Comprehensive School Reform in Middle Schools: The Effects of Different Ways of Knowing on Student Achievement in a Large Urban District

Marco A. Muñoz
Jefferson County Public Schools

Steven M. Ross and Aaron J. McDonald
The University of Memphis

Comprehensive school reform (CSR), a federally funded program, is designed to raise student achievement via the implementation of whole school reforms. Unfortunately, although close to 400 CSR models have been adopted by schools nationally, there is relatively limited empirical evidence from rigorous research studies regarding the effectiveness of CSR in general or of specific models in particular. Of additional importance is the extent to which CSR models can demonstrate positive effects over time. The purpose of this study is to evaluate the effects on student achievement and attendance, in an urban school district in Kentucky, of the Year 3 implementation of the Different Ways of Knowing (DWoK) for the Middle Grades model. Results indicated that achievement results on both the Comprehensive Test of Basic Skills and Kentucky Core Content Test were significantly superior for the DWoK schools relative to matched control schools; no positive impacts on student attendance occurred for Year 3. The implications of the study lies in revealing that learner-centered CSR models that integrate the arts with the core curriculum can produce observable achievement gains in both normed- and criterion-referenced tests.

As emphasized in the 2001 No Child Left Behind (NCLB) legislation and the Education Sciences Reform Act (ESRA, 2002), it is critically important to identify

Correspondence should be addressed to Marco A. Muñoz, Jefferson County Public Schools, 3332 Newburg Rd, Louisville, KY 40218. E-mail: marco.munoz@jefferson.kyschools.us.
what works in bringing all students to proficiency levels in core subjects. One philosophy is that enacting effective educational change must involve an entire school rather than a collection of isolated programs (Borman, Carter, Aladjem, & LeFloc'h, 2004; Desimone, 2002; Ross & Gil, 2004; Rowan, Camburn, & Barnes, 2004). The use of whole-school reform models was boosted substantially by the enactment in 1997 of the Comprehensive School Reform Demonstration (CSRD) program (U.S. Department of Education, 1999). In 1999–2000, CSRD funded more than 1,700 schools nationally with a minimum of $50,000 for each of 3 years (Hatch, 2001). To date, close to 400 models have been adopted using CSRD support (Desimone, 2002). Comprehensive school reform (CSR) programs encompass virtually all aspects of school operations, including instruction, assessment, classroom management, professional development, parental involvement, school management, and curriculum.

Although the philosophy and research-based grounding of CSR approaches should be conducive to schools improving student achievement, the evidence base in support of the impact of such models still remains limited (e.g., Berends, Chun, Schuyler, Stockley, & Briggs, 2002; Borman, Hewes, Overman, & Brown, 2003; Desimone, 2002; Herman, 1999). Recently, for example, Borman et al. (2003) conducted a meta-analysis on CSR effects and found an overall achievement advantage ranging from one-tenth to one-seventh of a standard deviation for the CSR models examined. Models classified as having the strongest evidence of yielding benefits were Direct Instruction, Success for All, and the School Development Program, with effects sizes (ES) in the +0.15 to +0.21 range. It is discouraging that 17 out of the 29 models were classified as having insufficient statistically reliable or generalizable results to judge effects.

The lack of convincing empirical evidence for an educational program may be more of a reflection of limited interest in or opportunities available for research on that program than evidence of its ineffectiveness. In concert with its emphasis on accountability for schools, NCLB (2001) explicitly advocated the use of rigorous, scientifically based research for determining which educational programs work to raise student achievement (Eisenhart & Towne, 2003; Feuer, Towne, & Shavelson, 2002). The purpose of this study, therefore, is to examine the effectiveness in raising student achievement of a newly developed CSR model, Different Ways of Knowing (DWoK) for the Middle Grades.

The middle grades clearly bring unique challenges for student learning and adjustment, associated with the major psychological, social, and biological changes that impact students during these years (Simmons & Blyth, 1987). Yet relatively few CSR and schoolwide reforms have been directed to middle school education (Bodilly, 1998; Bodilly & Berends, 1999). The major research question that guided this study was: What is the impact of the DWoK model on raising middle school student achievement? We addressed this question via a rigorous quantitative analysis of student achievement on state-mandated tests. A description of the DWoK for the Middle Grades model follows.

