement.* Denver, CO: Education Commission of the States.

Fortune, A. (2002). Closing the achievement gap: A look ahead. *Gaining ground: Achieving excellence in high poverty schools, 4*: 1. Washington, DC: Council of Chief State School Officers.

Fullan, M .G. (1991). *The new meaning of educational change.* New York: Teachers College Press.

Fullan, M. G., & Miles, M. B. (1994). Getting reform right: What works and what doesn't. In *Changing education: Resources for systemic change*, pp. 211-218. Washington, DC: U.S. Department of Education.

Garson, G. D. (2002). Factor analysis. Retrieved form the World Wide Web on August 8, 2002: http://www2.chass.ncsu.edu/garson/pa765/factor.htm

Goddard, R. (1998). *The effects of collective efficacy on student achievement in urban public elementary schools.* Doctoral dissertation, Ohio State University, Columbus.

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

Goddard, R. (2002). A theoretical and empirical analysis of the measurement of collective efficacy: The development of a short form. *Educational and Psychological Measurement, 62:* 1, pp. 97-110.

Goddard, R., Hoy, W. K., & Hoy, A. W. (2002). The effects of collective efficacy on student achievement in urban public elementary schools. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, April 2002.

Hord, S. M., Rutherford, W. L., Huling-Austin, L., & Hall, G. E. (1987). *Taking charge of change.* Alexandria, VA: Association for Supervision and Curriculum Development.

Howley, C., & Brown, P. (2001, April). To continue, press on: Sustaining school improvement. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA, April 10-14, 2001.

Howley-Rowe, C. (1999). *Engaging with school improvement: Initial and sustained involvement in the Quest network.* Charleston, WV: AEL.

Howley, C., & Riffle, M.J.S. (2002). *A Pilot Test of AEL's School capacity assessment.* Charleston, WV: AEL.

Johnston. R. C., & Viadero, D. (2000, March 15). Unmet promise: Raising minority achievement. *Education Week, 19:* 27. Retrieved from the World Wide Web on April 10, 2002: http://www.edweek.org/ew/ewstory.cfm?slug=27gapintro.h19

Kahle, J. B. (2002). Gender equity in science classroom. Retrieved from the World Wide Web on July 31, 2002: http://www.te-mat.org/kahle.shtml

Kruse, S., Louis, K. S., & Bryk, A. (1995). Let's build teachers' professional community. *WCER Highlights, 7:* 1. Retrieved from the World Wide Web on August 12, 2002: http://www.wcer.wisc.edu/publications/WCER_Highlights/Vol.7_No.1_Spring_1995/Teachers_prof_community.html

Lashway, L. (1998). Creating a learning organization. *ERIC digest.* Eugene, OR: ERIC Clearinghouse on Educational Management. Retrieved from the World Wide Web August 1, 2002: http://eric.uoregon.edu/publications/digests/digest121.html

Levine, D. U., & Lezotte, L. W. (1995). Effective schools research. In J. A. Banks & C. A. M. Banks (Eds.), *Handbook of research on multicultural education* (pp. 525-547). New York: Simon and Schuster.

Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal, 33:* 4, pp. 757-798.

Lumsden, L. (1997). Expectations for students. *ERIC digest.* Eugene, OR: ERIC Clearinghouse on Educational Management. Retrieved from the World Wide Web August 1, 2002: http://eric.uoregon.edu/publications/digests/digest116.html

McLeod, J. (1987). *Ain't no makin' it: Leveled aspirations in a low-income neighborhood.* Boulder, CO: Westview Press.

Meehan, M. L., & Cowley, K. S. (1998). Comprehensive evaluation of the Interdisciplinary Teamed Instruction Summer Institute. Charleston, WV: AEL.

Mid-Atlantic Equity Center. (1993). Beyond Title IX: Gender equity issues in schools. Chevy Chase, MD: Author. Retrieved from the World Wide Web, August 12, 2002: http://www.maec.org/beyond.html

Mills. C. W. (1959). *The sociological imagination.* New York: Oxford University Press.

