The purpose of this investigation was to examine the relationship between participation in contrasting school music programs and standardized test scores. Relationships between elementary (third- or fourth-grade) students' academic achievement at comparable schools, but with contrasting music programs as to instructional quality, were investigated. Relationships also were examined between middle school (eighth- or ninth-grade) students' academic achievement and their participation in school music programs that also differed in quality. Participants (N = 4,739) were students in elementary (n = 1,119) and middle schools (n = 3,620) from the South, East Coast, Midwest, and West Coast of the United States. All scores were standardized for comparison purposes. Analysis of elementary school data indicated that students in exemplary music education programs scored higher on both English and mathematics standardized tests than their counterparts who did not have this high-quality instruction; however, the effect sizes were slight. Analysis of middle school data indicated that for both English and math, students in both exceptional music programs and deficient instrumental programs scored better than those in no music classes or deficient choral programs; however, the effect sizes were not large.

Christopher M. Johnson and Jenny E. Memmott
University of Kansas

Examination of Relationships between Participation in School Music Programs of Differing Quality and Standardized Test Results

The extant literature is replete with investigations examining the effects of music study on the academic success of students. Researchers have paired music participation with various academic outcomes, including math and reading skills, as well as overall grade-
point averages. Though there are some investigations that look at elementary students and academic progress, those are far fewer.

Prior studies of elementary students have typically focused on those in third and/or fourth grade. In one such study, Kemmerer (2003) found that the number of hours spent in a general music class had no effect on reading-skill scores. However, closer examination showed that the difference of time actually spent in music instruction between the groups was less than 18 minutes per week, so the case for different amounts of instructional time did not seem strong enough to be consequential. Gregory (1988) found that third-grade students receiving music instruction through the "Leap into Music" curriculum made significant academic progress in mathematics. Another longitudinal study by Costa-Giomi (1999) showed that private piano lessons increased several measures of intelligence in the short term. However, those gains, as well as any academic gains, were not maintained through the entire 3-year span of the study. By the end, both the experimental and control groups were relatively equivalent on both measures—academic and intelligence scores.

Studies investigating connections between participation in music and general academic achievement have been ubiquitous in the literature. Most have demonstrated that participation in music parallels increased academic achievement (Perry, 1993). This relationship has been demonstrated with standardized tests in reading (Butzlaff, 2000; Neuharth, 2000), mathematics (Neuharth, 2000; Whitehead, 2001), grade-point averages (Miranda, 2001; Zanutto, 1997), SAT scores (Butzlaff, 2000; Cobb, 1997; Miranda, 2001), and ACT scores (Cobb, 1997; Miranda, 2001). Some of the aforementioned studies have found that academic achievement did not improve with music participation. Others have shown that music participation did not affect academic achievement more than did the other variables investigated, but significant academic gains were still noted (Andrews, 1997; Perry, 1993). None of the studies found that participation in music negatively influenced academic progress.

Perhaps the three most compelling reports on the relationship between music participation and academic achievement are by Cobb (1997), Catterall, Chapleau, and Iwanaga (1999), and Butzlaff (2000). Cobb (1997) examined the ACT registration forms of 17,099 test-takers and compared those who indicated that they had two or more classes or activities in music to those who had not. Findings indicated that individuals with a musical background had significantly higher ACT scores on the English, reading, and science subtests. Scores for math were also higher for all subgroups.

Catterall, Chapleau, and Iwanaga's 1999 investigation tracked approximately 25,000 students over the course of 10 years. Results indicated that, regardless of socioeconomic background, secondary school students involved in music had significantly higher standardized test scores—specifically mathematics proficiency—than students not involved in music. This study examined several standardized tests, including the SAT. Similarly, Butzlaff (2000) completed a meta-analysis of all studies wherein a reading standardized test followed music instruction. He documented a consistent correlation between reading ability and music instruction.

Though many claim that involvement in school activities in general aids academic progress, several studies have shown that not to be the case. In four investigations, music participation was the only activity shown to correlate significantly with academic progress (Miranda, 2001; Schneider, 2000; Trent, 1996; Underwood, 2000). Neither athletics nor any other extracurricular activity showed similar results.

The abundance of available literature demonstrates that the linkage between music participation and academic success has been an area of interest to many researchers. An overview of their work indicates a strong and reliable association between the study of music and performance on standardized reading and mathematics tests. Evidence exists that supports the concept of a two-way interactionist position, such that music might catalyze or deepen learning in other academic areas. None of the investigators whose work is included in this review have considered the quality of the music education received by the participants as a variable in their research.