DWoK is a 10-week project designed to improve students’ support for middle school students in improving students’ understanding and application of multiple intelligences in art, music, math, basic literacy, and science. The core of the program is to provide students with a motivating experience that is aligned with the curriculum.

DWoK is a 10-week project designed to improve students’ understanding and application of multiple intelligences in art, music, math, basic literacy, and science. The core of the program is to provide students with a motivating experience that is aligned with the curriculum.
THE DIFFERENT WAYS OF KNOWING MODEL

DWoK is a multiyear, comprehensive design for teacher quality and school improvement (see Gafe Institute, 2005). The original, elementary-grades model was developed in 1989 by the Gafe Institute, a nonprofit educational organization dedicated to school reform. As described by Herman (1999), DWoK builds on the "multiple intelligences" of students, to develop their skills in different areas (e.g., artistic, mathematical, social, language). Thematic units that integrate the learning of basic literacy and mathematics skills with artistic experiences are strongly emphasized. Specifically, according to Herman (1999), "The curriculum, which is organized around history and social studies, seeks to integrate the arts, literature, science, and math, and technology" (p. 59). DWoK met the rigorous standards needed to qualify as a New American Schools design and is a member of the Public Education Network.

DWoK for the Middle Grades was developed starting in 2000 under a $13 million contract with the U.S. Department of Education. Consistent with Turning Points 2000 (Jackson & Davis, 2000), the overall goal was to improve student achievement by providing educational experiences for young adolescents that are responsive to their academic, developmental, and social needs. The developer advocated building on the multiple intelligences of students, to develop their skills in various domains (e.g., logic and mathematics, language, social skills, and artistic skills).1 Accordingly, the approach is built around a variety of research bases, including those focusing on cognition, the effects of early and sustained intervention, motivation, and classroom environments. An additional foundation is research that supports usage of thematic, integrated instruction that incorporates artistic experiences (Northwest Regional Educational Laboratory, 2001).

The core assumption of the DWoK model is that all students, teachers, and school leaders have the capacity to develop expertise in any given subject or skill. All students can achieve proficiency—when schools and classrooms provide appropriate and quality instruction that meets the unique needs of different students and student groups. One key element used by DWoK for the Middle Grades to promote individual growth is varied instructional pathways. As indicated, another is integration of the visual, performing, literary, and media arts, to (a) promote in-depth, creative thinking and content acquisition across all disciplines; and (b) motivate students to think critically and gain deeper, long-lasting understandings. Additional model emphases include:

- standards-based planning in curriculum, assessment, and instruction;
- student inquiry and self-directed learning;
- a comprehensive schoolwide literacy program, including research-based strategies in reading and writing;

1See http://www.differentways.org.
• shared leadership for results; and
• organizational structures that support teaching, learning, and a positive school climate.

Previous research of the DWoK model for the elementary grades has reported positive results. Catterall (1995) followed 1,000 children over 3 years and found that greater exposure to DWoK was associated with higher gains. Students with 1 year of DWoK exposure showed gains of 8 percentage points in language arts achievement, and those with 2 years of exposure gained approximately 16 percentage points. Control students, on the other hand, showed much smaller changes in scores. During these 3 years, DWoK students also demonstrated increased motivation and engagement, whereas control students reported a decline in motivation. Other studies have confirmed these results and attest to the capability of the model as a means of improving student outcomes (Peterson et al., 1998; Petrosko, 1997).

Middle School CSR Reforms and Research

Given the richness of current reform efforts nationwide, it is surprising that there are very few reasonably rigorous studies on using CSR for improving urban students’ academic achievements (Nummery, 1998), particularly at the middle school level (Borman et al., 2003). According to a 2004 RAND study (Juvonen, Le, Kaganoff, Augustine, & Constant, 2004), nearly 9 million students attend public middle schools—schools that serve as an intermediary phase between elementary school and high school, typically consisting of Grades 6–8. The middle school years represent a critical time for young teens (Simmons & Blyth, 1987). Middle schools have been blamed for the increase in student behavior problems and cited as the cause of teens’ alienation, disengagement from school, and low achievement.

