

MEASURE OF SCHOOL CAPACITY FOR IMPROVEMENT (MSCI): EARLY FIELD TEST FINDINGS

Riffle, M. J. S.

Howley, C. W.

Ermolov, L. D.



P. O. Box 1348 • Charleston WV 25325 • 800-624-9120

Paper presented April 14, 2004 at the 85th Annual Meeting of the American Educational Research Association, San Diego, CA. For further information, please e-mail the authors at howleycw@ael.org, riflej@ael.org, or ermolovl@ael.org.

ABSTRACT

This study was conducted to confirm the validity and reliability of the 64-item Measure of School Capacity for Improvement (MSCI). The MSCI was designed to assess the degree to which schools possess the potential to become high performing learning communities, and was developed in response to a paucity of definition, operationalization, and assessment of school capacity in the education research and evaluation literature. The MSCI offers an operationalization of the concept of school capacity. The MSCI was administered to 1,274 professional staff affiliated with 12 elementary, 10 middle, and 13 high schools in Tennessee that were low performing, predominantly African American, and low SES. Results of factor analysis and estimates of internal consistency suggest that the MSCI and 58 of the original items, which comprise six subscales, are very reliable and has construct validity. Moreover, the MSCI holds some promise for providing a much-needed means for discerning schools with the resources, practices, and proclivities to successfully undertake serious development from those who might better focus their energies on first addressing the issues measured by the various MSCI subscales.

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).

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).

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).

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 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 (CSRDP). 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 Office of Educational Research and Improvement's 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, 1997; deMarrais & LeCompte; 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 is intended for administration to K-12 school professional staff. Data from administration of the survey are 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 will allow staff to identify dimensions of school capacity in need of further development in their schools. The instrument is intended for diagnostic use, for instance at the beginning of school reform efforts. It is also 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 previously validated. 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)

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, 1997; 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).

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 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 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 is 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 (deMarrrais & 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 North Carolina 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 (1999, 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 (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, 1997; 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.

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 is 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 one, strongly disagree, to four, 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 possess 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 correlate. 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.

The SCA has been 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. The current study reports the results of an early field test of this version of the instrument (renamed the Measure of School Capacity for Improvement or MSCCI) conducted in the spring of 2003.

METHOD

Participants

A total of 1,274 professional staff representing 35 schools (12 elementary, 10 middle, and 13 high school) from six districts completed the survey. Three hundred eighty-six respondents worked in an elementary school, 250 were from a middle school, and 638 were from a high school. The majority of respondents (n=912) were regular classroom teachers, with the remaining respondents fitting into the categories of special education teacher (n=110), counselor (n=43), principal/assistant principal (n=39), librarian/media specialist (n=25), and other (n=107). Approximately half of the respondents held a Master's, Master's + 15, or Master's + 30 or more (n=646), while slightly less than half held a Bachelor's, Bachelor's + 15, or Bachelor's + 30 or more (n=525). The remaining respondents (n=65) had a doctorate, categorized themselves as education specialist, or responded other.

Almost three-quarters of the respondents were female (n=885), while slightly more than a quarter were male (n=344). More than half of the respondents classified themselves as Black or African American (n=647) with slightly less classifying themselves as White (n=518). The remaining respondents (n=58) categorized themselves as Asian, Hispanic or Latino/a, Native Hawaiian or other Pacific Islander, American Indian or Alaska Native, or other.

About one-quarter of participants (n=295) 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=205) and one to three years (n=193). 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=433) with slightly less reporting that they had taught in their current school for four to six years (n=258). In relation to how long participants had worked in a particular district, almost one-quarter (n=285) had worked in the district between one and three years, while somewhat less had worked in the district between four and six years (n=239) and more than 25 years (n=202).

More than one-quarter of respondents (n=302) noted that they had taught their current subject from one to three years and a little less (n=218) had taught their current subject between four and six years. Similarly, 340 respondents noted that they taught their current grade from one to three years and 234 had taught their current grade from four to six years. In connection with this information, an overwhelming majority of respondents were certified in the grade(s) they currently teach as well as the subject area(s) they currently teach (n=1057 and n=1031, respectively).

