The Effect of Early Music Training on Child Cognitive Development

Terry D. Bilhartz, Rick A. Bruhn, and Judith E. Olson
Sam Houston State University

The relationship between participation in a structured music curriculum and cognitive development was studied with 71 4- through 6-year olds. Children were pre- and posttested with six subtests of the Stanford-Binet Intelligence Scale, fourth edition (SB) and the Young Child Music Skills Assessment (MSA). Approximately one half of the sample participated in a 30-week, 75-minute weekly, parent-involved music curriculum. Statistical analysis showed significant gains for participants receiving music instruction on the MSA and on the SB Bead Memory subtest. Four-order partial correlations analysis found musical treatment influence on Bead Memory scores when the participants were controlled for sex, ethnicity, parental education, and economic class. Treatment also produced higher scores on other SB measurements for select populations. This study suggests a significant correspondence between early music instruction and spatial–temporal reasoning abilities.

In a 1987 historiographical essay, Draper and Gayle (1987) detailed the history of American attitudes toward the social value of music education. According to their study, the idea that early musical instruction produces benefits beyond the realm of the arts has been circulating for more than a century. The primary justifications used to promote early musical training, however, have changed over time. Of the traditional justifications given to support childhood musical study, the only rationale that has increased significantly among supporters of the arts during the late twentieth century is the premise that music promotes cognitive development and abstract thought.

In recent decades, a number of researchers have suggested a link between musical and spatial reasoning abilities (Davidson & Scripp, 1989; Gromko & Poorman, 1998; Hassler, Birbaumer, & Neil, 1985; Hurwitz, Wolff, Bortnick, & Kokas, 1975; Leng, Shaw, & Wright, 1990; Rauscher, Shaw, Levine, Wright, Dennis, & Newcomb, 1997; Rideout & Taylor, 1997). Like other forms of intelligence, spatial reasoning involves the ability to establish relationships between items. Researchers often distinguish between two types of spatial abilities: spatial recognition and spatial–temporal reasoning. The former ability is a process that involves identifying

Direct all correspondence to: Terry D. Bilhartz, Department of History, Sam Houston State University, SHSU Box 2239, Huntsville, TX 77341 <his_tdb@shsu.edu>.

Journal of Applied Developmental Psychology 20(4): 615–636
ISSN: 0193-3973

and mentally sorting items according to size, shape, color, pattern, and so forth. Spatial–temporal reasoning, however, is a process that requires mentally maintaining images without the assistance of a physical model and then transforming and combining these images in ways that create a meaningful whole. This mental process of sequentially and spatially arranging items in useful ways is used to perform higher brain functions such as playing chess and solving advanced mathematical equations. According to a growing number of scholars, this ability—or an ability with very similar characteristics—is also used by musicians in the performance of musical tasks (Leng & Shaw, 1991; Patel, Peretz, Tramo, & Labreque, 1998; Sarnthein, von Stein, Rappelsberger, Petsche, Rauscher, & Shaw, 1997; Serafine, 1988; Sergent, Zuck, Terriah, & MacDonald, 1992).

Leng et al. (1990) developed a neurobiological model that can be used to explain the causal link between music and spatial–temporal reasoning. This model, known as the *Trion model*, proposes that the ability to compare and find relationships among patterns is predicated by spatial–temporal firing patterns of interconnected groups of neurons spread over large regions of the cortex. According to the developers of this model, musical activity such as playing an instrument requires the same neural firing patterns that are needed in the performance of other forms of spatial–temporal reasoning. Because these firing patterns can be enhanced through learning and repeated experiences, and because both musical reasoning and spatial intelligence rely on similar cortical pattern development, it was predicted that music exposure could be used to strengthen spatial reasoning, particularly if this exposure was given to young children whose cerebral cortexes were still maturing.

The development of the Trion model of the cortex has spurred a number of behavioral researchers to test the hypothesis that music and spatial task reasoning use similar neural mechanisms. In 1993, Rauscher, Shaw, and Ky (1993) found that college students scored significantly higher on a spatial task test after listening to 10 minutes of Mozart. The benefit of passive listening to music on spatial task performance, however, was short term, lasting no more than 10 to 15 minutes. In more extensive follow-up studies, Rauscher, Shaw, and Ky (1995) and Rauscher et al. (1997) linked nonpassive early musical exposure with long-term spatial–temporal reasoning. According to the results in Rauscher et al. (1997), after 6 months of treatment, preschool children who received both 10 to 15 minutes of private keyboard instruction twice weekly and 30 minutes of daily group singing scored higher on an object assembly test of spatial–temporal reasoning than three other control groups of children. The three control groups included children who received group singing instruction without private keyboard enhancements, children who received individualized computer instruction instead of the keyboard lessons, and children who received no additional music or computer enhancements. Because none of the groups improved significantly on a spatial recognition task, the researchers concluded that early keyboard instruction, coupled with exposure to group singing enrichments, enhances one specific form of intelligence—spatial–temporal reasoning abilities.

In a subsequent related study, Gromko and Poorman (1998) examined the effect of nonkeyboard early music training on spatial–temporal task performance. In this study, one group of preschoolers from a private Montessori school received

Table 1. Control and Experimental Group Locations and Numbers

Control Group Locations	*Experimental Group Locations*
Schools Representing Lower Income Households	
School 1: Head Start Location (n = 11)	School 2: Head Start Location (n = 12)
Schools Representing Middle Income Households	
School 3: Large Private Preschool (n = 9)	School 3: Large Private Preschool (n = 11)
School 4: Small Private Preschool (n = 3)	
Schools Representing Higher Income Households	
School 5: Small Private Preschool (n = 12)	School 6: Private Music Center (n = 13)

7 months of weekly 30-minute instruction in singing and playing songs with a 20-note set of songbells. A second group of children from the same school received no music treatment. The experimental group showed significantly greater improvement than the control group on the raw scores (but not on the age-adjusted scaled scores) of a Wechsler Performance IQ test (Wechsler, 1989).

Pioneering studies such as these have attracted the attention of a number of researchers and the national media in a diversity of fields ranging from American studies to neurobiology (Iwaki, Hayashi, & Hori, 1997; Lamb & Gregory, 1993; Mohanty & Hejmadi, 1992; Standley & Hughes, 1997). To date, most of the published studies have involved small samples drawn from a narrow population of upper middle-class children. Other population groups must be examined to determine to what degree demographic and familial factors impact the connection between music and spatial–temporal reasoning. Moreover, although Rauscher and her colleagues have demonstrated a strong link between early keyboard instruction and spatial–temporal reasoning performance, the Gromko and Poorman study, which offered nonkeyboard musical instruction to the experimental group, produced less definitive results. Do other forms and levels of musical experiences also produce cognitive gains in young children, and are these gains limited to the area of spatial–temporal reasoning? This project was designed to address several of these unanswered questions posed by this provocative research on the interactions between music and the brain.

Method

Participants

Seventy-one 4- and 5-year olds in east Texas counties were recruited to participate in the project. The 71 participants consisted of approximately equal numbers of boys and girls and represented diverse economic and ethnic backgrounds. The children attended classes at one of the following locations: two rural Head Start centers, four preschools in a small city (population 35,000), or a music center in the same city (see Table 1). Of this number, 23 children in the Head Start program represented lower income households, 20 children at a large preschool and 3 children at a small preschool in the city represented middle income households, and 25 children from another private preschool and the music center in the city represented higher income households.

Of the original 71 participants, 36 were selected for the experimental treatment group and 35 for the control group receiving no treatment. Students were assigned to groups for several reasons. For the lower income group, all children at one Head Start location were placed in the experimental group, whereas all the children at a second location 10 miles away were assigned to the control group. Several parents from the middle income locations did not want their children to participate in the experimental group, and thus their children were assigned to the control group. All other children at these centers were placed randomly into the groups. The children at the higher income preschool were assigned to the control group, whereas the students from the music center, also representing higher income households, were assigned to the experimental group.

