Mute Those Claims: No Evidence (Yet) for a Causal Link between Arts Study and Academic Achievement

ELLEN WINNER and MONICA COOPER

In the American educational climate of today, "basic" academic skills are valued while the arts are considered a frill. Many major urban school districts have cut back on arts education in order to strengthen academic subjects.[1] Even though most of our schools have some arts education, and even though most of our citizens say they want their children to be exposed to the arts in school,[2] only one in four students in American schools sings, plays an instrument, or performs plays in class each week.[3] When budgets are tight, the arts are almost always the first programs to be cut.

In reaction to the marginalization of the arts, arts educators and arts advocates have understandably sought evidence that the arts contribute to basic academic skills.[4] A strong belief has grown among policymakers and arts advocates that the arts can play a powerful role in education because skills and attitudes learned through the arts can help children in non-arts, academic areas of learning.

Here are a few recent quotations that express this instrumental view of what the arts can do. According to testimony presented to the U.S. House of Representatives, "... studies dating back to 1989 have revealed that students involved in music programs show improved reading abilities, and higher math and science scores.... Because participation in music generates neural connections, it benefits those brain functions that aid the abstract reasoning that math and science require."[5] A letter to the *New York Times* by the violinist Isaac Stern states that "Teaching the arts to the very young, particularly music because it is such a natural thing for a child to sing or dance and to sense rhythm, also helps them excel beyond all norms in logic, memory and mathematics, and there are generally accepted studies and statistics to back this up."[6] Secretary of Education Richard Riley states that "The arts teach young people how to learn by giving them the first step: the desire to learn."[7] Senator Alan K. Simpson states that "A love of art helps the learning process. It helps you learn. All studies tell us that."[8] And a 1996 report put out by the President's Committee on the Arts and Humanities states that "At-risk youth show increased motivation to learn

Journal of Aesthetic Education, Vol. 34, Nos. 3-4, Fall/Winter 2000

and improve academically when participating in arts education programs outside of school."[9]

These claims are typically made without much thought to what underlying mechanism might account for such a causal link. There are two very different kinds of explanations for why learning in the arts might generalize to learning in an academic subject area. As explained below, the most direct link from learning in the arts to learning in other disciplines is a link in cognitive structure; a more circuitous link would be a motivational one.[10]

The Cognitive Structure Argument

Some cognitive structures developed by learning in the arts might be the same as some needed to do well in academics. If so, cognitive skills learned though the arts could be applied to learning in an academic area. Candidate skills and structures that might be learned in the arts and transferred to academic disciplines include focusing, close observation, critical, divergent, or independent thinking, problem solving, and problem finding. Such transfer is unlikely to occur unless students are made aware of the possibility of transfer. For example, a teacher might point out the importance of close observation in the visual arts and note that this same kind of skill can be applied to the study of science.[11]

The Motivational Argument

There are a number of ways in which learning in the arts might stimulate motivational and attitudinal changes that could then spill over into academic studies.

Entry Points. When used as entry points into an academic area, the arts may lead otherwise unmotivated or non-academically oriented students to develop an interest in the academic subject area in question. This is usually what is meant by "teaching through the arts" or "integrating the arts into the curriculum." Arts integration usually takes the form of the arts being brought into academic disciplines rather than the reverse; typically the integration occurs around a common theme or question: for example, music notation can be used as a way into ratios; drama can be used as a way into history; making a work of art can be used as a way into writing.[12] This kind of integration may be applied most effectively to subjects that naturally and authentically combine the arts with other disciplines. For instance, language arts are central to drama; music theory involves ratios; ceramics involves chemistry; cultural history includes the study of art and music; and Advanced Placement courses in art now often involve writing about one's portfolio.

Self-confidence. Participation in the arts might boost the self-confidence or status of students who discover they can perform well in an art form.[13] Such self-confidence may engender a more positive attitude towards school, which in turn could lead to greater participation, effort, and attention.

Perseverance. The arts may serve as a vehicle to develop discipline and perseverance in working on long-term projects, and these attitudinal characteristics may then be transferred to other subject areas.

High Standards. Because students in the arts often must perform their works publicly or put them on display, they may develop high standards for their work. The value placed on high standards may then transfer to other subjects.