Newmann, F. M., King, M. B., & Youngs, P. (2001, April). Professional development that addresses school capacity: Lessons from urban elementary schools. Presented at the annual meeting of the American Educational Research Association, New Orleans, April 28, 2000.

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). School instructional coherence: Benefits and challenges. Chicago, IL: Consortium on Chicago School Reform.

Newmann, F. M., & Wehlage, G. G. (1996). Conclusion: Restructuring for authentic achievement. In F. M. Newmann, Ed., *Authentic achievement: Restructuring schools for intellectual quality* (pp. 286-301). San Francisco: Jossey-Bass.

Nilsen, K. L. (1999). Assessing pre-intervention capacity for change in a high-need rural school district. Charleston, WV: AEL.

Northwest Regional Educational Laboratory. (1997). Closing the achievement gap requires multiple solutions. Retrieved from the World Wide Web on March 13, 2001: http://www.nwrel.org/cnorse/infoline/may97/article5.html

Ogbu, J. (1983). Minority status and schooling in plural societies. *Comparative Education Review, 27*: 2, 168-90.

O'Neil, J., & Tell, C. (1999). Why students lose when "tougher standards" win: A conversation with Alfie Kohn. *Educational Leadership, 57:* 1. Retrieved from the World Wide Web on August 19, 2002: http://www.ascd.org/pdi/demo/diffinstr/oneiltell.html

Paley, V. G. (1979). *White teacher.* Cambridge, MA: Harvard University Press.

Perez, C. (2000). Equity in standards-based elementary mathematics classrooms. Cambridge, MA: TERC. Retrieved from the World Wide Web, August 12, 2002: http://www.terc.edu/wge/equity.html

Pohan, C. A., & Aguilar, T. E. (2001). Measuring educators' beliefs about diversity in personal and professional contexts. *American educational research journal, 38:* 1, pp. 159-182.

Raywid, M. A. (1993). Finding time for collaboration. *Educational leadership, 51:* 1, 30-34.

Rickford, A. (2001). The effect of cultural congruence and higher order questioning on the reading enjoyment and comprehension of ethnic minority students. *Journal of education for students placed at risk, 6*: 4, 357-387.

Riffle, M.J.S. & Howley, C. W. (2003, July). *Pilot test of AEL's School Capacity Assessment.* Paper presented at the 12th Annual Consortium for Research on Educational Accountability and Teacher Evaluation National Evaluation Institute, Louisville, KY.

Riffle, M.J.S., Howley, C., & Ermolov, L.D. (2004, April). *Measure of School Capacity for Improvement (MSCI): Early Field Test Findings.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Riordan, C. (1997). *Equality and achievement: An introduction to the sociology of education.* New York: Longman.

Rose, A. (1999). Teaching effectively and equitably. *The National Teaching and Learning Forum, 8:* 4.

Rose, P. I. (1990). *They and we: Racial and ethnic relations in the United States*, 4th edition. New York: McGraw-Hill.

Rowan, B, & Miracle, A. W., Jr. (1983). Systems of ability grouping and the stratification of achievement in elementary schools. *Sociology of Education, 56:* 3, pp. 133-144.

Sadker, M., & Sadker, D. (1994). *Failing at fairness: How America's schools cheat girls.* New York: Charles Scribner's Sons.

Sarason, S. B., & Lorentz, E. M. (1998). *Crossing boundaries: Collaboration, coordination, and the redefinition of resources.* San Francisco: Jossey-Bass.

Schwartz, W. (2001a). Closing the achievement gap: Principles for improving the educational success of all students. *ERIC digest.* New York: ERIC Clearinghouse on Urban Education. Retrieved from the World Wide Web on March 13, 2002: http://eric-web.tc.columbia.edu/digests/dig169.html

Schwartz, W. (2001b). Strategies for improving the educational outcomes of Latinas. *ERIC digest.* New York: ERIC Clearinghouse on Urban Education. Retrieved from the World Wide Web on August 6, 2002: http://eric-web.tc.columbia.edu/digests/dig167.html

Sizer, T. R. (1999). No two are quite alike. *Educational Leadership, 57:* 1. Retrieved from the World Wide Web on August 19, 2002: http://www.ascd.org/pdi/demo/diffinstr/sizer.html

Takaki, R., Ed. (1987). *From different shores.* New York: Oxford University Press.