The ages of the children in September 1997 ranged from 51 to 72 months. There were no significant age differences between the control and experimental groups (one-way ANOVA: $F(1,69) = 2.448$; $p < .122$). Approximately 70% of the participants were White, 17% African American, 7% non-Black Hispanic, and 6% Asian. There were no significant differences in the ethnic composition of the control and experimental groups ($\chi^2 = 6.454$; $Df = 3$; $p < .092$). However, the limited availability of participants of appropriate age, coupled with the unwillingness of parents to consent to the treatment in some instances, prevented an equal distribution of boys and girls into the control and experimental groups. There were 22 boys and 13 girls in the control group, and 14 boys and 22 girls in the experimental group. Boys were significantly underrepresented in the experimental group from the preschools that represented middle-income level households ($\chi^2 = 9.295$; $Df = 1$; $p < .002$). Because boys often achieve marginally higher scores than girls of similar age on abstract reasoning tests (Linn & Peterson, 1985; Thorndike, Hagen, & Sattler, 1986b), the underrepresentation of boys in the experimental group will not compromise the conclusions of the study if a significant effect of music training on abstract reasoning is found.

At the beginning of the study, 77% of the control participants and 67% of the experimental participants lived in two-parent households. These differences were not statistically significant ($\chi^2 = .963$; $Df = 1$; $p < .327$). The mean number of years of schooling of the parents of the participants of both groups was approximately 13 years. Approximately 28% of the mothers and 30% of the fathers of the participants had completed 4 or more years of college. There were no significant differences between the control and experimental groups in the education levels of the mothers or in the education levels of the fathers of the participants in the study (one-way ANOVA: Mothers' education: $F(1,69) = .286$; $p < .594$; Fathers' education: $F(1,65) = 1.343$, $p < .251$).

After the participants were recruited and assigned to the experimental and control groups, the parents or guardians of the 71 participants were asked to complete a questionnaire about their children. The questionnaire sought information about the history and familial background of the child, including date of birth,

Table 2. Inferred Abilities for Selected Stanford-Binet Intelligence Scale: Fourth Edition Subtests

Subtest	*Inferred Abilities*
Vocabulary	Vocabulary development, verbal expression, concept formation, and meaningful long-term memory effect
Memory for Sentences	Verbal comprehension, knowledge of English syntax, short-term auditory memory, and attention
Bead Memory	Visual analysis; visual imagery; visual memory, sequencing, chunking, or clustering strategies; attention; flexibility; and manual dexterity
Pattern Analysis	Part-to-whole synthesis, visual analysis, spatial visualization, planning ability, visual–motor coordination, manual dexterity, and time pressure effect
Quantitative	Mathematical concepts/computation and the ability to analyze word problems
Copying	Visual imagery, spatial visualization, visual perception, visual–motor coordination, attention, and manual dexterity

previous time in preschool, sex, ethnicity, the number and ages of the child's siblings, and the marital status, occupation, years of education, and native language of the child's mother and father. The instrument also included 22 agree/disagree questions that assessed the caregivers' opinions regarding the child's interests, health, behavior, interest in music, activities, and future plans. Caregivers from all 71 participants returned the questionnaires. Among the original 71 participants, it was noted that 12 students had already received some Kindermusik (an age-appropriate, holistic curriculum designed to develop musical listening, movement, and singing skills) instruction before the study. Nine of these students were in the experimental group, and three were in the control group.

Descriptions of Pre- and Posttests and Treatment

Before the initiation of music treatment, the 71 participants were given a battery of SB subtests and an MSA. The SB (Thorndike, Hagen, & Sattler, 1986a) is a widely respected measure of cognitive abilities with good concurrent validity, high reliabilities, excellent standardization, good administration procedures, and helpful scoring criteria (Hendershott, Searight, Hatfield, & Rogers, 1990; Sattler, 1992; Spruill, 1987). According to its test developers, each SB subtest measures a variety of abilities contributing to overall cognitive ability (Delaney & Hopkins, 1987; Thorndike et al., 1986b). To collect data that maximized the range of tested abilities and minimized the required testing time, the researchers selected six of the eight SB subtests that are available for this age level. The subtests used in this study, along with the test developers' descriptions of the inferred abilities of each subtest (Delaney & Hopkins, 1987), are presented in Table 2.

Other researchers conducting confirmatory factor analyses on the SB for this age level have reported that the test measures three principal cognitive abilities: the "g" or general abilities factor, a verbal factor, and a nonverbal abstract reasoning factor. These studies indicate that the Vocabulary and Memory for Sentences subtests have a high loading on the verbal factor, whereas the Bead Memory, Quantitative, and Pattern Analysis subtests load more highly on the abstract reasoning factor (Gridley & McIntosh, 1991; Kaplan & Alfonso, 1997; Kline, 1989; Laurent, Swerdlik, & Ryburn, 1992; Thorndike, 1990). The Copying subtest has high test specificity and weak loadings on the abstract reasoning factor (Boyle, 1990; Thorndike et al., 1986a) All six subtests load on the general abilities factor. (A subsequent confirmatory factor analysis of the data created in this study corroborates these general conclusions.) Composite scores derived from the subtests that load highly on each of the three principal factors measured by SB were also used to explore for effects of music treatment on the development of general abilities, verbal abilities, and abstract reasoning abilities.

The MSA was developed and used for the first time in this study. The MSA composite score was derived from four musical skills subtests designed to measure the ability to: (a) maintain a steady beat, (b) recall and reproduce rhythmic patterns, (c) recall and reproduce vocal pitches, and (d) discriminate between tones played on a glockenspiel. Descriptions of the MSA along with its scoring protocols are provided in the Appendix.

Six advanced graduate students at Sam Houston State University and one member of the research team from the graduate counseling program administered the SB subtests. The administration and scoring of each SB was double checked by the research team member trained in the SB. A professional musician and music teacher not otherwise connected with the research project administered the MSA.

From September 1997 through May 1998, the 35 children in the control group attended their respective preschools but received no additional in-class music treatment. Meanwhile, the 36 students in the experimental group participated in a Kindermusik for the Young Child Year 1 Pilot Program (Swears, 1998). Three classes of approximately 12 students per class met for 75 minutes once weekly for 30 weeks. Two licensed Kindermusik (KM) educators taught the classes at a Head Start Center and two other educational facilities.

Kindermusik is a program of music and movement for children from birth to 7 years. Its goal is to encourage children to explore the world creatively through their voices, bodies, and minds. The weekly lessons for the age group in this study involve vocal exploration and matching pitch, singing, playing percussion instruments and the glockenspiel, exploring and notating basic rhythms, learning to read and write music on a treble staff, composing, and developing coordination and balance through movement. The curriculum encourages direct parent or caregiver involvement in the program. Caregivers agree to attend with their children the entire music class on the first and sixteenth lessons, and the last 15 minutes of all other lessons. In addition, the children receive materials that provide tools for repeating and expanding classroom activities at home. Caregivers agree to guide the learning process by assisting with weekly home assignments. Each child is encouraged to listen daily to a CD that includes songs, games, and dances from class. During the second semester, the home assignments also involve playing games and practicing songs and patterns on a glockenspiel.

Before the initiation of treatment, parents or guardians of the participants in the experimental group agreed to meet the KM caregiver guidelines. To monitor compliance levels with these standard curriculum expectations, the music educators kept a weekly record of the attendance of the participants and the caregivers and of the completion of the homework assignments.