Test-retest Participants

A total of 174 professional staff representing schools (three elementary, 2 middle, and 2 high schools) from three districts completed the survey for test-retest purposes. Eighty-four respondents worked in an elementary school, 47 were from a middle school, and 43 were from a high school. The majority of respondents (n=128) were regular classroom teachers, with the remaining respondents fitting into the categories of special education teacher (n=12), principal/assistant principal (n=8), librarian/media specialist (n=6), counselor (n=3), and other

(n=17). Approximately half of the respondents held a Master's, Master's + 15, or Master's + 30 or more (n=86), while slightly less held a Bachelor's, Bachelor's + 15, or Bachelor's + 30 or more (n=80). The remaining respondents (n=3) had a doctorate, categorized themselves as education specialist, or responded other.

More than three-quarters of the respondents were female (n=135), while slightly less than one-quarter were male (n=39). More than three-quarters of the respondents classified themselves as White (n=136) with less than one-quarter classifying themselves as Black or African American (n=34). The remaining respondents (n=3) categorized themselves Hispanic or Latino/a, or other.

About one-quarter of participants (n=40) 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=29) and one to three years (n=23). 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=39) or four to six years (n=39). In relation to how long participants had worked in a particular district, almost one-quarter (n=33) had worked in the district between four and six years, while somewhat less had worked in the district between one and three years (n=29) and more than 25 years (n=28).

More than one-quarter of respondents (n=38) noted that they had taught their current subject from one to three years and a little less (n=33) had taught their current subject between four and six years. Similarly, 37 respondents noted that they taught their current grade from one to three years and 37 had taught their current grade from four to six years. In connection with this information, an overwhelming majority of respondents were certified in the grade(s) they currently teach as well as the subject area(s) they currently teach (n=154 and n=144, respectively).

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 four-point Likert-type scale ranging from one indicating "Not at all True" to four indicating "Almost Always True." For the remaining items, professional staff are asked to rate how often each item is true for their school using a similar four-point Likert-type scale ranging from one indicating "Never True" to four indicating

“Frequently True.” Participants are also asked to respond to additional demographic items. The survey is formatted for machine scoring.

The AEL Continuous School Improvement Questionnaire (AEL CSIQ) is a 60-item, machine scannable, field-tested and validated instrument used to help school staff gauge its performance on six dimensions related to continuous school improvement (Meehan, Cowley, Craig, Balow, & Childers, 2002). It consists of six subscales described below.

- **Shared Leadership.** This subscale reflects the extent to which leadership is viewed as being shared. It assesses whether school administrators dominate decision making or if there are mechanisms for involving teachers, students, and parents. Opportunities for leadership development among the members of the school community are assessed, as are the degree to which information is shared and the extent to which school administrators listen and solicit the input of others.
- **Effective Teaching.** This subscale ascertains the extent to which teacher practice is aligned with research on effective teaching. It assesses whether teachers actively engage students in a variety of learning tasks, pose questions that encourage reflection and higher order thinking, expect students to think critically, and use teaching strategies designed to motivate students.
- **School/Family/Community Connections.** This subscale assesses the extent to which parents and community members are involved and feel part of the school. It reflects the degrees to which they are kept informed, meaningful partnerships exist, communication is open, and diverse points of view are honored and respected.
- **Purposeful Student Assessment.** This subscale reflects the extent to which student assessment data are meaningful; are used by teachers to guide instructional decisions; and are communicated to and understood by the greater school community, including teachers, parents, students, and other members of the community.
- **Shared Goals for Learning.** This subscale assesses the extent to which the school has clear, focused goals that are understood by all members of the school community. In addition, it reflects whether shared goals affect what is taught and how teachers teach, drive decisions about resources, focus on results for students, and are developed and “owned” by many rather than a few.
- **Learning Culture.** This subscale reflects whether the culture of the school promotes learning by all—students, staff, and administration. It reflects the extent to which the school emphasizes learning rather than passive compliance, is a safe but exciting place to be, and encourages curiosity and exploration. In addition, it indicates the extent to which teachers have opportunities and encouragement to reflect on practice, work with others, and try new ways of teaching.

Data Collection

Some 2,200 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 March 2002 and sent with the copies of the instrument and envelopes. Please refer to the Appendix copies of each.

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. This procedure was followed in four of the seven schools. For the remaining three schools, the TN EE collected the surveys after each administration without the use of envelopes (i.e., the surveys from the first administration were all packaged together, as were the surveys from the second administration).