In the study entitled Focus on the Wonder Years: Challenges Facing the American Middle School, RAND researchers identified major concerns in middle school education and recommended ways to improve students’ academic achievement, communicate with parents, and support teachers and principals (Juvonen et al., 2004). Although middle schools are not in the state of crisis described in some media accounts, and credible reform efforts are under way, the continuing lackluster performance of many of these schools might be explained, in part, by inadequate implementation of the middle school concept (Felner et al., 1997; Williamson & Johnston, 1999). Core practices such as interdisciplinary team teaching and advisory programs often tend to be weakly implemented with little attention to the underlying goals. A sufficient level of fidelity to many of the reform practices is not possible without substantial additional attention, resources, and long-term support. Based on prior research, a general hypothesis is that whole-school reform, in general, tends to be more successful using an
externally-developed change model than locally-developed (home grown) programs (Nunnery, 1998). As an externally developed CSR model, DWoK therefore seems worthy of investigation as a possible means of increasing the academic success of at-risk middle school students.

Research Context

The context for our study was all DWoK middle schools (n = 3) in the Jefferson County Public Schools district (JCPS) in Louisville, Kentucky. The district is located in a large metropolitan area and has 150 schools serving more than 96,000 students. JCPS educates a high percentage of at-risk urban students with high poverty levels. The district has a student assignment plan based on managed choice, which facilitates the racial desegregation of its schools by providing students with transportation from their home neighborhoods to other parts of the district.

This study is a continuation of a preliminary study of the first 2 years of DWoK implementation in JCPS. Specifically, Ross, McDonald, Goldfeder, and Munoz (2004) analyzed student-level achievement scores for the three DWoK and matched control schools on the Kentucky state assessment system (i.e., Comprehensive Test of Basic Skills, CTBS, and Kentucky Core Content Test, KCCT) and school district measures (i.e., demographics, Stanford Diagnostic Reading Test). Findings showed generally positive trends, included some significant advantages for DWoK, especially in reading, in both years. Effect sizes for statistically significant comparisons ranged from +0.10 to +0.17.

METHOD

A quasi-experimental design, in which each of the three DWoK schools was matched to a similar control school, was employed in our study of 3rd-year DWoK effects. Examining program impacts on norm-referenced and criterion-referenced test scores was the focus of the study and basis for the methodology described below.

Participants and Matching Procedure

The sample included three DWoK middle schools and three matched control schools. The former schools adopted DWoK as part of their effort to increase student achievement, and at the time of this research, had been implementing the model for 3 years (from school year 2001–2002 to school year 2003–2004).

A two-level matching scheme (school- and student-level) was used to increase the internal validity of the study (Cook & Campbell, 1979). The school-level matching involved checking for similarity in terms of poverty (i.e., percentage of
students eligible for free or reduced-price lunch), student mobility, attendance, special education (Exceptional Child Education), and single-parent households. A second basis for the first-level matching was the state’s Accountability Index (AI). The AI is a combination of the academic index and the nonacademic index. First, a weighted academic index is calculated for all content-area tests administered within a school (e.g., reading, mathematics). Then, after adding weighted non-academic measures (e.g., attendance) and a weighted norm-referenced test, the CTBS, the school’s AI for a particular year is determined. The weights used for calculating the AI vary for the different content area tests and for school level (elementary, middle, high), but sum to 1.0. The AI is considered a high-stakes system because rewards and sanctions are attached to results.

The three treatment schools began implementing DWoK in the 2001–2002 school year. Table 1 illustrates the key characteristics by which the treatment and control schools were matched at the baseline year (2000–2001). As shown, the matches were based on six variables, including previous diagnostic test scores in reading, gender, race, and participation in the free or reduced-price lunch program. The comparability of the matches was evaluated using chi-square tests for categorical variables and ANOVAs for continuous variables (e.g., diagnostic test). Table 2 outlines pattern of achievement for the baseline years for both treatment and control schools.