In May 1998, the six SB and four MSA subtests were readministered to 66 of the original 71 participants. Two experimental group children dropped out of the KM program after the first semester. These participants, however, returned in May for the post-testing and their scores were included among those in the experimental group. Five other children (2 control and 3 experimental) were not posttested because they had either moved away from the community or were absent from their preschools during the May posttesting period. Although the pretest scores of all 71 participants were used to convert the MSA raw scores into age-normed MSA scores, the results reported in this study are based on the test scores of the 66 students (33 control; 33 experimental) who received complete pre- and posttesting.

The Null Hypothesis and Design Modifications

The null hypothesis stated that early musical training does not produce measurable growth in cognitive, spatial–analytical, and musical skills. Using the null format, it was hypothesized that there would be no significant differences between the pretest or posttest mean scores of the control and experimental groups on each of the SB and MSA subtests and composite scores. The researchers anticipated that the null hypothesis would be rejected for the MSA subtests because the KM curriculum offered to the experimental group was designed specifically to produce improvement in child musical development. A rejection of the null hypothesis for the MSA tests would bring assurance that some measure of music treatment was delivered to the experimental children via the structured KM classes. The more critical question to be investigated was whether the independent variable—weekly music training in a structured curriculum—also produced measurable cognitive growth on nonmusical tasks.

The original design was to use the difference in means for independent observations test (*t*-test) to determine the effect of treatment on each of the SB and MSA pre- and posttest outcomes and to use ANOVA procedures to examine the effect of selected demographic variables on subtest performance. Because several independent *t*-tests would be performed simultaneously, Bonferroni corrective methods were used to account for the number of comparisons being performed, thus reducing the probability of obtaining spurious positive results.

During the treatment period, the researchers observed significantly different compliance patterns among the participants in the three KM classes. Although there were no significant differences in the weekly attendance records of the participants at the three locations, after the first 6 weeks of the study, parental participation in the weekly KM classes fell sharply at the Head Start facility, and to a lesser degree at the middle income center. Similarly, compliance with the out-of-class KM curriculum assignments was significantly lower among the participants at the middle income center than among those at the higher income center and was almost nonexistent among the participants at the Head Start Program. The participants in the higher income experimental group completed nearly twice as many out-of-class assignments as the participants in the middle income experimental group, and approxi-

Table 3. KM Attendance/Compliance Patterns Among Higher, Middle, and Lower Income Participants Throughout the 30-Week Treatment Period

	Class Averages for Weeks					
Location	*1–6*	*7–12*	*13–18*	*19–24*	*25–30*	*Total*
Higher Income Location (n = 13)						
% Children Attending	96	86	77	80	68	81
% Parents Attending	86	81	74	78	68	78
% of Classes With More Than 50% Unattended Children	0	0	0	0	0	0
% of Out-of-Class Assignments Returned	71	65	58	52	46	59
Middle Income Location (n = 11)						
% Children Attending	94	79	64	67	61	73
% Parents Attending	89	77	62	62	58	70
% of Classes With More Than 50% Unattended Children	0	0	0	0	0	0
% of Out-of-Class Assignments Returned	50	55	09	32	20	35
Lower Income Location (n = 12)						
% Children Attending	89	88	78	85	78	83
% Parents Attending	61	31	22	17	14	29
% of Classes With More Than 50% Unattended Children	0	83	100	83	83	70
% of Out-of-Class Assignments Returned	38	21	06	02	06	15

mately four times the number of assignments as the participants at the Head Start facility. In-class parental participation levels were almost three times higher at the middle and the higher income centers than at the Head Start facility. Continual attempts by the teachers to encourage parental participation and homework follow through produced minimal results. A summary of these patterns is reported in Table 3. Because of these differences among the classes within the experimental group, the researchers decided to test for the effects of music treatment at each income level as well as for the effects of treatment on the combined sample.

In addition, the researchers theorized that if music treatment affects posttest outcomes on any subtest, then the differences between the experimental and control group posttest scores on that subtest should be greater for children who received higher levels of music treatment than for children, also assigned to the experimental group, who received less music treatment. Thus in addition to performing t-tests that compare mean scores for the experimental and control groups for the various income levels, it was decided to create a series of KM compliance variables that would discriminate between those in the experimental groups who met and failed to meet specified KM compliance standards. Unlike the original experimental/

control group variable (which divided participants into treatment and nontreatment groups but did not require minimum participation levels), the KM compliance variables limited the treatment group to those who passed set minimum compliance standards. To be included among those in the "minimal treatment" category, participants had to pass 40% attendance and 20% parental attendance standards. The "low compliance" category required 50% student attendance, 30% parental attendance, and 10% homework completion. Other variables demanded 10% higher participation rates in each of the three KM compliance categories. For instance, the category "modest compliance" included those with at least 60% attendance, 40% parental attendance, and 20% homework completion rates. Similarly, "satisfactory compliance," "above average compliance," and "high compliance" demanded participation rates, respectively, of 70-50-30%, 80-60-40%, and 90-70-50%. The researchers hypothesized that on subtests affected by music treatment, the difference in means scores between the treated and the nontreated would increase as the criteria levels for compliance were raised.

Before conducting these difference in means tests, an analysis was performed to determine the possible effect of test examiner on SB pre- and posttest outcomes. This analysis revealed significant differences among the scores of the test examiners on the SB Copying subtest. No test examiner effect was identified for the other SB and MSA subtests. A review of the literature on this particular SB subtest found that other researchers have reported similar scoring problems on the Copying subtest, especially when examining children of this age group (Boyle, 1990; Choi & Proctor, 1994; Mason, 1992; Prewett, 1992). Given these developments, the research team decided to eliminate the Copying subtest scores from the study. The five remaining subtests were used to calculate the General Abilities Composite. The Verbal Reasoning Composite was derived from the Vocabulary and Memory for Sentences subtests, and the Abstract Reasoning Composite (ARC) from the Bead Memory, Pattern Analysis, and Quantitative subtests.

Results

Analysis of Pretest Scores

There were no significant differences in the pretest SB and MSA outcomes between the experimental and control groups. However, when higher compliance standards were invoked to define music treatment, significant differences in pretest scores appeared between the experimental and control groups. This difference emerged because the children in the experimental group who failed to meet the minimum compliance standards also generally scored below average on the SB pretests. Consequently, one is cautioned against interpreting higher posttest SB outcomes for those in the higher criteria treatment groups as evidence causally linking music training and cognitive development. The higher posttest scores for those participants are more likely an artifact of the tendency for children with highly involved parents to perform better on tests than children whose parents are less involved. To demonstrate a link between music treatment and cognitive development, the experimental participants should not only score higher than the

Table 4. Summaries of *t*-Test Results: Mean △ Value Differences Between Groups That Met and Failed to Meet Select KM Minimum Compliance Criteria

	Select KM Compliance Criteria					
	No Minimum		*Low Compliance*		*High Compliance*	
	Mean	*Probability*	*Mean*	*Probability*	*Mean*	*Probability*
SB						
Vocabulary	1.85	.175	2.10	.151	1.46	.561
Memory for Sentences	−2.58	.079	−1.25	.415	0.93	.665
VR Composite	−0.73	.733	0.85	.712	2.39	.531
Pattern Analysis	−1.88	.357	1.24	.821	7.69	.104
Quantitative Reasoning	3.03	.112	0.26	.898	3.43	.234
Bead Memory	3.58	.061	6.82	.001	9.39	.001
AR Composite	0.83	.515	2.77	.063	6.84	.016
GA Composite	0.10	.976	1.22	.215	3.90	.031
MSA						
Steady Beat	6.09	.005	4.05	.080	1.86	.744
Rhythmic Pattern	4.95	.012	3.49	.092	1.84	.393
Vocal Pitch	4.14	.050	1.88	.398	0.73	.768
Aural Discrim	0.87	.716	1.07	.668	3.24	.391
MSA Composite	5.37	.003	3.17	.100	2.27	.424

Note: Group 1: met minimum KM compliance criteria; Group 2: failed to meet minimum KM compliance criteria. MEAN = Mean △ Value Group 1–Mean △ Value Group 2.

control participants on the posttests, but they should also demonstrate greater improvement than the control population during the testing interval itself. To measure rates of change (improvement versus declension) during the testing interval, Δ variables were created for each subtest, with the Δ value of a subtest being the difference between the posttest standard age score and the pretest standard age score of participants on the particular subtest. Positive Δ values indicate more rapid improvement during the testing interval than the national norm on the subtest. Negative Δ values indicate slower than expected development of the participants during the testing interval. Those showing no change in their pre- and posttest standard age score (that is, participants with a Δ value score of 0) improved their performance on the age-normed SB standardized test at precisely the national average.