Because of the mixed results of previous investigations and the lack of research using instructional quality as an independent variable, this study was conceived as an attempt to mitigate this identified gap in the literature. Therefore, the purpose of this investigation was to examine any relationships between participating in high- or low-quality school music programs and standardized test results in a variety of geographical locations across the United States. One part of the study investigated the relationship between third- or fourth-grade students' academic achievement at schools that were similar in size, socioeconomic status, and other factors, but had music programs characterized as exemplary or of deficient quality. A second part of the study examined the relationship between eighth- or ninth-grade students' standardized test scores and their participation in school music programs recognized as either high-quality or low-quality. Examination of these secondary school scores also included an analysis of whether these students were involved in instrumental or choral music, or if they had no involvement in the music programs at their respective schools.

METHOD

Participants

Test scores from the 2004–2005 school year were analyzed from 4,799 third-, fourth-, eighth-, and ninth-grade students from five states representative of regions in the South, East Coast, Midwest, and West Coast of the United States. The total number of elementary student participants, in either third or fourth grade, was 1,119. The total number of middle school students, in either eighth or ninth grade, who participated in the study was 3,620. From the South (n = 1,761), test scores were examined from participants in two elementary
schools \((n = 352)\) and three middle schools \((n = 1,409)\). For the West Coast \((n = 424)\), test scores were analyzed from two elementary schools \((n = 84)\) and two middle schools \((n = 340)\). Test scores from the East Coast \((n = 1,085)\) included data from two elementary schools \((n = 410)\) and two middle schools \((n = 672)\). From the Midwest \((n = 1,403)\), student test scores from two elementary schools \((n = 275)\) and four middle schools \((n = 1,199)\) were used.

Participating elementary and middle schools were identified by university music education professors familiar with the school districts in each geographic region, and, consequently, familiar with the caliber of the music education programs at each school. In each region, two elementary schools and two middle schools were initially identified—one elementary school and one middle school with exemplary music education programs, and one elementary school and one middle school considered to offer less-than-optimal music education. However, in several middle school situations, schools were identified as having excellent instrumental music but substandard choral music. In those cases, scores from other schools where the choral program (and not necessarily the instrumental program) was optimal were obtained.

The music education professors selecting the schools for participation were all published and accomplished researchers. Selection criteria given to the professors indicated that schools chosen were to be as similar as possible in every regard except in the quality of their music programs. If possible, the schools should be from a single district, should be equal in socioeconomic status, and be similar in all other environmental factors. The differences in the music programs were determined by guidelines established using the National Standards of MENC: The National Association for Music Education.

The schools were to have been on opposite ends of the curve when the faculty questioned chose where to place their student teachers. All decisions were confirmed with an onsite visit and observation. The schools chosen either agreed or declined to participate. When schools opted out of participation, other schools were chosen using the same original criteria, and these alternate schools were pursued until data were obtained. School district and/or building administrators were provided a description of the project that explained the selection process, including the school-selection criteria, as approved by the authors' Institutional Review Board. No decision-making personnel inquired about which category any music programs represented.

All students in third- or fourth-grade or eighth- or ninth-grade who took the standardized test specific to their school district/state were included in the study. Permission to use student test scores and additional data was obtained through contact with individual schools and school districts before the start of the study. All participant data were completely anonymous; therefore, gender, age, socioeconomic background, and other identifying characteristics were not factored into the results of this study.

Materials

The standardized tests analyzed in this investigation were those administered to meet state assessment requirements stipulated by No Child Left Behind (NCLB) legislation. This resulted in a total of 6 different English and 5 different mathematics tests at each level. Though one can argue that differences exist between any set of different tests, NCLB has resulted in standardization of K–12 assessment to a considerable degree. Printed or electronic copies of the test results for each school were obtained for use in this study. Some school systems required substantial paperwork in order to release scores; other schools faxed data as a result of telephone conversations. All school systems were strongly encouraged not to include any personal or identifying data for any students with their scores. While some schools provided sets of scores matched by students, such that an English score could be paired with a mathematics score, other schools provided completely separate lists of each set of scores, making such matching impossible.

Design

The study's design for the elementary level used a sample containing two independent groups of students. All children received general music instruction in their schools. Group 1 included those students who took their third- or fourth-grade state-mandated assessments at a school where the quality of the music instruction was deemed exemplary by knowledgeable music education faculty. Conversely, Group 2 was composed of elementary students who took their standardized tests at a school where the quality of the music instruction was considered less than ideal.