Instrumentation

Only the students who had complete data encompassing the Stanford Diagnostic Reading Test (SDRT), CTBS, and KCCT Reading and Arts and Humanities scores in Grades 6, 7, and 8, served as the basis for the second, student-level matching component (Rossi, Freeman, & Lipsey, 1999). All tests were group-administered. The SDRT test was administered to sixth, seventh, and eighth graders at the beginning of the school year (2003–2004); the CTBS Reading subtest was administered to the sixth graders at the end of the year. At the end of the 2003–2004 school year, sixth graders took the CTBS, seventh graders took the KCCT Reading subtest, and eighth graders took the KCCT Arts and Humanities subtest. A summary of the tests and analyses employed (described in a section to follow) is provided in Table 3.

Design and Procedure

The research design employed a matched treatment-control school pre- and posttest design (Cook & Campbell, 1979; Rossi, Freeman, & Lipsey, 1999). Data at the school level for both treatment and control schools were analyzed using descriptive statistics. However, to control for prior achievement, data at the student level were analyzed via ANCOVA, with the treatment condition as the between-subjects factor; the SDRT pretest scores as the covariate; and sixth-grade
TABLE 1
School-Level Matching Data for Treatment and Control Schools at Base

<table>
<thead>
<tr>
<th>School</th>
<th>N</th>
<th>Poverty (%)</th>
<th>Mobility (%)</th>
<th>Attendance (%)</th>
<th>ECE (%)</th>
<th>Single household (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWoK A</td>
<td>857</td>
<td>47.3</td>
<td>11.3</td>
<td>93.1</td>
<td>14.1</td>
<td>60.6</td>
</tr>
<tr>
<td>Control A</td>
<td>962</td>
<td>43.2</td>
<td>10.6</td>
<td>94.7</td>
<td>11.0</td>
<td>50.7</td>
</tr>
<tr>
<td>DWoK B</td>
<td>915</td>
<td>39.1</td>
<td>10.6</td>
<td>93.2</td>
<td>13.1</td>
<td>50.8</td>
</tr>
<tr>
<td>Control B</td>
<td>625</td>
<td>41.0</td>
<td>13.2</td>
<td>93.1</td>
<td>13.8</td>
<td>53.5</td>
</tr>
<tr>
<td>DWoK C</td>
<td>945</td>
<td>38.9</td>
<td>12.7</td>
<td>94.4</td>
<td>13.8</td>
<td>54.5</td>
</tr>
<tr>
<td>Control C</td>
<td>831</td>
<td>43.9</td>
<td>12.0</td>
<td>92.6</td>
<td>14.0</td>
<td>53.5</td>
</tr>
</tbody>
</table>

*Note.* Accountability Index includes academic (e.g., reading, math, science, and social studies) and non-academic = Different Ways of Knowing; ECE = Exceptional Child Education.
TABLE 2
School-level Prior Patterns of Achievement for Treatment and Control Schools

<table>
<thead>
<tr>
<th>School</th>
<th>Accountability Index</th>
<th>Average Accountability Index (1999–2001)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1999</td>
<td>2000</td>
</tr>
<tr>
<td>DWoK A</td>
<td>62.9</td>
<td>65.5</td>
</tr>
<tr>
<td>Control A</td>
<td>59.7</td>
<td>62.5</td>
</tr>
<tr>
<td>DwoK B</td>
<td>58.7</td>
<td>58.0</td>
</tr>
<tr>
<td>Control B</td>
<td>55.8</td>
<td>56.3</td>
</tr>
<tr>
<td>DwoK C</td>
<td>60.2</td>
<td>56.3</td>
</tr>
<tr>
<td>Control C</td>
<td>47.7</td>
<td>48.8</td>
</tr>
</tbody>
</table>

Note. Accountability Index includes academic (e.g., reading, math, science, and social studies) and non-academic (e.g., attendance, retention) indicators. DWoK = Different Ways of Knowing.