Analysis of Δ Value Scores

MSA Subtests. As anticipated, the posttest and Δ value scores on the MSA composite, the MSA steady beat subtest, the MSA rhythmic pattern subtest, and the MSA vocal pitch subtest were significantly higher for the combined experimental group than for the control group. (See Table 4 for summaries of the effect of select levels of treatment on MSA and SB Δ values.) The only MSA subtest not significantly affected by treatment was the aural discrimination subtest, a test designed to identify the development of perfect pitch. However, the effect of treatment on MSA scores was not uniform across all income groups. For example, among lower income children only, the experimental children out-performed the control children on only the rhythmic pattern subtest, whereas among middle income children, treatment produced significantly higher scores on only the vocal pitch subtest. In contrast, the higher income children exposed to music treatment out-performed the other higher income children on all the MSA subtests except the aural discrimination subtest. The experimental children from all income levels who met the higher compliance criteria also had higher posttest scores and generally showed greater improvement during the testing period than the control children. These results confirm that compliance with the KM curriculum produces musical development in young children and that the greater the compliance, the greater the gains in musical skills. The results also suggest, however, that considerably less music treatment was delivered to the children in the lower income groupings than in the higher ones.

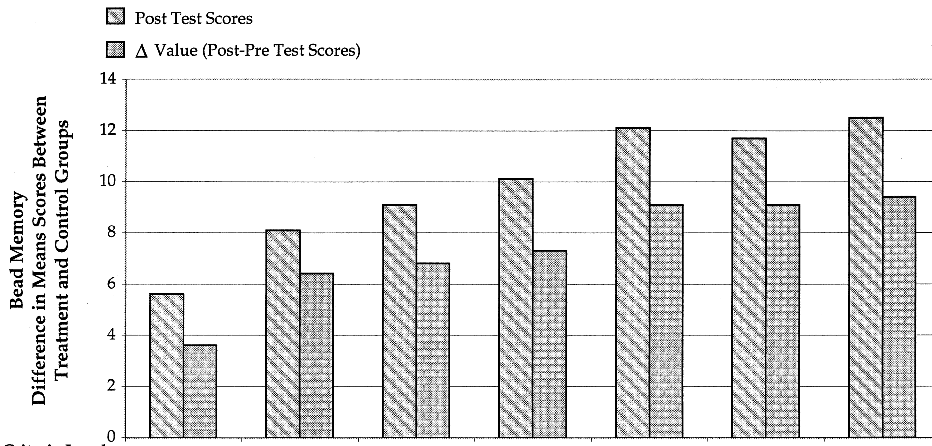
Stanford-Binet Subtests. The following overview describes the impact of music treatment at different income and compliance levels on each of the SB tests.

Vocabulary: There were no significant differences between the experimental and control groups in Vocabulary Δ values for the combined sample ($p < .175$), although experimental group children from high income households did show greater vocabulary improvement during the testing period than control group children from high income households ($p < .035$). Higher levels of music compliance did not significantly affect the rate of improvement on the vocabulary subtests.

Memory for Sentences: There were no significant differences in Memory for Sentences Δ values between the experimental and control children of any income group. Higher KM treatment criteria levels also did not produce significant changes. Music treatment made no observable impact on Memory for Sentences outcomes.

Bead Memory: Experimental group children from middle and higher income households showed significantly greater improvement on their Bead Memory scores than the control participants from middle and high income households (one-way ANOVA: non-Head Start Exp: $F(1,43) = 6.29$; $p < 016$). Although there was no significant difference in Bead Memory Δ values between the experimental and control groups among low income households, the Head Start children who met minimal KM compliance standards did have marginally higher Δ values than the Head Start children who did not receive this level of music treatment.

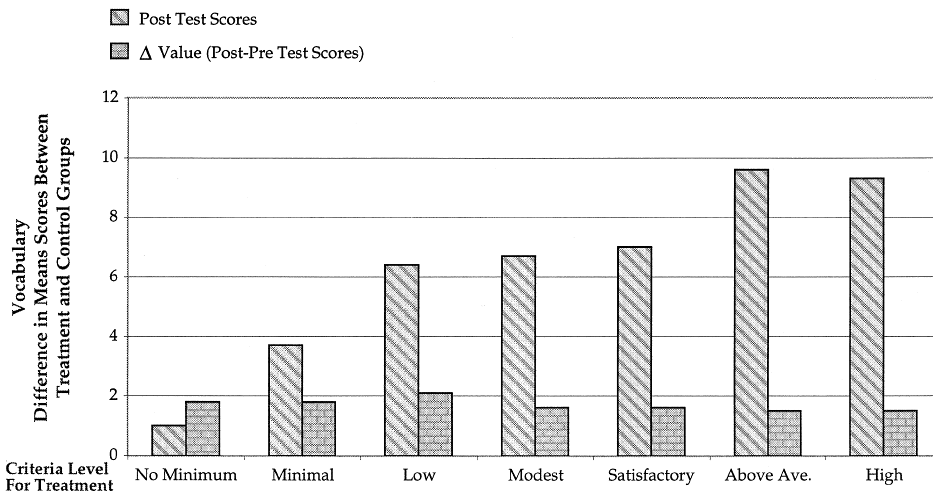
High KM compliance affected Bead Memory improvement for children in all income groups. Approximately one half (16 of 33) of the experimental participants improved significantly (Δ values greater than or equal to 4) on their Bead Memory standard age scores during the testing period. This group of children attended an average of 87% of the classes, had parental assistance during 80% of the lessons, and returned 60% of the weekly homework assignments. In all three compliance categories, these levels of participation were significantly greater than the mean participation rates of the experimental children who did not show Bead Memory improvement. These children, on the average, attended 74% of the music classes, received parental assistance in 37% of the lessons, and returned 25% of the homework assignments.



Criteria Level For Treatment	No Minimum	Minimal	Low	Modest	Satisfactory	Above Ave.	High
Size of Criteria Qualifiers Group (Q)	Q=33	Q=27	Q=23	Q=21	Q=18	Q=13	Q=8
Size of Criteria Non-Qualifiers Group (NQ)	NQ=33	NQ=39	NQ=43	NQ=45	NQ=48	NQ=53	NQ=58
Difference in Means Probability on Bead Post Test	.011	.000	.000	.000	.000	.000	.002
Bead Δ Value	.061	.001	.001	.000	.001	.001	.011

Figure 1. Difference in mean Bead Memory scores for treated vs. control groups by treatment criteria levels.