Upon receipt of the test-retest surveys that were not collected according to the procedures designed to ensure confidentiality and matching instruments, AEL staff used self-reported identification numbers, as well as handwriting analysis to pair as many of the surveys as possible for use in the test-retest analysis. Those surveys with different identification numbers but matching handwriting samples were force matched to the first administration to create a pair for analysis. If the surveys had different identification numbers and different or no handwriting samples, the surveys were not used for test-retest purposes. Finally, while cleaning the data, staff noticed that the demographic information did not match for some of the pairs. Upon this discovery, staff decided to force match the demographic data to that reported on the first administration. This decision was made based on a belief that respondents were more likely to be honest on the first administration, as well as the increased instance of missing data on the second administration.

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. Concurrent validity was explored by calculating correlations between and shared variances of subscales of the MSCI and the CSIQ.

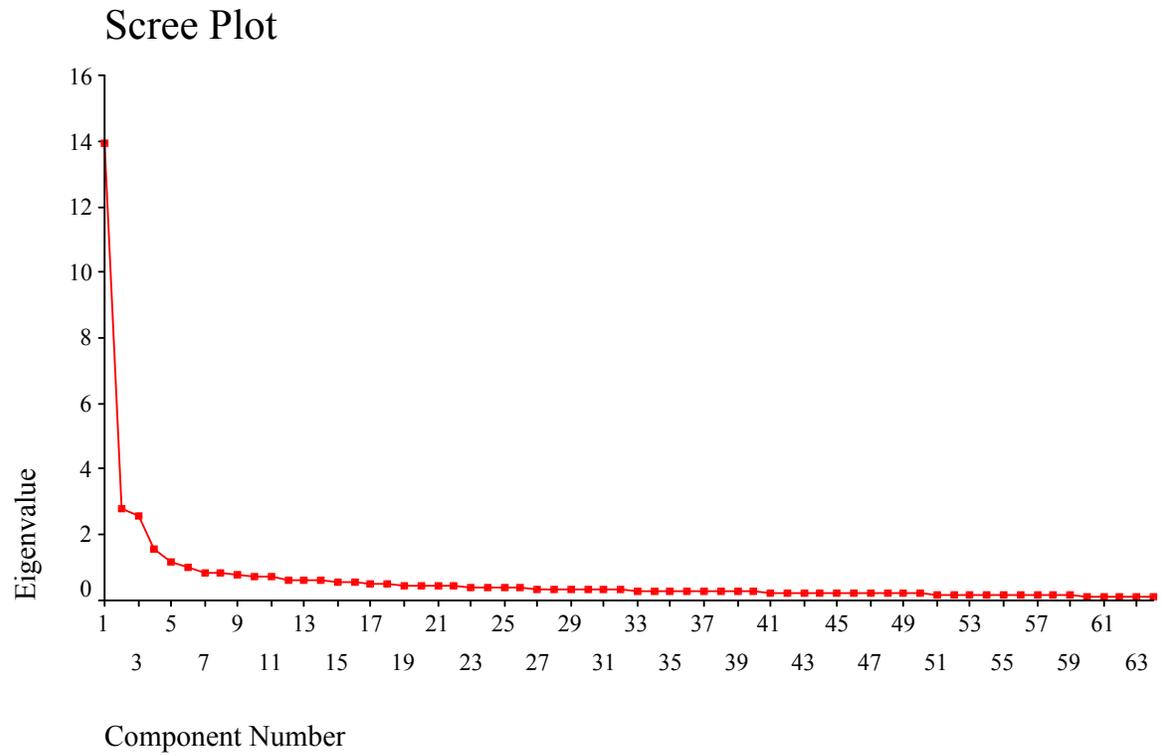
FINDINGS

Internal Consistency Reliability

Internal consistency of the MSCI and its eight subscales was estimated with the Cronbach's alpha coefficient. The alphas showed that the MSCI itself ($\alpha = .97$) and its subscales were very reliable, with alphas ranging from .79 for the Program Coherence and Technical Resources Scales to .91 for Differentiated Instruction. Remaining alphas were .85 for Collective Professional Capacity, .83 for Peer Reviewed Practice, .86 for Anti-Discriminatory Teaching, Responsive Pedagogy, and Expectations for Student Performance.