The design for the middle school level required more variables. Although the dependent measures were the same, each middle school student was coded as an instrumental music student, choral music student, or nonmusic participant for the 2004–2005 school year. Like their elementary counterparts, middle school students involved in music were coded based on whether they received music instruction of exemplary or deficient quality.

Procedure

Data collection for the elementary test scores consisted of entering the anonymous academic scores into a database. For the middle school data, in addition to entering each middle school student's standardized test scores, each student's music participation, as defined by enrollment in a music class during the 2004–2005 school year, was recorded.

RESULTS

This study was designed to investigate relationships between elementary and middle school students' academic achievement, as mea-
measured by scores on state-mandated standardized tests of English and mathematics designed to meet the requirements of the No Child Left Behind legislation and by the instructional quality of their school music programs. To compare the different tests adopted by the different states, it was necessary to convert them to a comparable scale. Thus, all scores were standardized (transformed mathematically) into z-scores.1

Analysis of the data from the elementary school part of the study examined each dependent measure separately. Because many of the schools reported their student English and mathematics data separately, there was no way to match them for a more complex analysis. Therefore, scores were examined using univariate analysis of variance (ANOVA) procedures, with two independent variables, geographic region and quality of music instruction. Because of the large N, the alpha level was set at .01 for all statistical tests.

Results for the dependent measure English indicated that there were significant differences for region \( F(3, 945) = 7.40, p < .001, \) partial \( \eta^2 = .023 \), instruction quality \( F(1, 945) = 34.22, p < .001, \) partial \( \eta^2 = .035 \), and the two-way interaction \( F(3, 945) = 6.12, p < .001, \) partial \( \eta^2 = .019 \). Because of the significant interaction, the data were graphed (see Figure 1). What is first noticeable in this graph is that the schools with the high-quality music programs on the West Coast had English scores lower than those of the schools whose music programs were deemed inferior. However, the three remaining comparisons indicate a difference approaching 24%.2

Similar results were found for the mathematics scores, again using univariate ANOVA procedures with independent variables of region and differing quality of music instruction. Results indicated that there was no significant difference by region \( F(3, 1022) = 1.63, p = .180, \) partial \( \eta^2 = .005 \), but significant differences were noted for instructional quality \( F(1, 1022) = 30.49, p < .001, \) partial \( \eta^2 = .029 \). In spite of this difference, a significant two-way interaction was not found \( F(3, 1022) = 2.87, p = .035, \) partial \( \eta^2 = .008 \). Means are illustrated in Figure 2. In this graph, all four of the elementary schools with high-quality music programs scored better than those whose programs were considered to be of lower quality. This difference is exaggerated in the South and East regions, where the differences exceed a half standard deviation.

The magnitude of the differences on both of these dependent measures can be seen in the overall means for the two types of music instruction (see Table 1). These differences, represented by z-scores, indicate a magnitude of difference equal to 22% better English scores and 20% better mathematics scores for the excellent music programs.

Analysis of the middle school data involved many of the same methods as those used for the elementary data. English scores were examined using univariate ANOVA procedures with two independent variables, region and type/quality of music instruction. The latter was categorized into 5 levels: exemplary instrumental, exemplary choral, deficient instrumental, deficient choral, and no music.
Table 1  
Overall Mean z Scores for Elementary School Data

<table>
<thead>
<tr>
<th>Group</th>
<th>English Scores</th>
<th>Mathematics Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent Music</td>
<td>.59</td>
<td>.68</td>
</tr>
<tr>
<td>Deficient Music</td>
<td>.01</td>
<td>.14</td>
</tr>
</tbody>
</table>

Results indicated that there were significant differences for region \(F(3, 3142) = 46.40, p < .001\), partial \(\eta^2 = .042\), type/quality \(F(4, 3142) = 52.30, p < .001\), partial \(\eta^2 = .062\), and the two-way interaction \(F(12, 3142) = 3.65, p < .001\), partial \(\eta^2 = .014\). Because of the significant interaction, the data were graphed (see Figure 3). These data illustrate that students at schools with excellent music programs generally performed better on standardized tests than students at schools with lower-quality music offerings. Students at schools with poorer instrumental programs outscored the students who had no music at all, and the students who participated in poor choral programs scored the worst in every region.