A summary of these and related findings are illustrated in Figures 1–7. Figure 1 shows that the average Bead Memory Δ value scores were higher for those in the experimental than the control group and that the degree of improvement increased linearly with the degree of music treatment. The average magnitude of the improvement among those who met minimal compliance standards was 6.4 points, equivalent to an increase from the fiftieth percentile on the SB standardized test to above the seventy-eighth percentile. For children meeting higher compliance criteria, the improvement was greater, reaching 9.1 points for those who met satisfactory compliance standards, equivalent to a jump from the fiftieth to above the eighty-seventh percentile. The histograms in Figures 3–7 provide visual comparisons of the range of Δ scores among children in the control group, among children in the experimental group, and among children in the experimental group who completed 50% or more of the KM curriculum expectations. As these frequencies indicate, those assigned to the experimental group were twice as likely to show significant improvement on Bead Memory during the testing interval than the control participants. Moreover, those who attended at least one half of the classes with their caregivers or who completed at least one half of their out-of-class music assignments were more than three times as likely to show significant improvement on Bead Memory during the treatment period than those in the control group.



Size of Criteria Qualifiers Group (Q)	Q=33	Q=27	Q=23	Q=21	Q=18	Q=13	Q=8
Size of Criteria Non-Qualifiers Group (NQ)	NQ=33	NQ=39	NQ=43	NQ=45	NQ=48	NQ=53	NQ=58
Difference in Means Probability on Vocabulary Post Test	.598	.065	.001	.001	.002	.000	.002

Figure 2. Difference in mean Vocabulary scores for treated vs. control groups by treatment criteria levels.

Improvement in Bead Memory scores also was predicted by the variable named PREMUSIC. The 12 children exposed to early KM treatment before the beginning of this study (PREMUSIC) improved more during the treatment period than children who did not receive early music treatment (one-way ANOVA: premusic $F(2,63) = 5.75$; $p < .005$). Of these children, three were in the control group and thereby received no additional music treatment, three received middle levels of KM treatment, and six received higher levels of KM treatment.

Pattern analysis: There were no significant differences in improvement rates on the Pattern Analysis subtest between the experimental group and the control group at any income level. Kindermusik compliance also did not generally predict improvement in Pattern Analysis scores, although students who met high compliance standards did have Δ values approximately 8 points higher than the other children in the study ($p < .012$). Children who received KM instruction before the beginning of this study also had Δ values more than 6 points higher than the other children in the study, a noteworthy yet insignificant trend ($p < .060$). This data suggests, but does not confirm, a link between early or intensive KM exposure and Pattern Analysis gains.

Quantitative reasoning: The rate of improvement among the experimental group during the testing period was not significantly higher than the rate of improvement

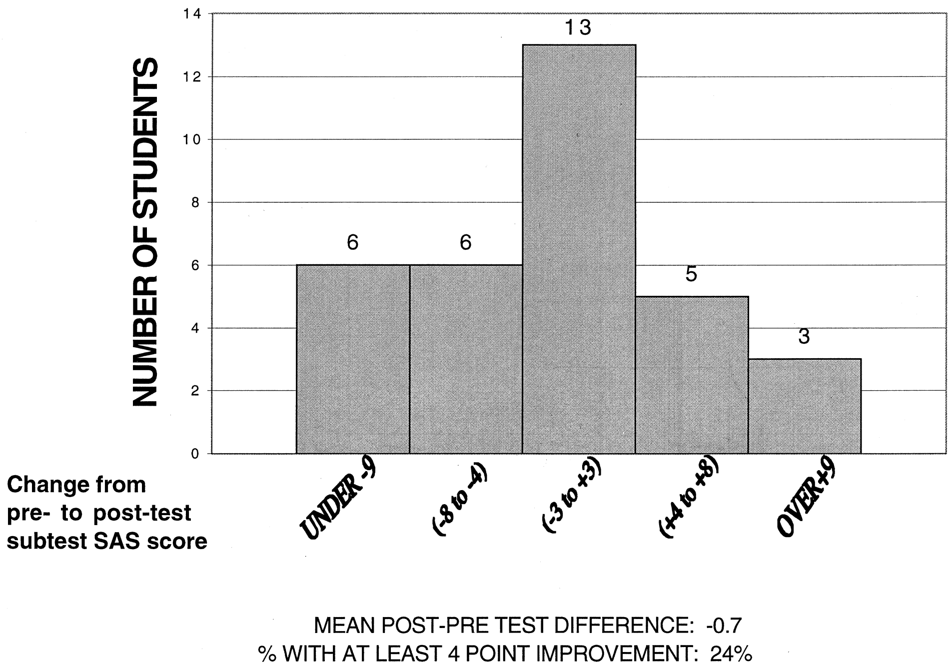
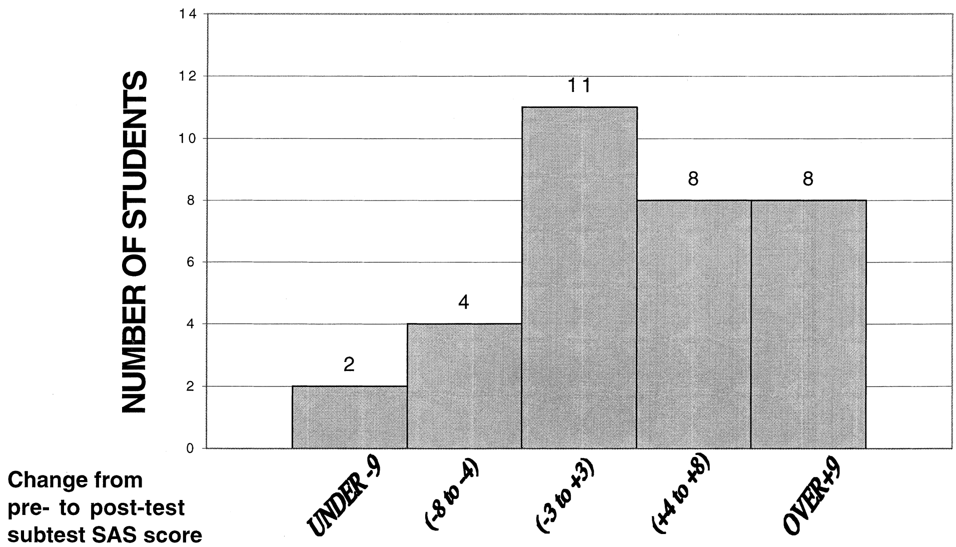


Figure 3. Bead Memory Δ value ranges for children in the control group ($n = 33$).

of the other children. High KM compliance also did not significantly affect rates of quantitative reasoning gains.

Stanford-Binet Composite Scores. Table 5 presents a summary of the correlations between selected music treatment variables and the posttest and Δ value composite variables that measure the three principal factors assessed by the SB subtests. The data demonstrates that the music treatment variables correlate more strongly with high performance on the abstract reasoning than on the verbal reasoning components of the SB. Although the effect of music instruction on gains in the verbal reasoning component was unsubstantiated, those who received even low levels of music treatment generally showed significant improvement during the treatment interval on their Abstract Reasoning Composite (ARC) scores (one-way ANOVA: low compliance independent, $F(1,64) = 4.667$; $p < .034$). Similar improvement also was demonstrated among those who received KM exposure previous to the beginning of this study (one-way ANOVA: PREMUSIC independent, $F(2,63) = 6.87$; $p < .002$). Some of this effect, however, can be attributed to nontreatment influences. For instance, when four-order partial correlation procedures were used to control for the influence of the child's sex, ethnicity, parental education, and household income, the only treatment variables that correlated significantly with ARC improvement were the above average and high compliance music treatment variables.



MEAN POST-PRE TEST DIFFERENCE: +2.9
% WITH AT LEAST 4 POINT IMPROVEMENT: 48%

Figure 4. Bead Memory Δ value ranges for children in the experimental group (n = 33).

Discussion

The results of this study lend support to the hypothesis that there is a significant link between early music instruction and cognitive growth in specific nonmusic abilities. Even minimally musically treated children in this study scored significantly higher than the control children on one measurement of abstract reasoning ability, the SB Bead Memory subtest. Moreover, the improvements in Bead Memory scores were greatest for those who participated most fully in the music exercises.