TABLE 3
Summary of Analyses Comparing DWoK to Control Schools on Student Achievement

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Year</th>
<th>Analysis</th>
<th>Covariate(s) &lt;TH&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sixth-grade CTBS</td>
<td>2003–2004</td>
<td>MANCOVA</td>
<td>Previous achievement (beginning of sixth-grade SDRT)</td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td>and gender</td>
</tr>
<tr>
<td>Language Arts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seventh-grade KCCT</td>
<td>2003–2004</td>
<td>ANCOVA</td>
<td>Previous achievement (beginning of seventh-grade SDRT)</td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eighth-grade KCCT</td>
<td>2003–2004</td>
<td>ANCOVA</td>
<td>Previous achievement (beginning of eighth-grade SDRT)</td>
</tr>
<tr>
<td>Humanities and Arts</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. KCCT = Kentucky Core Content Test; CTBS = Comprehensive Test of Basic Skills; SDRT = Stanford Diagnostic Reading Test; DWoK = Different Ways of Knowing.

CTBS Reading, Language Arts, and Math scores, seventh-grade KCCT Reading scores, and eighth-grade Arts and Humanities scores as the dependent variables (see Table 3).

Results

Sixth-Grade Analysis Using CTBS Subtests

Analyses of 2003–2004 CTBS subtests compared DWoK and control schools on the sixth-grade subtests of Reading, Language Arts, and Mathematics. A
### TABLE 4
Unadjusted and Adjusted Means and Standard Deviations for DWoK and Matched-Control Comparisons on CTBS 6th-Grade Scale Scores in 2003–2004 (Year 3)

<table>
<thead>
<tr>
<th>Comparison Group and Subtest</th>
<th>M</th>
<th>M_{adj}</th>
<th>SD</th>
<th>ES*</th>
<th>ES*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWoK</td>
<td>652.29</td>
<td>649.93</td>
<td>43.21</td>
<td>+0.20</td>
<td>+0.09</td>
</tr>
<tr>
<td>Control</td>
<td>643.50</td>
<td>646.02</td>
<td>43.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWoK</td>
<td>653.28</td>
<td>650.95</td>
<td>44.77</td>
<td>+0.16</td>
<td>+0.04</td>
</tr>
<tr>
<td>Control</td>
<td>646.74</td>
<td>649.23</td>
<td>38.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWoK</td>
<td>651.06</td>
<td>648.71</td>
<td>48.97</td>
<td>+0.09</td>
<td>-0.02</td>
</tr>
<tr>
<td>Control</td>
<td>646.92</td>
<td>649.44</td>
<td>46.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Effect sizes were computed from the unadjusted means. *ES = Effect size.*

**Note.** DWoK n = 926; Control n = 867. DWoK = Different Ways of Knowing.

MANCOVA was employed in which covariates were prior achievement ($t = 9.78, p = .002$) and gender ($t = 5.37, p = .02$), both of which were found to differ significantly between program and control groups.

Descriptive unadjusted statistics for each group on the three CTBS subtests are summarized in Table 4. As can be seen, the DWoK mean was directionally higher than the control group mean on all three tests. The MANCOVA (Wilks' Lambda) yielded a significant overall program effect, $F(3, 1787) = 2.93, p < .05$. Follow-up univariate tests were significant in Language Arts, $F(1, 1789) = 7.29, p = .007$, but not in Reading, $F(1, 1789) = 1.67, p = .20$, or Mathematics, $F(1, 1789) = .20, p = .65$. As shown in Table 4, the effect sizes associated with adjusted means were positive for reading and language arts but small in magnitude (ranging from +0.04 to +0.09). Those associated with unadjusted means were slightly higher, ranging from +0.09 to +0.20.

**Seventh-Grade Analysis Using KCCT Reading Test Scores**

Initial analyses, conducted to verify the similarity of the program groups, failed to show significant differences on any of the key demographic variables examined: race ($\chi = 6.98, p = .14$), free or reduced-price lunch status ($\chi = 1.01, p = .60$), and gender ($\chi = .71, p = .40$). To determine group equivalence on prior achievement, we conducted an ANOVA comparing the 2003 SDRT means for the DWoK and control groups. Results favored DWoK ($F = .77, p = .38$) directionally but not significantly.
TABLE 5
Unadjusted and Adjusted Means and Standard Deviations for DWoK and Matched-Control 7th-Grade and 8th-Grade Comparisons on KCCT Scale Scores in 2003–2004 (Year 3)