Bonding. Arts that have a collaborative component, such as drama, dance, music, or mural making, may help students come to feel part of a group, and hence to feel less alienated from school. Such groups are opportunities for focusing on learning and may be the foundation for creating a community of learners.

Positive Mentors. Given the fact that arts teachers often work with students individually, like coaches, participation in the arts may provide students with positive mentors who may convey the importance of staying in school and working hard.[14]

Stress Reduction. Participation in the arts may result in relaxation and stress reduction. This may allow students to return refreshed and motivated to their academic studies.

Could the Link Be an Epiphenomenon?

There is another reason why the arts may be linked to positive academic outcomes. It is possible that schools that decide to grant the arts a central role in the curriculum also make other kinds of reforms in the way that academic subjects are taught. Schools that value the arts may also promote innovative, inquiry-oriented, project-based academic work. Schools that value the arts may attract the best kinds of academic teachers—energetic, innovative, and imaginative. Finally, schools that value the arts may attract certain kinds of students—those from families who value the arts. And families who value the arts may also value academic achievement.

In this case, the link between arts and academic achievement would be epiphenomenal. The arts would simply be indicative of other aspects of the school that are themselves directly linked to and causally implicated in academic improvement.[15] This kind of explanation is open to empirical investigation, but we know of no systematic research evaluating this possibility.

In this climate of rising claims, the same secondary and tertiary sources are quoted over and over. It is time to take a careful and dispassionate look at what the primary research shows. It is also time to consider whether it serves the cause of the arts well to advance utilitarian arguments in their favor, or whether we should instead justify the arts in terms of their inherent (and unique) value.

In what follows, we analyze the primary research testing the claim that study of the arts is associated with improved academic outcomes. Unlike most of the papers in this issue, the studies analyzed here are ones assessing the effects of more than one of the four major art forms (music, visual arts, drama, dance). The effects of specific arts forms could not be disentangled. We have included two kinds of studies: those that are purely correlational, testing the claim that students who study the arts are higher academic achievers than those who do not; and those that are experimental, testing the claim that academic outcomes improve as a function of studying the arts (with or without evidence about the underlying mechanism that might be involved).

Ideally only experimental studies should be used to examine whether studying the arts leads to improved academic outcomes, since correlational studies demonstrate only that an association exists but do not demonstrate anything about direction of causality. The existence of a correlation cannot tell us whether the arts cause academic improvement, whether academic achievement causes involvement in the arts, or whether study of the arts and achievement in academics are both caused by some third independent factor. However, we included both correlational as well as experimental studies because many of the claims made by policymakers today are based on correlational findings. One goal of this study was to compare the overall effects of correlational vs. experimental studies.

We searched exhaustively for all relevant studies (published in English) that appeared from 1950 to 1998. We were unable to find any experimental studies that provided a test of which causal mechanism might underlie academic improvement as a function of arts study. Thus, the research we review below tells us only (1) whether there is a correlation between arts study and academic achievement; and (2) whether academic achievement improves when students are exposed to the arts. The research that has been carried out on this question has not been designed to reveal the mechanism underlying any improvement shown by the experimental studies.

Method

In order to identify all relevant studies, published as well as unpublished, we searched seven electronic data bases from their inception through 1998: Arts and Humanities Index (1988-1998), Dissertation Abstracts International (1950-1998), Educational Resource Information Clearinghouse (1950-1998), Language Linguistics Behavioral Abstracts (1973-1998), MedLine (1966-1998), PsychLit/PsychINFO (1984-1998), and Social Science Index (1988-1998). The search terms we entered were as follows: art (first string), instruct or train or educat or program (second string), learn or academ or cognit or achiev or intell or IQ (third string), measur or find or outcome or effect (fourth string), and transfer (fifth string). In addition, we conducted handsearches of 41 journals from 1950 to 1998 (listed in Table 1 of the introductory paper in this issue) that publish articles in education, development, and the arts. We checked the bibliographies of all identified articles; and we sent requests to over 200 arts education researchers for unpublished data or manuscripts not yet published (for which we received a modest rate of return).