This link between music treatment and Bead Memory scores is of particular importance because this subtest measures both visual imagery and sequencing strategies, mental processes that Leng et al. (1990) have theorized require the same neural firing patterns that are needed in the performance of musical activity. The Bead Memory subtest requires students to recall and manipulate an assortment of beads of various colors (red, white, and blue) and shapes (round, ellipsoid, cone, and saucer). In the easier tasks, the child must recognize and remember the shape and color of one, and then two beads. At the next level of difficulty, the child looks at a picture of several beads placed on an upright plastic stick. The stimulus is then removed and the child is asked to select the right color and shape of bead(s) from a box, and place the beads on the upright stick in the exact order as the picture. Successive tasks increase the number of beads, thus testing the child's ability to remember longer sequences. The evidence of this study demonstrates that children

Table 5. Relations Between Select Music Treatment Variables and SB Composite Variables: Zero-Order Correlations and 4-Order Partials Controlling for Participant's Sex, Ethnicity, Parental Education Levels, and Household Income

	SB Composite Vars					
	Verbal Reasoning		*Abstract Reasoning*		*General Abilities*	
	Post-test	△ *Value*	*Post-test*	△ *Value*	*Post-test*	△ *Value*
Zero-Order Correlations, KM Compliance Vars						
Premusic	.535***	.068	.498***	.429***	.542***	.397***
No Minimum	.016	−.041	.211*	.082	.140	.047
Minimal	.184	−.043	.327**	.202	.286**	.146
Low	.334**	.046	.417***	.261*	.408***	.236*
Modest	.324**	.027	.432***	.226*	.413***	.198
Satisfactory	.356**	.109	.522***	.359**	.484***	.347**
Above Average	.365**	.145	.531**	.457***	.493***	.444***
High	.273*	.089	.470***	.440***	.415***	.404***
4-Order Partials[a], KM Compliance Vars						
Premusic	.328**	.022	.258*	.210	.326**	.192
No Minimum	−.049	−.095	.232*	.025	.129	−.032
Minimal	.119	−.105	.330**	.125	.276*	.041
Low	.236*	.012	.352**	.124	.347**	.103
Modest	.217*	−.002	.375**	.078	.353**	.060
Satisfactory	.184	.092	.435***	.201	.377**	.206
Above Average	.126	.160	.369***	.286*	.305**	.309**
High	.059	.106	.353**	.325*	.262*	.310**

Note: *p = .050. **p = .010. ***p = .001.
[a] 4-Order Partials controlled for sex, ethnicity, parental education, and household income.

trained to produce music vocally and on a glockenspiel—sequential training that uses and develops kinesthetic, aural, and visualization skills—become better able to perform the abstract reasoning tasks measured by the SB Bead Memory subtest.

The improvements among the treated children in Bead Memory scores and, correspondingly, in the ARC can be partly but not fully attributed to nontreatment factors. When four-order partial correlations procedures were used to control for the influence of parent's education, household income, and child's sex and ethnicity, the correlation between lower levels of music treatment and ARC improvement was weakened to the extent that it failed to pass a significance test at the predetermined 95% confidence level (see Table 5). However, the effect of high KM treatment levels on Bead Memory Δ scores remained significant even at a 99% confidence level when controlled for these four demographic variables. These findings indicate that the rates of improvement among treated participants on the Bead Memory subtest and on the ARC were not always consistent across all demographic groups. The socioeconomic variations, however, followed predictable patterns that correspond directly with intensity of received music treatment. Although the average time in class was approximately the same for the experimental children in the Head Start facility as for the children in the other experimental groups, owing to the lack

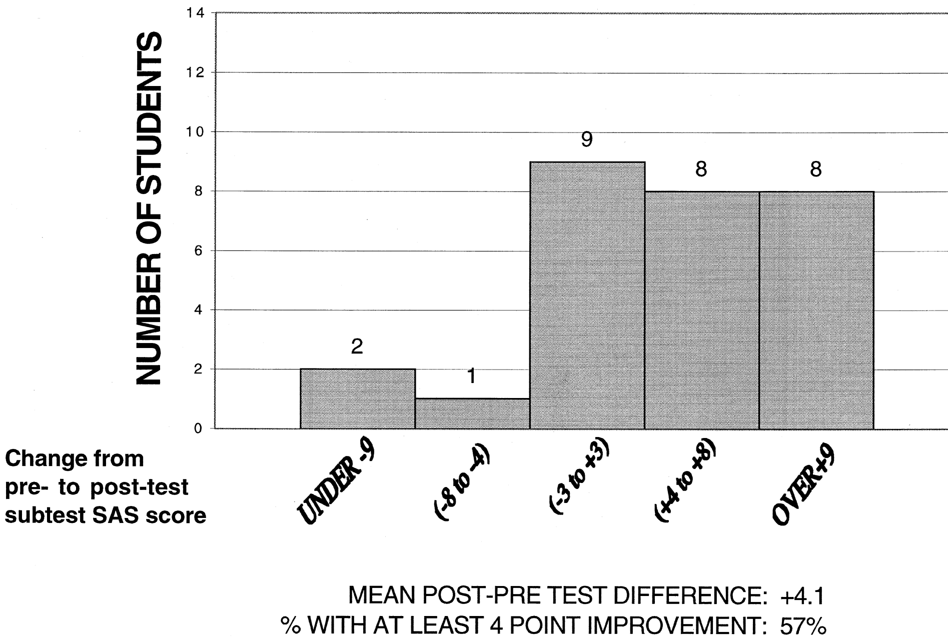


Figure 5. Bead Memory Δ value ranges for children who attended more than 50% of KM classes (n = 28).

of caregivers during the treatment period at the Head Start location, the quality of in-class instructional time was much less among these children. These children, along with the children in the middle income group, also received less out-of-class music treatment than the experimental children in the higher income group. Evidences of lower levels of received music treatment among the Head Start children were observed both in their attendance/compliance records and in their lower MSA posttest and Δ value scores. Given these differing levels of received treatment, it is not surprising to find a stronger link between KM attendance patterns and the ARC among children in higher income households, the participants who received greater exposure to music treatment.

This study did not find a significant effect of music treatment on the other SB measurements of abstract reasoning, although some evidence does suggest a possible link between high levels of KM instruction and Pattern Analysis gains. It is interesting to note that the SB Pattern Analysis subtest is similar to the Wechsler Block Design subtest that Rauscher and her colleagues used to measure spatial recognition abilities. In the SB Pattern Analysis subtest, the child is asked to create a pattern with blocks that matches a pattern displayed on blocks or a picture. Success at this task involves visual perception, part-to-whole synthesis, and manual dexterity, but not short-term memory or sequencing strategies. In the Rauscher studies, no correlation was found between music exposure and improvement on the Wechsler Block Design subtest, and consequently it was concluded that the benefits of music instruc-