<table>
<thead>
<tr>
<th>Comparison Group and Test</th>
<th>N</th>
<th>M</th>
<th>M\textsubscript{adj}</th>
<th>SD</th>
<th>ES\textsuperscript{a}</th>
<th>ES\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003–2004 KCCT Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWoK</td>
<td>892</td>
<td>516.74</td>
<td>516.80</td>
<td>37.46</td>
<td>+0.28</td>
<td>+0.28</td>
</tr>
<tr>
<td>Control</td>
<td>815</td>
<td>506.98</td>
<td>506.91</td>
<td>32.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003–2004 KCCT AH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWoK</td>
<td>889</td>
<td>533.51</td>
<td>533.44</td>
<td>83.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>828</td>
<td>497.27</td>
<td>497.34</td>
<td>60.48</td>
<td>+0.48</td>
<td>+0.48</td>
</tr>
</tbody>
</table>

*Note.* DWoK = Different Ways of Knowing; KCCT = Kentucky Core Content Test; AH = Arts and Humanities KCCT subtest.

\*Effect sizes were computed from the unadjusted means. \*\*Effect sizes were computed from the adjusted means.

The ANCOVA performed on KCCT Reading yielded significance for both the SDRT covariate, $F(1, 1704) = 18.24, p < .001$, and most importantly, the Program effect, $F(1, 1704) = 34.32, p < .001$, ES = +0.28.

---

Eighth-Grade Analysis Using KCCT Arts and Humanities Test

As in the preceding analyses of Reading outcomes, an ANCOVA using the baseline (2003) SDRT scores as a covariate, was conducted on the KCCT Arts and Humanities (AH) subtest scores. Results indicated that the SDRT covariate was significant, $F(1, 1714) = 13.51, p < .001$, and the Program effect was highly significant, $F(1, 1714) = 104.47, p < .001$, ES = +0.48, favoring DWoK over control. Table 5 displays a summary of the KCCT (a) Reading and (b) Arts and Humanities student achievement analyses.

---

**SUMMARY**

The Year 3 results were generally supportive of DWoK effects on academic achievement. On the CTBS Language Arts subtest in the sixth grade, the KCCT Reading subtest in the seventh grade, and KCCT Arts and Humanities subtest in the eighth grade, DWoK schools significantly surpassed the matched control schools. No statistically significant differences were observed in the Arts and Humanities subtest.

---

**Student Achievement**

The analytical results of the Program (D\*W) treatment effect on student achievement outcomes were mixed. While the program showed promise in improving both Reading and Arts and Humanities scores, further investigation is needed to determine the long-term impact of the DWoK intervention at these schools.

---

The results of the Achievement Program Evaluation (APE) study were revealed at the annual conference on educational innovation. The findings suggested that the DWoK approach could potentially improve student engagement and academic performance, but further research is needed to confirm these preliminary results.
TABLE 6
Means and Standard Deviations for DWoK and Matched-Control Comparisons on Attendance Rates

<table>
<thead>
<tr>
<th>Comparison Group and DWoK</th>
<th>M(%)</th>
<th>SD</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002–2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWoK</td>
<td>95.10</td>
<td>5.21</td>
<td>9.31</td>
<td>.002</td>
</tr>
<tr>
<td>Control</td>
<td>94.50</td>
<td>5.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003–2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWoK</td>
<td>94.48</td>
<td>6.33</td>
<td>1.06</td>
<td>.303</td>
</tr>
<tr>
<td>Control</td>
<td>93.85</td>
<td>6.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. DWoK n = 1649; Control n = 1505. DWoK = Different Ways of Knowing.*

Schools. No statistically significant differences, however, were found on the sixth-grade CTBS Reading and Mathematics subtests.

Student Attendance

The analytical design employed on attendance data was an ANCOVA, consisting of Program (DWoK vs. control) as the between-subjects factor with previous year attendance rates as the covariate. The attendance rate was computed using the following formula:

\[
\left(\frac{m-a}{m}\right) \times 100
\]

In the formula, school membership \( (m) \) for the school year minus attendance \( (a) \) for the school year is divided by membership \( (m) \) for the school year. The resultant dividend is then multiplied by 100 to obtain the attendance rate expressed in percentages.