Similar results were found for the mathematics scores, using univariate ANOVA procedures with the independent variables of region and type/quality of music instruction. Results indicated that there were significant differences for region \(F(3, 2954) = 115.35, p < .001\), partial \(\eta^2 = .150\), type/quality \(F(4, 2954) = 48.19, p < .001\), partial \(\eta^2 = .061\), and the two-way interaction \(F(12, 2954) = 10.13, p < .001\), partial \(\eta^2 = .040\). The graphed data (see Figure 4) illustrate that students in schools with excellent music programs in two of the four regions performed better on standardized tests than students in schools with poorer music programs. The West Coast and Midwest data, however, are a notable exception to the data trends associated with good instructional programs. Interestingly, and across the board, the students in schools with poorer instrumental programs outscored the students who had no music at all, and the students who participated in poor choral programs generally scored the lowest.

The magnitude of the differences on both of these dependent measures can be seen in Table 2, which provides the overall means for the four types of music instruction and no music instruction. The results indicated three approximate groupings: (1) excellent instrumental and choral programs grouped with deficient instrumental programs, (2) students not participating in any school music, and (3) deficient choral programs. The magnitude of the differences with regard to English was as much as 19% from the top grouping to the nonmusic participants group, and an additional 13% of the popula-

Figure 3. English Test Results for Middle School Students by Region.

Figure 4. Mathematics Test Results for Middle School Students by Region.
Table 2
Overall Mean z Scores for Middle School Data

<table>
<thead>
<tr>
<th>Group</th>
<th>English Scores</th>
<th>Mathematics Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent Instrumental</td>
<td>.90</td>
<td>1.02</td>
</tr>
<tr>
<td>Excellent Choral</td>
<td>.81</td>
<td>.70</td>
</tr>
<tr>
<td>Deficient Instrumental</td>
<td>.69</td>
<td>.80</td>
</tr>
<tr>
<td>No Music</td>
<td>.33</td>
<td>.44</td>
</tr>
<tr>
<td>Deficient Choral</td>
<td>-.02</td>
<td>.03</td>
</tr>
</tbody>
</table>

**DISCUSSION**

Several aspects of this study should be discussed to help with the interpretation of the differences found. First is the sampling of the schools included in this study. As described in the methods section, schools were chosen by local music professors very familiar with both the schools in their area and appropriate research methodology. This was done to ensure that the focus of school selection was on the quality of their music programs, rather than other possible means of assessing differing schools and school systems.

One question that should be addressed is whether other variables (particularly socioeconomic status) might have confounded the results. The possibility of other confounding variables factoring into these results is much more likely to have occurred with the elementary schools than with the middle schools. Since each elementary school was treated as an intact group, all of the students in each school were considered participants in either an excellent or a deficient music program. However, that was not the case with the middle school students. All the middle school participants either participated in some type of music or had no music at all. Furthermore, many of the middle schools had excellent band programs and deficient choral programs or vice versa. Thus, it was not possible to consider “school” as a unit of analysis. When these factors are considered, and the results at the two levels are compared, it would seem that whatever other differences may have been present between the elementary schools were minimal.

This finding parallels those of Catterall, Chapleau, and Iwanaga (1999), who reported that, regardless of socioeconomic background, students involved in music had significantly higher standardized test scores than students not involved in music. Coupled with the integrity of the professors who assisted with participant selection, school differences, though always present and very real, seem to have played only a small part in the differences found in this investigation.

Another aspect of this investigation that should be addressed is the treatment of the data. The data collected were rather simple. Though more data could have been collected for any given student, an increase in the depth of the data would have greatly decreased the number of subjects that could have been obtained. One other confounding factor of this study was the vast geographical representation. By deciding to strive for four widely diverse regions, we encountered different tests as dependent measures. Though the tests themselves were, in all probability, measuring the same student skills and abilities, they used different measures to do so. These differing measures required that all of the test results be standardized to z-scores. It must be acknowledged that any standardization of scores necessitates an arithmetic calculation, which by definition contains some potential for error and reduces some precision. We believe, however, that the benefits of the broad sample outweigh the potential deficits resulting from the lack of a common achievement test.

All of the aforementioned having been acknowledged, it is evident, based on the results of this study, that rather large differences exist between the test scores of the students in the excellent and deficient music programs. Though the statistically significant differences can be credited in part to the large number of participants involved in the study, these differences are difficult to ignore. The probability that any of these differences are due to chance alone is less than one in 1,000 for each of the four statistical comparisons. If one is to accept the methodology used in the selection of the schools and the validity of the standardized tests the states administered, then these results seem noteworthy.