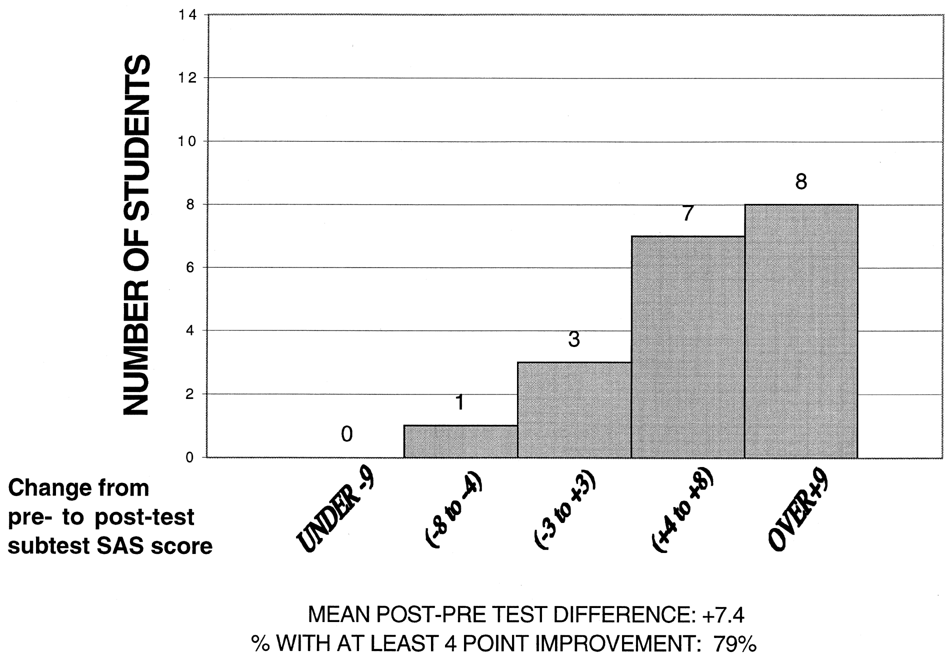


Figure 6. Bead Memory Δ value ranges for children with caregivers who attended more than 50% of KM classes (n = 19).

tion were limited to spatial–temporal (not spatial recognition) reasoning abilities. The fact that this study did not find a significant link between low levels of KM treatment and Pattern Analysis gains corroborates with Raucher's general conclusions. We, however, remain unconvinced that the correspondence between high levels of treatment and early KM exposure and improved Pattern Analysis performance was coincidental. More research is needed to determine if early music treatment produces gains in cognitive abilities measured by the SB Pattern Analysis subtest.

Finally, although the children who attended music classes and met KM compliance expectations tended to score higher than the other children on their SB Vocabulary subtests, there were no significant differences in the Vocabulary Δ values between the treated and the nontreated children, except among the high income children (see Figure 2). Consequently, the evidence in this study does not confirm a causal link between early music instruction and verbal reasoning abilities. The researchers hypothesize that the higher Vocabulary posttest scores of the children who participated fully in the music program were a consequence of enhanced verbal interactions between those children and their caregivers, including but not limited to the child–parent interaction that occurred while completing the out-of-class KM assignments. A detailed discussion of the effects of parental and environmental factors on cognitive and musical development is beyond the scope

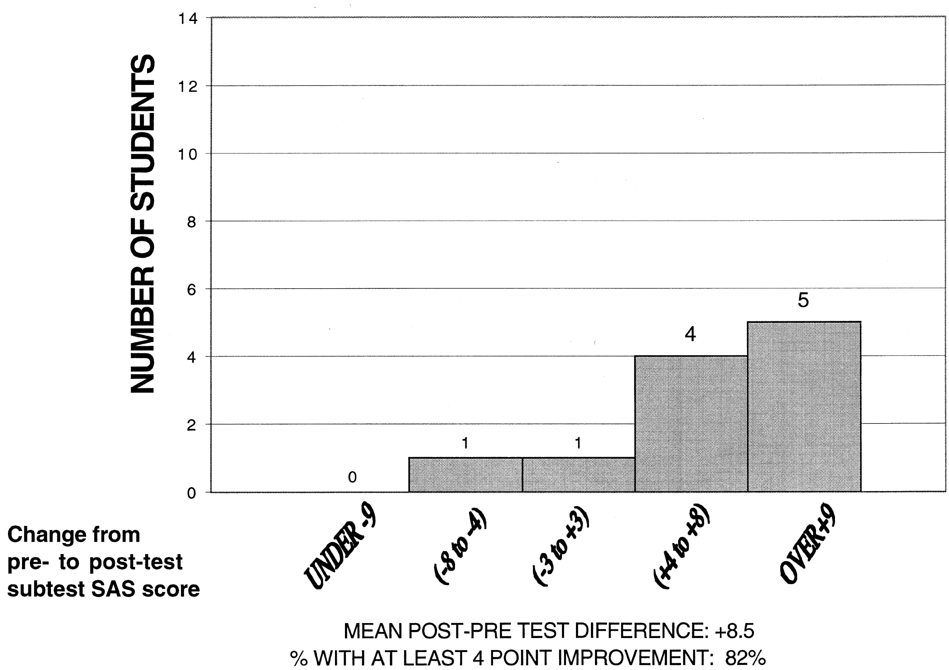


Figure 7. Bead Memory Δ value ranges for children who completed more than 50% of homework assignments (n = 11).

of this article. However, the strong correlations found in this study between levels of parental involvement in the music program and posttreatment scores on both SB Verbal and Abstract Reasoning measurements underscore the importance of parental activities as an influence in the cognitive development of young children. (A manuscript discussing the impact of family relationship variables on SB and MSA outcomes is in progress.)

In conclusion, promoters of the arts have assumed a linkage between music instruction and cognitive development for more than a century. Until recently, the evidence supporting this linkage was anecdotal. Today, however, social science research validates this long-held premise. Rauscher and her colleagues in their 1997 study demonstrated that preschool keyboard instruction produced improved scores on an Object Assembly task, a Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1989) measurement of spatial–temporal reasoning. The present study, which uses different forms of music treatment and different outcome measurements, demonstrates that KM instruction enhances the abstract reasoning abilities measured by the SB Bead Memory subtest.

Perhaps the greater questions for future behavioral researchers concern what particular forms of music treatment, in what amounts, and at what ages produce the greatest cognitive gains. The powerful predictive powers of the variable PRE-

MUSIC (a measurement of KM exposure that predated the beginnings of this study) on future Bead Memory scores, and perhaps on Pattern Analysis scores, suggest that accelerated cognitive gains in abstract reasoning abilities may result from exposing infants and toddlers to select early musical experiences. Additional work needs to be carried out to determine the optimal times for introducing young children to the benefits of music. Knowledge gained from such pursuits will provide parents, educators, and state and national education policy makers with crucial information needed to maximize the benefits among our nation's young.

Acknowledgments: This research was supported by grants from Sam Houston State University (SHSU) and Kindermusik International, Inc. (KI). KI supplied the student materials and provided a stipend for the licensed KM instructors, Kelly Sharp and Patty Bilhartz. We would like to thank the preschool directors, Marla Shelly, Bobette Beggs, and Sheila Yarborough, for their assistance in recruiting the children and for their cooperation and support throughout the project. Appreciation also is given to the music tester Ellen Brown, and to SHSU graduate students Tom Clark, Chuck Johnson, Christie Gamblin, Vanessa Mc Elmeel, Melissa Templeton, and Kim Abernethy for their work in testing the children. We also wish to thank James Horn of SHSU's Computer Services Center for his technical assistance, and Cecil Hallum of SHSU's Math Department for reviewing earlier drafts of this manuscript.

Appendix

Young Child Music Skills Assessment (MSA)

The MSA consists of four subtests and was used for the first time in this study.

In the Vocal Pitch subtest, the child was asked to recall and vocally reproduce patterns of two or three tones sung by the tester using the syllable "bahm." Each attempt was scored on a 4-point system. Zero points were scored for failing to match the pattern, one point for matching one of the tones, two points for matching two of the three tones, three points for matching the contours but not the exact pitches of the pattern, and four points for accurately matching the complete tonal pattern.

The Steady Beat subtest required the child to tap a steady beat with rhythm sticks while listening to music. Low scores were awarded to those who were unable to maintain the steady beat for at least 7 measures, mid scores for maintaining the beat for 7–11 measures, and high scores for maintaining the beat for 12–16 measures.

In the Rhythmic Pattern subtest, the child was asked to recall and rhythmically reproduce patterns played on rhythm sticks by the tester. Each of the four patterns gradually increased in difficulty. Weighted scores on the Steady Beat subtest and the Rhythmic Pattern subtest were combined and used as a measure of rhythm and steady beat skills.