Construct Validity--Factor Analysis

Factor analysis using principal component analysis with oblimin rotation was conducted. Factors were expected to be closely related to each other as they were all hypothesized to be related to school capacity for improvement. Initial results revealed ten factors with eigenvalues greater than one. However, two factors accounted for less than 3% of the total variance and consisted of fewer than three items. Therefore, a secondary factor analysis was conducted using the same method, but forcing eight factors. Results of this analysis revealed that factor loadings ranged between .34 and .86 for all items. However, two factors were composed of three items each and made very small contributions to the 45% total variance accounted for (1.4 and 2.5% respectively). Thus, six of the eight factors appeared to be fairly robust.



Interestingly, all of the items designed to assess expectations for student performance and all but one of those that were designed to assess collective teacher capacity loaded on the first factor, with loadings ranging from .68 to .86. This suggested that teachers' expectations for students' performance are closely tied to their beliefs regarding the collective capacity of the faculty to teach students effectively and appropriately. Therefore, it appears that the underlying construct is that of collective professional capacity, albeit professional capacity that includes an evaluation of students' academic capacities.

The second factor consisted of four items designed to assess peer-reviewed practice with loadings that ranged from .65 to .67. Thus, this subscale name was retained. The third factor, like the first, consisted of all the items designed to assess Anti-discriminatory Teaching, seven of the items designed to assess Responsive Pedagogy, and one item from the Peer Reviewed Practice subscale. All of the items had loadings that ranged from .52 to .64 and their content appeared to reflect the degree to which faculty understand diversity and engage in classroom practices that equitably support the learning of all students. In the early stages of the development of the SCA/MSCI, Howley and Riffle (2002) had originally conceived these items as part of a larger set of 38 that they named Equitable Practice. Therefore, this subscale name was reinstated to describe the third factor.

The fourth factor consisted of four items from the Technical Resources subscale with loadings ranging from .48 to .51, and all reflected the availability of adequate materials and equipment. None of the items that assessed having sufficient time allotted to engage in professional sharing or collaboration loaded on this scale. The fifth factor consisted of two of the items from the original Technical Resources scale that reflected sufficient time for professional exchanges, as well as six of the items from the program coherence subscale and one item from the Peer Reviewed Practice subscale. Factor loadings ranged from .43 to .47. Upon reflection, these items all appeared to relate to the degree of coordination among a school's programs for student and staff learning, focus on clear learning goals, and are sustained over time. Therefore, the subscale name Program Coherence was retained for this factor. The sixth and final factor consisted of one item from the original Collective Professional Capacity scale that reflected school staff's persistence if a child did not seem to want to learn, and all of the items originally designed to assess Differentiated Instruction. Thus, this subscale name was retained for the sixth and final scale with factor loadings ranging from .37 to .43. Each of the two excluded factors consisted of one item each from the original Program Coherence, Peer Reviewed Practice, and Technical Resources subscales, but neither set of items appeared to reflect any consistent construct.

Table 1

Factor Loadings* for Revised MSCI Subscales

No. Items	<i>Collective Professional Capacity</i>	<i>Peer Reviewed Practice</i>	<i>Equitable Practice</i>	<i>Technical Resources</i>	<i>Program Coherence</i>	<i>Differentiated Instruction</i>
1	.86	.67	.64	.51	.47	.43
2	.85	.67	.64	.49	.47	.41
3	.80	.66	.63	.49	.46	.41
4	.80	.65	.63	.48	.46	.41
5	.79		.63		.46	.40
6	.77		.63		.45	.38
7	.77		.63		.44	.38
8	.76		.62		.44	.38
9	.75		.57		.43	.37
10	.73		.57			
11	.72		.56			
12	.71		.54			
13	.71		.54			
14	.71		.53			
15	.70		.53			
16	.68		.52			

*Loadings equal to or greater than .30 are reported

Correlations Among Factors and Total MSCI Scores

As shown in Table 2, the correlations between each of the newly created subscales and total MSCI scores ranged from .39 for the Equitable Practice and Technical Resources subscales, to .83 for the Collective Professional Capacity and Differentiated Instruction subscales. As one would expect, all of the subscales were well related to the total MSCI score, with Technical Resources and Peer Reviewed Practice having the smallest correlations with the total MSCI. This is not surprising, given that the sample consists of respondents from low-performing schools where these may not be available or deemed important.