Conversely, the effect size associated with these differences is somewhat less impressive. For the elementary participants, the effect size was approximately 3% of the total differences noted for the middle school students, effect size was slightly higher, at approximately 6%. Taken independently, this is a very small difference, and somewhat negates the probabilities noted above. However, when considered in conjunction with previous research investigating the causes of academic success, the effect size becomes appreciably more consequential. Previous research has indicated that 75% of the variance in academic success can be attributed to family variables, such as years of parental education, family income, and the presence or lack of a mother in the household, as well as individual characteristics such as race (Goldhaber & Brewer, 1997). This leaves 25% of the variance in academic success unaccounted for. If the quality of music instruction accounts for 3–6% of that 25%, then this effect size is much more noteworthy. In this context, the magnitude figures cited in Tables 1 and 2 seem to be more consequential.

It would be naive to claim that high-quality music experiences caused these differences. First, this study examined relationships, not causality. Furthermore, if differing experiences with music do result
in such differences, then surely that would have been evident long before this study. However, there does seem to be what could be considered an important relationship.

One facet of these results that may be somewhat disturbing for music educators would be the results of the middle school students who participated in deficient choral programs. As noted, those students did poorest on standardized tests. However, after careful examination of the curricula of the schools in this study, we speculate that some of these choral music programs might be used by school administrators as academic "holding cells," serving as low-cost scheduling or curricular "fixes," perhaps in lieu of study halls, into which students are placed regardless of interest or musical aptitude.

While there may be many potential explanations for the overall results of this study, three seem most likely. First, it is possible that schools that are diligent in hiring excellent music teachers might also be diligent in hiring excellent teachers across the board. A second possible explanation might be that excellent music programs disproportionately attract academically gifted students. While other students will also inevitably participate in music programs, these more average students might show measurable benefits from being in an environment with those gifted students. A third explanation for these results is that many organizational skills and learning strategies that are generally and naturally present in high-quality and more challenging music programs can aid students in the acquisition of knowledge in other subjects. The skills of thinking in an organized way and developing an ability to strategize and process logically very well may transfer from one area to another.

Finally, it is crucial to note that this project has revealed a relationship between quality of music instruction and academic performance. This finding agrees with previous research showing that music supports academic achievement (Butzlaff, 2000; Cobb, 1997; Miranda, 2001; Neuharth, 2000; Perry, 1995; Whitehead, 2001; Zanutto, 1997). However, since student participation in other activities was not examined in the form of variables in this study, there is no basis for comparing the effects of participation in music with various other activities that could also potentially affect academic achievement (Schneider, 2000; Trent, 1996; Underwood, 2000). Although the relationship between the quality of music education and academic performance appears to be strong, there is nothing in this study that should imply causation. There is no evidence to suggest, for example, that if one were to participate in a good band program, one’s standardized test scores in English and mathematics would improve. Nor should one presuppose that if a school’s chorus program improved, standardized test scores in that school would rise.

The statistically strong relationship shown in this study is simply that—a strong relationship. However, it must not be forgotten that the purpose of a fine music education is not to improve English test scores, and one should no more study music to improve English scores than one should study English to improve music scores. The reason to study music is to enhance the quality of one’s life through the myriad opportunities and experiences that music study provides.

The results of this study warrant further investigation. While replicating this study using students from different schools would be a welcome addition to the literature, there also is a need for qualitative investigations examining more completely the meaning of "participation" to individual students. Investigation of how participation in out-of-school experiences such as children’s choirs, private piano lessons, home music instruction, and religious music instruction might affect academic success would be studies of consequence. Although any type of experimental study that included a control group from whose members music was withheld would be unconvincing, a study where schools have decided to actively improve the quality of their music programs, with subsequent examination of school tests scores pre and post, might yield some very important additions to the body of literature examining music education and academic performance.

NOTES

1. The z-score represents the magnitude of a score’s deviation from the mean and indicates the score’s relative placement withing a normal distribution with a mean of zero and a standard deviation of one.

2. This percentage is based on the area under the normal curve. This area is calculated by subtracting the area of the curve below the low z-score from the area of the curve under the higher z-score. This difference is the percentage of the area between the scores. For example, in Table 1, the z-score of .01 has an area of .5040 under it. A z-score of .39 has an area of .7224. The difference of these scores indicates a difference of 21.84 of the area under the normal curve, or approximately 22%.

REFERENCES


Submitted May 11, 2006; accepted December 20, 2006.