In the Aural Discrimination subtest, selected tones were played on a glockenspiel by the tester, and the student was asked to remember the sounds of the tones. Then, the tester played additional tones on the glockenspiel, and the student was asked if these tones matched the tones played earlier. The advanced levels of this test were difficult and were designed to identify children with exceptional aural abilities.

A composite raw score was created by adding the weighted scores from the vocal pitch, steady beat, rhythmic pattern and aural discrimination subtests. The pretest raw scores for the MSA composite and the four MSA subtests were converted into z scores for three age groups: children under 56 months, children 56 to 59 months, and children more than 60 months. These age brackets correspond to the age brackets used in the age-normed SB. Similarly, z scores were created from the posttest raw scores for children in each of the three age brackets. Finally, to enable MSA comparisons with the SB subtests, the MSA pre- and posttest z scores were adjusted to have a mean of 50 and an *SD* of 8, the scale units of the SB subtests.

References

Boyle, G. J. (1990). Stanford Binet IV Intelligence Scale: Is its structure supported by LISREL congeneric factor analyses? *Personality and Individual Differences, 11*(11), 1175–1181.

Choi, H., & Proctor, T. (1994). Error prone subtests and error types in the administration of the Stanford-Binet Intelligence Scale: Fourth edition. *Journal of Psychoeducational Assessment, 12*(2), 165–171.

Davidson, L., & Scripp, L. (1989). Education and development in music from a cognitive perspective. In D. J. Hargreaves (Ed.), *Children and the arts* (pp. 59–86). Milton Keynes, UK: Open University Press.

Delaney, E., & Hopkins, T. (1987). *Examiner's handbook: An expanded guide for fourth edition users.* Chicago: Riverside.

Draper, T., & Gayle, C. (1987). An analysis of historical reasons for teaching music to your children. In J. C. Perry, I. W. Perry, & T. W. Draper (Eds.), *Music and child development* (pp. 194–205). New York: Springer-Verlag.

Gridley, B. E., & McIntosh, D. E. (1991). Confirmatory factor analysis of the Stanford-Binet: Fourth edition for a normal sample. *Journal of School Psychology, 29*(3), 237–248.

Gromko, J., & Poorman, A. (1998). The effect of music training on preschoolers' spatial-temporal task performance. *Journal of Research in Music Education, 46*(2), 173–181.

Hassler, M., Birbaumer, N., & Neil, A. (1985). Musical talent and visual-spatial abilities: Longitudinal study. *Psychology of Music, 13*, 99–113.

Hendershott, J., Searight, H., Hatfield, J., & Rogers, B. (1990). Correlations between the Stanford-Binet fourth edition and the Kaufman Assessment Battery for Children for a preschool sample. *Perceptual and Motor Skills, 71*(3), 819–825.

Hurwitz, I., Wolff, P., Bortnick, B., & Kokas, K. (1975). Nonmusical effects of the Kodaly music curriculum in primary grade children. *Journal of Learning Disabilities, 8*, 45–51.

Iwaki, T., Hayashi, M., & Hori, T. (1997). Changes in alpha band EEG activity in the frontal area after stimulation with music of different affective content. *Perceptual & Motor Skills, 84*, 515–526.

Kaplan, S. L., & Alfonso, V. C. (1997). Confirmatory factor analysis of the Stanford-Binet Intelligence Scale: Fourth edition with preschoolers with developmental delays. *Journal of Psychoeducational-Assessment, 15*(3), 226–236.

Kline, R. B. (1989). Is the fourth edition Stanford-Binet a four-factor test? Confirmatory factor analyses of alternative models for ages 2 through 23. *Journal of Psychoeducational Assessment, 7*(1), 4–13.

Lamb, S., & Gregory, A. (1993). The relationship between music and reading in beginning readers. *Educational Psychology, 13*, 19–26.

Laurent, F., Swerdlik, M., & Ryburn, M. (1992). Review of validity research on the Stanford-Binet Intelligence Scale: Fourth edition. *Psychological Assessment, 4*(1), 102–112.

Leng, X., & Shaw, G. L. (1991). Toward a neural theory of higher brain function using music as a window. *Concepts in Neuroscience, 2*, 229–258.

Leng, X., Shaw, G. L., & Wright, E. (1990). Coding of musical structure and the Trion model of cortex. *Music Perception, 8*, 49–62.

Linn, M. C., & Peterson, A. C. (1985). Emergence and characterization of sexual differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479–1498.

Mason, E. M. (1992). Percent of agreement among raters and rater reliability of the Copying subtest of the Stanford-Binet Intelligence Scale: Fourth edition. *Perceptual and Motor Skills, 74*(2), 347–353.

Mohanty, B., & Hejmadi, A. (1992). Effects of intervention training on some cognitive abilities of preschool children. *Psychological Studies, 37*, 31–37.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language, 61*(1), 123–144.

Prewett, P. N. (1992). Short forms of the Stanford-Binet Intelligence Scale: Fourth edition. *Journal of Psychoeducational Assessment, 10*(3), 257–264.

Rauscher, F., Shaw, G., & Ky, N. (1993). Music and spatial performance. *Nature, 365*, 611.

Rauscher, F., Shaw, G. & Ky, N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letter, 185*, 44–47.

Rauscher, F., Shaw, G., Levine, L., Wright, E., Dennis, W., & Newcomb, R. (1997). Music training causes long-term enhancement of preschool children's spatial-temporal reasoning. *Neurological Research, 19*, 2–8.

Rideout, B. E., & Taylor, J. (1997). Enhanced spatial performance following 10 minutes exposure to music: A replication. *Perceptual and Motor Skills, 85*(1), 112–114.

Sarnthein, J., von Stein, A., Rappelsberger, P., Petsche, H., Rauscher, R. H., & Shaw, G. L. (1997). Persistent patterns of brain activity: An EEG coherence study of the positive effect of music on spatial-temporal reasoning. *Neurological Research, 19*, 107–116.

Sattler, J. M. (1992). *Assessment of children's intelligence and special abilities* (3rd ed.). Boston: Allyn & Bacon.

Serafine, M. (1988). *Music as cognition: The development of thought in sound.* New York: Columbia University Press.

Sergent, J., Zuck, E., Terriah, S., & MacDonald, B. (1992). Distributed neural network underlying musical sight-reading and keyboard performance. *Science, 257*(5066), 106–109.

Spruill, J. (1987). Stanford-Binet Intelligence Scale, fourth edition. In D. Keyser & R. Sweetland (Eds.), *Test critiques: Volume VI* (pp. 544–559). Kansas City, Missouri: Test Corporation of America.

Standley, J. M., & Hughes, J. E. (1997). Evaluation of an early intervention music curriculum for enhancing prereading and writing skills. *Music Therapy Perspectives, 15*(2), 79–85.

Swears, L. (1998). *Kindermusik for the young child: Introductory information for young child curriculum.* Greensboro, North Carolina: Kindermusik International, Inc.

Thorndike, R., Hagen, E., & Sattler, J. (1986a). *The Stanford-Binet Intelligence Scale: Fourth edition, Guide for administering and scoring.* Chicago: Riverside.

Thorndike, R., Hagen, E., & Sattler, J. (1986b). *The Stanford-Binet Intelligence Scale: Fourth edition, technical manual.* Chicago: Riverside.

Thorndike, R. M. (1990). Would the real factors of the Stanford-Binet fourth edition please come forward? *Journal of Psychoeducational Assessment, 8*(3), 412–435.

Wechsler, D. (1989). *Wechsler preschool and primary scale of intelligence—revised: Manual.* San Antonio: Psychological Corp.