Table 2

Inter-Correlations of the Six Revised MSCI Subscales

Subscales	<i>Collective Professional Capacity</i>	<i>Peer Reviewed Practice</i>	<i>Equitable Practice</i>	<i>Technical Resources</i>	<i>Program Coherence</i>	<i>Differentiated Instruction</i>
<i>Collective Professional Capacity</i>	--					
<i>Peer Reviewed Practice</i>	.42*	--				
<i>Equitable Practice</i>	.61*	.42*	--			
<i>Technical Resources</i>	.43*	.53*	.39*	--		
<i>Program Coherence</i>	.58*	.66*	.57*	.64*	--	
<i>Differentiated Instruction</i>	.83*	.45*	.69*	.49*	.60*	--
<i>Total MSCI</i>	.88*	.65*	.82*	.64*	.81*	.89*

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

Reliability estimates based on Cronbach's alpha for the revised MSCI improved slightly such that the overall alpha was .97. The 16-item Collective Professional Capacity scale had a new alpha of .94. Equitable Practice (16 items) and Differentiated Instruction (9 items) also appeared highly reliable with alphas of .92 and .91 respectively. Peer Reviewed Practice (4 items) and Program Coherence (9 items) were slightly less reliable with alphas of .84 and .86 respectively. The Technical Resources subscale was reliable with an alpha of .78, although less so than each of the other revised subscales.

Test-Retest Reliability

The correlation between total MSCI scores on the two administrations of the survey was .87 ($p = .000$) based on 125 respondents who completed all items. Thus, participants' responses on the two testing occasions appear to have remained quite stable over time. Correlations by subscales' mean scores from the two administrations are presented in the table below and range from .68 for Anti-Discriminatory Teaching to .86 for Technical resources. Thus, original subscale scores appear to have adequate reliability over time.

Table 3

Descriptive Information and Correlation Coefficients for the Original Eight Subscales

Subscales	1 st Administration			2 nd Administration			Correlation Coefficient
	N	Mean	SD	N	Mean	SD	
Collective Professional Capacity	174	2.99	.55	174	2.95	.54	.77*
Expectations for Student Performance	172	3.00	.64	173	2.94	.58	.76*
Peer Reviewed Practice	174	2.91	.65	174	2.90	.62	.77*
Equitable Practice	174	3.16	.52	174	3.18	.54	.78*
Anti-Discriminatory Teaching	174	3.41	.50	174	3.43	.49	.68*
Technical Resources	174	2.59	.66	174	2.62	.61	.86*
Program Coherence	174	2.97	.55	174	2.94	.48	.75*
Differentiated Instruction	172	3.12	.62	173	3.07	.59	.76*

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

Results of a similar analysis based on 58 items and the six factors revealed by the factor analysis show improved stability in subscale scores over time ranging from .76 to .85. Overall, the changes had no apparent effect on the stability of total scores on the MSCSI ($r=.87$, $p=.000$). As before, the Technical Resources subscale had the most stability across the two occasions ($r = .85$, $p < .01$).

Table 4

Descriptive Information and Stability of the Six Revised MSCSI Subscales Across Administrations

Revised MSCSI Subscales	1 st Administration			2 nd Administration			Correlation Coefficient
	N	Mean	SD	N	Mean	SD	
Collective Professional Capacity – FACTOR 1	174	2.95	.58	174	2.91	.55	.77*
Peer Reviewed Practice – FACTOR 2	174	2.65	.86	173	2.60	.81	.77*
Equitable Practice – FACTOR 3	174	3.34	.49	174	3.35	.49	.76*
Technical Resources – FACTOR 4	174	2.79	.78	174	2.82	.72	.85*
Program Coherence – FACTOR 5	174	2.95	.64	174	2.93	.58	.79*
Differentiated Instruction – FACTOR 6	172	3.13	.61	173	3.07	.58	.76*

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

Concurrent Validity

The MSCI was expected to be predictive of scores on the AEL CSIQ because the capacity for school improvement is likely to predict successful engagement in continuous school improvement efforts. As expected, the correlation between overall mean scores on the two instruments was .68 ($p = .000$), with the MSCI accounting for some 47% of the variance in the CSIQ. Correlations between the subscales of the two instruments ranged from .36 between Peer Reviewed Practice and Effective Teaching, to .61 between Differentiated Instruction and Effective Teaching. As shown in Table 5, all correlations were significant and positive. To facilitate the recognition of patterns among these relationships, Table 6 presents the shared variance (r^2) between the MSCI and CSIQ subscales.