Analyses were performed comparing the DWoK and control schools on attendance in the year 2002–2003 and the Year 2003–2004. Means, standard deviations, \( F \), and \( p \) values for program and control schools by year are presented in Table 6. Descriptive outcomes indicated that attendance rates at DWoK schools were higher than the control schools by about 0.60 of a percentage point each year. However, although the difference was statistically significant in Year 2 \( (p = .002) \), it did not reach significance in Year 3 \( (p = .303) \).

DISCUSSION

The results of this study showed generally favorable third-year achievement outcomes for the DWoK for the Middle Grades CSR model. It is important that, con-
sistent with a recent review by RAND of whole-school reforms (Juvonen et al., 2004), the potential value of interventions that address both academic achievement and developmental needs of young teens is suggested. Unlike reform models that incorporate prescribed curricula or pedagogy, such as Success For All (Slavin & Madden, 2001) or Direct Instruction (Engleman, Becker, Carne, & Gersten, 1988), DWoK attempts to engage students, via teacher-designed lessons, on creative, learner-centered activities that integrate curricula, especially with the arts. Expected outcomes include high engagement, increased achievement, meaningful learning, and connection of skills to cultural and real-world events. However, given the narrow curricular focus and intensity of high-stakes testing environments (Linn, 2003; Shepard, 2000), as well as the greater time needed for implementation of process-oriented (as opposed to prescription-based) models (Datnow, Hubbard, & Melan, 2002), a concern might be whether student achievement on state-mandated tests could be noticeably impacted after only 3 years (see Desimone, 2002; Fullan, 2000; Levin, 1993).

Given this rationale, it is noteworthy that the three DWoK schools demonstrated a significant overall advantage over their matched control schools on sixth-grade CTBS language arts, seventh-grade KCCT Reading, and eighth-grade KCCT Arts and Humanities. Effect sizes with adjusted means were small for CTBS sixth grade (range = +0.02 to +0.09); those associated with unadjusted means were slightly higher, ranging from +0.09 to +0.20. Notably, the effect sizes for KCCT Reading in seventh grade and KCCT Arts and Humanities in eighth grade (+0.28 and +0.48) were fairly large.

To understand the magnitude of the effect sizes, we followed the comprehensive approach to effect size interpretation that utilizes multiple criteria and benchmarks (Cooper, 1981). First, and most generally, we compared the overall CSR effect size to Cohen's (1988) definitions of a small effect (d = 0.20) and a large effect (d = 0.80) within the behavioral sciences. Cohen pointed out that the relatively small effects of around +0.20 were most representative of fields that are closely aligned with education, such as personality, social, and clinical psychology. Similarly, Lipsey and Wilson (1993) concluded in a compendium of meta-analyses that psychological, educational, and behavioral treatment effects of modest values of even $d = 0.10$ to $d = 0.20$ should not be interpreted as trivial. More specifically, the DWoK program effect sizes exceeded the levels associated with the more successful CSR models identified in the research literature (Borman et al., 2003). Perhaps these differential grade-level patterns reflect the cumulative benefits of DWoK implementation over seventh graders 2-year school enrollment.

Why and how student learning improved is less clear based on these data. Interviews with DWoK coaches, conducted in a separate study of the same schools (Ross, Wilson-Releyea, et al., 2004), suggested that DWoK positively impacted student sharing, engagement, and enthusiasm. Of specific interest in that study were DWoK's design components, effectiveness, implementation quality and progress,
COMPREHENSIVE SCHOOL REFORM IN MIDDLE SCHOOLS

and perceived impact. Positive responses were indicated in most of these areas. Overall, the coaches’ perceptions pointed to the importance of understanding program implementation fidelity and promoting long-term teacher empowerment for change in all externally developed comprehensive school reform models (see Datnow & Castellano, 2000a, 2000b).