Tauber, R. T. (1998). Good or bad, what teachers expect from students they generally get! *ERIC digest.* Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education. Retrieved from the World Wide Web on August 6, 2002: http://www.ericsp.org/pages/digests/good_or_bad.html

Thompson, C. L., & O'Quinn, S. D. (2001). First in America special report: Eliminating the black-white achievement gap. Chapel Hill, NC: North Carolina Education Research Council.

Thorne, B. (1995). *Gender play: Girls and boys in school*. New Brunswick, NJ: Rutgers University Press.

Tomlinson, C. A. (1995). Differentiating instruction for advanced learners in the mixed-ability middle school classroom. *ERIC digest.* Reston, VA: ERIC Clearinghouse on Disabilities and Gifted Education.

Tomlinson, C. A. (1999a). The differentiated classroom: Responding to the needs of all learners. Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A. (1999b). Mapping a route toward differentiated instruction. *Educational leadership, 57:* 1. Retrieved from the World Wide Web on August 19, 2002: http://www.ascd.org/pdi/demo/diffinstr/tomlinson2.html

Tomlinson, C. A. (2000). Differentiation of instruction in the elementary grades. *ERIC digest.* Champaign, IL: ERIC Clearinghouse on Elementary and Early Childhood Education. Retrieved from the World Wide Web on January 11, 2002: http://ericir.syr.edu/plweb-cgi/obtain.pl

Tomlinson, C. A., & Kalbfleisch, M. L. (1998). Teach me, teach my brain: A call for differentiated classrooms. *Educational Leadership, 56:* 3. Retrieved from the World Wide Web on August 19, 2002: http://www.ascd.org/pdi/demo/diffinstr/tomlinson.html

University of Minnesota, Diversity Work Group. (2002). Survey of campus climate and diversity: Summary, 1997-98. Retrieved from the World Wide Web on January 11, 2002: http://webhome.crk.umn.edu?~mprada/Survey.html

University of North Carolina. (2001). School of Education and Teachers Academy: Conceptual framework for professional education. Greensboro, NC: Author. Retrieved from the World Wide Web: http://soe.uncg.edu/affiliates/teachers_academy/documents/uncg_cf_ovr_sum.pdf

van der Bogert, R. (Ed.) (1998). *Making learning communities work: The critical role of leader as learner.* San Francisco: Jossey-Bass.

Vernon-Feagans, L., Hammer, C. S., Miccio, A., & Manlove, E. (2001). Early language and literacy skills in low-income African American and Hispanic children. In S. Neuman and D. Dickinson (Eds.), *Handbook of early literacy research*. New York: Guildford Press, pp. 192-210.

Viadero, D. (2000, Marsh 29). Minority gaps smaller in some Pentagon schools. *Education Week, 19:* 29. Retrieved from the World Wide Web on April 10, 2002: http://www.edweek.org/ew/ewstory.cfm?slug=29dodd.h19

Viadero, D., & Johnston, R.C. (2000, March 22). Lags in minority achievement defy traditional explanations. *Education Week, 19:* 28. Retrieved from the World Wide Web on April 10, 2002: http://www.edweek.org/ew/ewstory.cfm?slug=28causes.h19

Viadero, D., & Johnston, R.C. (2000, April 5). Lifting minority achievement: Complex answers. *Education Week, 19:* 30. Retrieved from the World Wide Web on April 10, 2002: http://www.edweek.org/ew/ewstory.cfm?slug=30gap.h19

Willis, P. (1981) *Learning to labor: How working class kids get working class jobs*. New York: Columbia University Press.

Wilson, R. J., & Martinussen, R. L. (1999). Factors affecting the assessment of student achievement. *Alberta Journal of Educational Research, 45:* 3, 267-277.