Table 5

Correlations Between Subscales of the AEL CSIQ and the Revised AEL MSCI

<i>Revised AEL MSCI Subscales</i>	<i>AEL CSIQ Subscales</i>					
	<i>Learning Culture</i>	<i>School / Family / Community Connections</i>	<i>Shared Leadership</i>	<i>Shared Goals for Learning</i>	<i>Purposeful Student Assessment</i>	<i>Effective Teaching</i>
<i>Collective Professional Capacity</i>	.55* ($n=478$)	.53* ($n=478$)	.45* ($n=478$)	.44* ($n=478$)	.48* ($n=478$)	.54* ($n=477$)
<i>Peer Reviewed Practice</i>	.42* ($n=472$)	.39* ($n=472$)	.40* ($n=472$)	.40* ($n=472$)	.43* ($n=472$)	.36* ($n=471$)
<i>Equitable Practice</i>	.55* ($n=477$)	.45* ($n=477$)	.41* ($n=477$)	.45* ($n=477$)	.47* ($n=477$)	.52* ($n=476$)
<i>Technical Resources</i>	.45* ($n=478$)	.44* ($n=478$)	.41* ($n=478$)	.43* ($n=478$)	.45* ($n=478$)	.38* ($n=477$)
<i>Program Coherence</i>	.61* ($n=478$)	.58* ($n=478$)	.56* ($n=478$)	.62* ($n=478$)	.60* ($n=478$)	.56* ($n=477$)
<i>Differentiated Instruction</i>	.60* ($n=476$)	.49* ($n=476$)	.42* ($n=476$)	.49* ($n=476$)	.51* ($n=476$)	.61* ($n=475$)

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

Table 6

Shared Variance (r^2) between the Revised MSCI and AEL CSIQ Subscales

<i>Revised AEL MSCI Subscales</i>	<i>AEL CSIQ Subscales</i>					
	<i>Learning Culture</i>	<i>School / Family / Community Connections</i>	<i>Shared Leadership</i>	<i>Shared Goals for Learning</i>	<i>Purposeful Student Assessment</i>	<i>Effective Teaching</i>
<i>Collective Professional Capacity</i>	.30	.28	.20	.19	.23	.29
<i>Peer Reviewed Practice</i>	.18	.15	.16	.16	.18	.13
<i>Equitable Practice</i>	.30	.20	.17	.20	.22	.27
<i>Technical Resources</i>	.20	.19	.17	.18	.20	.14
<i>Program Coherence</i>	.37	.34	.31	.38	.36	.31
<i>Differentiated Instruction</i>	.36	.24	.18	.24	.26	.37

Generally speaking, the Program Coherence Subscale of the MSCI shared the largest portion of the variance with each of the CSIQ subscales, and Peer Reviewed Practice the least. **Collective Professional Capacity** was most closely related to Learning Culture, School/Family/Community Connections, and Effective Teaching. This is not surprising given that a faculty's belief in its shared capability to positively influence student learning, including their expectations that for student performance are likely to be related to their perceptions and understanding of the community in which students live, contribute to a positive, safe, school culture, and to their effectiveness as teachers.

Peer-Reviewed Practice assessed the frequency with which teachers and supervisors observe staff's classes to provide meaningful feedback and improve teaching. Initially, the comparatively low correlation with effective teaching seems counterintuitive. However, this may be an artifact of administering the two surveys only to a sample of very low performing schools. Although opportunities for teachers to observe each other and collaborate effectively are somewhat limited in most schools today, when such deprivatization occurs it is often because the schools either were or are currently receiving some kind of external facilitation or professional development designed to improve their teaching. In the current political climate, funding for professional development is provided primarily to the lowest performing (in theory the least effective) schools whose capacities to engage in school improvement may be hampered by more basic issues such as a shortage of qualified teachers, or deteriorating building facilities. This may explain the small proportion of shared variance between Effective Teaching on the AEL CSIQ and Peer Reviewed Practice on the Revised MSCI.

Program Coherence evaluated the extent to which a school's programs for student and staff learning are coordinated, focused on clear learning goals, and sustained over time. Therefore, as shown in Tables 5 and 6, one might expect this construct to be strongly tied to Shared Goals for Learning and Purposeful Student Assessment on the AEL CSIQ. Moreover, the foundational nature of program coherence is supported by the tendency of the construct to be closely related to all of the constructs measured by the subscales of the CSIQ.