The three DWoK participating schools were the only ones using the model in the district. In a prior qualitative study (Ross, Wilson-Relyea, et al., 2004), interviews regarding DWoK Middle Grades implementation in Kentucky (encompassing two rural and the present three urban schools) were conducted with two artist-educators and two DWoK Instructional Leaders who worked closely with all five schools. Due to their confidentiality agreements with the schools, they were willing to discuss their perceptions about implementation quality only globally or generically (e.g., “Three schools were strong on component X, whereas all five were moderate to strong on component Y, etc.”). Important to this study, therefore, was the consensual global evaluation that, although the five schools differed in implementation quality, all had adequate implementations to represent the middle grades model in an evaluation study. There was also agreement that DWoK implementation was stronger overall for the rural than urban sample. However, several respondents who had previously worked with the DWoK elementary school model in varied contexts described this disparity and the level of implementations as typical of what would occur in other locations.

The qualitative study further suggested positive DWoK effects on teacher instruction, planning, sharing, and involvement (Ross, Wilson-Relyea, et al., 2004). Strongest among these was the confidence that the teachers placed in their own ability to provide organized, meaningful, and individually adaptive instruction. Weakest was their confidence in the school administration to provide them with the time, resources, and expertise to improve their own performance and that of the school in general. Assuming these impressions to be valid, the combination of student enthusiasm for learning and high teacher involvement would certainly be predictive of observable achievement gains.

However, as shown in prior research with New American School models, specifically in Memphis (Ross, 2001) and San Antonio (Berends, Chun, Schuyler, Stockley, & Briggs, 2002), early success may be realized as part of extraordinary commitment to and expectations for reform. A key problem in implementing and institutionalizing reform that seems to have plagued most prior systemic reform efforts is demonstrating sustainability over time (Desimone, 2002; Rowan et al., 2004). Subsequent research is therefore needed to follow up these early, promising results with the DWoK middle grades model. To extend our study and answer some important questions that still remain, such studies should involve rigorous examination of model implementation and process outcomes via triangulated data sources reflecting participant/stakeholder perceptions, activities, and classroom practices. Unique to DWoK, compared to other widely used CSR models, is its em-
phasis on integrating the arts with the core curriculum, thus ostensibly making learning more meaningful and engaging for students. Although there is a popular belief that arts education enhances student achievement in reading, writing, and mathematics (Paige & Huckabee, 2005), supportive evidence from rigorous experimental research is still lacking. Future studies on DWoK for the Middle Grades should address this need. Additional, broader issues of research interest, as suggested by our study and the Ross et al. (2004) research on DWoK, include degree of CSR ownership among veteran and novice teachers, levels of adaptation at the classroom level, and the prescriptiveness of the CSR model.

Although clearly encouraging, our findings are qualified by several factors. First, due to the inability to employ a randomized experimental design, firm causal conclusions cannot be reached. However, our systematic matching procedure and statistical controls should have established sufficiently strong internal validity for attributing differential group change more to the interventions than to potentially confounding intervening factors (Cook & Campbell, 1979). A second limiting factor was the limited number of schools involved in the study. Although the matching process used may ensure internal validity, so few schools limits the external validity and hence the policy relevance of the conclusion drawn. A third limiting element was the shrinkage of sample sizes for both the DWoK and control groups due to incomplete testing data and student mobility. As a consequence, the representativeness of the sample (i.e., threat to external validity) might have been reduced.

A final limitation is associated with a line of inquiry referred to as effective schools research (for a review, see Teddlie & Reynolds, 2000). Effective school researchers have shown that some schools have poor leadership and make little progress (Brookover, Beady, Flook, Schweitzer, & Wisenbaker, 1979; Edmonds, 1979); in those scenarios, educational leaders do not major interventions like DWoK. Because the faculty is in disarray, leadership will not embark on school reform challenges. Predictably, the school taking on a reform enthusiastically over a couple of years will do better than the comparison school. The specifics of programs like DWoK account for some, but not all, of the differences found in the statistical analyses. Other factors, such as disparity in leadership and school climate, might account for some of the differentiation between treatment and control schools.

Whole school reform has thus far proven more challenging and less successful at the middle school than at the elementary school level (e.g., Desimone, 2002). To facilitate positive change, one influential element appears to be strong coordination between the policies and activities of district and state organizations and foundations (Juvonen et al., 2004). A second element is having a systematic, research-based framework to motivate and guide curricular and instructional improvements in middle school classrooms. Our results suggest that arts integration in general, and DWoK in particular, have potential to foster effective classroom changes and are deserving of further study.
REFERENCES