Our search yielded 1135 research "records." Because our search criteria were liberal, many records that turned up were clearly irrelevant (e.g., those on the liberal arts). Others were potentially relevant but lacked data (e.g., advocacy pieces) or lacked control groups. Forty-four studies were potentially relevant. After applying a set of strict inclusion criteria, we were left with 31 useable studies, from which we calculated 66 effect sizes, as documented in Table 1. (From here on, we refer to each effect size as a separate study.) We selected studies that examined instruction in "the arts" in general, rather than instruction in a particular art form. In addition, the studies selected had to assess some form of non-arts, academic achievement as their outcome, and they had to be in English. Studies had to have a comparison or control group. Hence, one-group pretest-posttest studies were not included, as such studies have many threats to internal validity and produce effect sizes that are over six times larger than those of well-designed studies.[16] Table 2 reveals that most of the studies retained were "fugitive": that is, they were unpublished (in the form of unpublished dissertations, technical reports, and conference presentations) rather than appearing in peer-reviewed publications.

Studies with insufficient data to calculate an effect size, along with studies excluded due to lack of control group, were included in a table classified by direction of finding (positive vs. null or even negative [where the arts group performed worse than the control group]), and were then submitted to a binomial test to determine whether their results were positive more often than might be expected by chance (i.e., greater than 50%).

Table 1: Number of Research "Records" Identified vs. Number of Studies Used

Number Records Identified	Number of Studies Meeting Our Inclusion Criteria	Number of Effect Sizes Calculated
1135	31	66

Table 2: Number of Studies Used in Meta-analyses According to Publication Outlet

Unpublished	Published in Non-Peer Reviewed Publication	Published in Peer Reviewed Journal
28	2	1

Note: Unpublished papers include: doctoral dissertations, masters theses, technical reports, conference presentations, and unpublished data.

Ultimately we identified two broad categories of studies for meta-analysis: correlational and experimental studies. We looked separately at composite, verbal, and mathematical outcomes. Composite outcomes were ones in which verbal and math outcomes were summed. Composite outcomes were used only when separate verbal and math outcomes were not reported. This yielded five meta-analyses: Correlational Composite, Verbal, and Mathematical; Experimental Verbal and Mathematical. Studies assessing the relationship between arts education and outcomes that were classified as motivational (such as attendance, college aspirations, election to class office, etc.) were also tabulated but not submitted to a meta-analysis, since in most cases insufficient data were reported to compute an effect size. Of course, even if a link between arts education and motivation to attend and be engaged in school were demonstrated, one would still need to demonstrate that such increased motivation led to increased achievement.

Coding Procedure

Each study was coded by two independent coders (the first author and a research assistant) in terms of a variety of basic characteristics. The coders disagreed on 24 out of 425 codings, yielding a 5.6% rate of disagreement. In all cases the correct coding was fully resolved by rechecking the text.

Outcome. Studies were coded for whether they assessed composite, verbal, or math outcomes.

Year of Publication. Studies were coded for year of publication.

Research Design. Studies were coded as having one of three kinds of designs: correlational, experimental-matched, and experimental-unmatched.[17]

Correlational studies were ones that assessed academic achievement in students who had voluntarily chosen (self-selected) to study the arts. No pretest was given, and there was thus no way to determine whether students who did and did not choose the arts differed academically prior to studying the arts. No causal conclusions about the effects of the arts can be drawn from correlational studies.

Experimental-Matched studies were ones in which arts and comparison groups were matched prior to their exposure to the arts. These included (only two) true-experimental studies (in which students were assigned to arts and control groups randomly) and quasi-experimental studies (in which an intact arts classroom was compared to an equivalent classroom receiving no special arts). Quasi-experimental matched studies all attempted to match the ability level of students in the arts and control groups by statistically controlling for pre-existing differences. This was done in one of several ways: by means of analysis of covariance or regression in which preexisting differences are covaried or regressed; by analyzing gain scores; entering pretest and posttest scores as factors in an analysis of variance so that differences in mean gain between groups could be analyzed; or by carrying out the analysis directly on pre- posttest difference scores. It is only such studies that allow causal conclusions about the impact of the arts on academic achievement.

Experimental-Unmatched studies lacked a repeated measures pre-/posttest design and based their analyses solely on the posttest. In most cases researchers made an admirable attempt to find a comparison group in a similar school and from a similar socioeconomic background. Nonetheless, we cannot be certain that the groups were equivalent in ability prior to arts exposure.[18]

Age of Participants. Studies were coded as testing preschool/kindergarten level, elementary school level (grades 1-5), middle school level (grades 6-8), high school level (grades 9-12), or college level students.