Technical Resources measured the availability to faculty of planning time, working equipment, technology, instructional materials, facilities, and professional resource materials, such as journals. Schools with the best resources may also be those that are more strongly supported by their communities (e.g., by engaging in more fundraising or facility improvement activities). The presence of important technical resources also seems likely to contribute to a positive school climate in which faculty and students feel safe, both physically, and in terms of opportunities to experiment and explore new instructional methods for example. This seems particularly likely because the MSCSI defines Technical Resources as including both physical and collaborative types. As shown in Tables 5 and 6, the weaker relationship between Technical Resources and Effective Teaching is not entirely surprising since it is widely realized that such resources are beneficial and may facilitate engagement in continuous school improvement without being absolutely essential for its' success.

The **Equitable Practice** subscale assesses the degree to which faculty understand diversity and engage in classroom practices that equitably support the learning of all students. It includes school staff's responsiveness to their students' communities, the creation of equitable classroom environments, and pluralistic language and text use. Accordingly, one would expect this subscale to be strongly related to School/Family/Community Connections, as well as the creation and maintenance of a positive Learning Culture. In addition, equitable practices that recognize student diversity are likely to contribute to Teacher Effectiveness, particularly in high minority and/or low income schools like those included in the present sample. These relationships are supported by the data.

Finally, **Differentiated Instruction** evaluates the extent to which faculty modify their instructional strategies and grouping arrangements to meet the learning needs of students. Such behaviors are likely to be strongly related to Teacher Effectiveness, and the existence of a Learning Culture that encourages this kind of flexibility and experimentation with instruction. These relationships are evident in the data reported in Tables 5 and 6.

Descriptive Statistics

Mean scores and standard deviations for both the original and revised subscales of the MSCSI are presented in Table 7 and are all relatively high, ranging from 2.6 to 3.3 on the four point Likert scales. Given that these participants were all professionals at very low performing schools, all of these results must be interpreted with caution. However, the tendency of scores for low performing schools to cluster around the third point of a four-point scale suggests that a four-point scale is not sufficiently sensitive to reflect differences between participants' ratings of their schools' capacities to improve. It seems likely that using a six or seven point Likert scale

would be more sensitive and would allow participants to make more finite distinctions in their responses to the survey items.

Table 7

Mean Scores Based on the Revised and Original MSCI Subscales

Subscales	Revised			Original		
	N	Mean	SD	N	Mean	SD
Collective Professional Capacity--FACTOR1	1165	2.87	0.58	1274	2.93	0.55
Student Performance Expectations				1269	2.89	0.64
Peer Reviewed Practice--FACTOR2	1217	2.62	0.80	1272	2.87	0.62
Equitable Practice/Responsive Pedagogy--FACTOR3	1156	3.27	0.51	1274	3.12	0.53
Anti-Discriminatory Teaching				1272	3.31	0.54
Technical Resources--FACTOR4	1230	2.81	0.70	1274	2.66	0.58
Program Coherence--FACTOR5	1179	2.96	0.58	1273	2.97	0.52
Differentiated Instruction--FACTOR6	1211	3.04	0.57	1269	3.03	0.58

Educational or Scientific Importance of the Study

The significance of this study lies in its validation of an instrument assessing school capacity for improvement. It both offers an operationalization of the concept of school capacity, a notion which has received much citation but little substantiation, and provides a means for discerning schools with the resources, practices, and proclivities to successfully undertake serious development from those who might better focus their energies on first addressing the issues measured by the various MSCI subscales.

Suggested Directions for Future Work

Results of this field test suggest that the MSCI and 58 of the original items that comprise six subscales have a high degree of internal consistency, are quite stable over time, and are predictive of successful engagement in continuous school improvement as measured by the AEL CSIQ. Although these results are promising, the instrument's ability to identify schools with a great deal of capacity for reform and thus stronger likelihood of being described as high performing, is as yet only partially clear. The samples used to validate the MSCI have tended to include mostly low performing schools, which is likely to bias results. Ideally, the MSCI will prove valid and reliable for schools found at all points on the continuum between low and high performing. Therefore, it is recommended that future studies of the MSCI include the establishment of validity with more varied samples including known groups of schools that perform at high, moderate, and lower levels. In addition, it appears that the four-point Likert type response options on the MSCI may not generate sufficient variance to distinguish between low and high performing schools. Therefore, further psychometric studies of the MSCI should offer respondents a wider range of response options of perhaps up to six points. Finally, to improve the utility of the MSCI for researchers and practitioners alike, norms should be established against which schools may compare themselves as they undertake change efforts.

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.
- 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 from 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.
- 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.ericasp.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.

APPENDIX

October 31, 2003

Dear Exemplary Educator:

We want to thank you very much for your help distributing and administering the AEL CSIQ and MSCI surveys to the schools with which you work. Enclosed you will find sufficient copies of the MSCI for each professional staff member in your county that is employed in the school to complete it twice, as well as extra copies for each school. Half of the copies are printed on green paper and half on pink paper. We have also enclosed two sizes of brown envelopes (9x12 and 10x13) to use in the administration of the survey along with large white envelopes that you can use to return the surveys to AEL.

We ask that you administer the MSCI on two separate occasions. The first time the **green** MSCI should be completed along with the CSIQ. After completion, each staff member should place his or her completed surveys in one of the small (9x12) brown envelopes provided, seal the envelope, and write his or her name across the seal on the back. We ask that you collect these envelopes and keep them sealed in a safe place for approximately two weeks.

At that time, we would like you to return his or her sealed envelope to each staff member and instruct them to open the envelope, ensure that the surveys contained therein are indeed his or her own by looking at the ID number each provided, then take those surveys and seal them in a second small (9x12) brown envelope that you provide. This helps to ensure participants that their responses have remained confidential and that their names will not be associated with the data from this point onward.

After each participant has sealed his or her surveys in the second brown envelope, we ask that you distribute the **pink** copy of the MSCI to the participants and ask them to complete it again. When each has completed the pink MSCI, each person should seal his or her newly completed survey, along with the envelope containing the two previously completed surveys in a third larger (10x13) brown envelope.

It is important that participants be asked to avoid looking at their original survey responses and look only at their ID numbers on the surveys. We are interested in finding out how reliable responses on this survey are over time and how much they may

be influenced by uncontrollable factors associated with the circumstances of the MSCI administration.

When you have collected all participants' large (10x13) brown envelopes (each of which will contain the small (9x12) brown envelope with the two previously completed surveys, and a newly completed, pink copy of the MSCI), we ask that you seal these large (10x13) brown envelopes in the white envelope provided and return them to AEL.

These instructions are summarized in bulleted form on the enclosed document. However, if you have any questions about the process please contact Joy Riffle, Lisa Ermolov, or Jim Craig, at AEL at 1-800-624-9120.

Thank you again for your help. If you have any questions or concerns about the surveys, or the process for their completion please do not hesitate to contact us.

Sincerely,

Joy Riffle
Research and Evaluation Specialist

Cc: Jim Craig
Merrill Meehan
Lisa Ermolov
Steve Moats

Instructions for EEs Administering the MSCI

1. Administer the **green** MSCI along with the CSIQ.
2. After completion, each staff member places his or her completed surveys in a small (9x12) brown envelope, seals the envelope, and writes his or her name across the seal on the back.
3. You collect these envelopes and keep them sealed in a safe place for approximately two weeks.
4. After the two weeks have gone by, you meet with staff again, and return each participant's sealed envelope to him or her.
5. You instruct participants to open the envelopes and look at the ID numbers they provided to verify that the surveys contained therein are indeed their own.
6. Participants are then instructed to take those previously completed surveys and seal them in a new small (9x12) brown envelope so that their names are no longer associated with the data. This helps to ensure participants that their responses have remained confidential and that their names will not be associated with the data from this point onward.
7. Distribute the **pink** copy of the MSCI to the participants and ask them to complete it.
8. **It is important that participants be asked to avoid looking at their original survey responses and look only at their ID numbers on the surveys.** We are interested in finding out how reliable responses on this survey are over time and how much they may be influenced by uncontrollable factors associated with the circumstances of the MSCI administration.
9. When everyone has completed the **pink** MSCI, each person should seal this pink copy and the small (9x12) envelope containing the two previously completed surveys in a large (10x13) brown envelope.
10. Collect all participants' large brown envelopes (each of which will contain a small brown envelope with the two previously completed surveys, and a newly completed, pink copy of the MSCI).
11. Seal these large brown envelopes in the large, expandable white envelope provided and return them to AEL.
12. The envelopes that have names on them should be empty and can then be destroyed. They should **not** be returned with the data to AEL.