Participant Characteristics. Studies were coded as testing average students or ones from some special group: for instance, low SES; high or low academic ability; or talent in the arts.[19]

Integrated vs. Separate. Studies were coded as assessing the effects of studying the arts in arts classes vs. the effects of programs in which the arts were integrated into the curriculum. A clear distinction could not be made between these two kinds of approaches because almost all of the programs in which the arts were integrated into the curriculum also taught the arts as separate disciplines.

Duration of Arts Exposure. Programs were coded as long in duration if they lasted longer than one academic year.[20] Programs were coded as short in duration if they lasted less than one academic year. Unfortunately we were unable to code programs for the amount of arts that students received per week, as many studies did not supply this information.

Sample Size. Studies were coded for size of sample.

Outlet. Studies were coded by where they appeared, as follows: peer-reviewed journal, non-peer-reviewed journal, doctoral dissertation, masters thesis, technical report/unpublished paper, conference presentation, published (unanalyzed) raw data, and unpublished raw data.

Perhaps the most important coding category of all is quality of the arts program. If arts learning transfers to academic learning, undoubtedly such transfer will be related to the quality of the arts teaching. One piece of evidence for this comes from a study conducted in England demonstrating that the relationship between arts experience and academic performance varied widely by school.[21] Unfortunately, we had no way of determining this critically important variable, as no study announced that the arts program under evaluation was poor. We were forced to use duration of arts exposure as a proxy for quality of program, but we are not satisfied with this measure as a measure of quality of arts instruction.

Calculating Effect Sizes

For each study we calculated one or more effect sizes (r). An effect size r is a measure of association, or correlation, between two variables. It indicates the strength of the relationship between these two variables—in our case, between exposure to the arts and academic achievement.[22] All subsequent calculations used the Fisher's transformation of the Pearson r (Zr) to account for the non-normal distribution of the rs.[23] However, we report the untransformed mean effect sizes.

We chose r as our effect size estimate rather than the often used d for a number of reasons. First, to calculate d it is necessary to have means and standard deviations. However, the studies included here often reported only significance tests (e.g., t, F) and associated p levels. A second reason for using r rather than d is that r is more flexible. For instance, in a study with three groups, one can calculate an r based on a hypothesized ordering of means; d cannot be used in such a situation.[24]

Each effect size was based on a separate, independent sample of participants. If a study reported more than one result with each result from a different group of participants (e.g., different age groups), we calculated an effect size for each group and then entered each as a separate "study." Effect sizes were calculated in eight different ways, depending on the type of information provided by the study, and studies were coded for the method by which the effect size was calculated. Methods for calculating effect sizes in this paper as well as the papers that follow, along with formulae, are given in Table 3. Results that were in the predicted direction (arts group performing better than the control) were reported as positive effect sizes. Results in the opposite direction were entered as negative.

Correlational Studies

Our first three meta-analyses were performed on correlational studies for composite, verbal, and mathematics outcomes, respectively.

Correlational Studies Composite Outcomes: Verbal + Math Summed

Our first meta-analysis was performed on correlational studies, listed in Table 4, whose outcomes were undifferentiated, composite academic measures summing or averaging performance over verbal and mathematical measures. We excluded studies whose outcomes were IQ scores, even though these are "composite" measures of verbal and mathematical ability, and looked only at achievement measures. (We demonstrated that excluding these studies did not alter our results, however.)[25]

We identified six relevant studies. One of these studies however, was dropped from the analysis since it was the only one whose subjects were preschool children. Dropping this study did not, however, alter the mean effect size reported below.[26] All but one of the other studies assessed academic

Table 3: Methods for Computing Effect Size *r*s

Method	Formula	When Used and Notes on Computations and Formulas
F-Method	$r =, \sqrt{F / F + df}$ from Rosenthal 1984/1991, p. 19, formula 2.17	When *F* values are provided and have only 1 *df* in the numerator.
Maximum Possible Contrast *F*-Method (MPC-*F*)	See Rosenthal & Rosnow, 1985, pp. 74-77 for computing *r*-contrast.	When more than 2 groups were compared, yielding *F* with more than 1 numerator *df*. Contrast weights summing to zero are assigned to means to test a theory (i.e., a higher weight is assigned to means of groups expected to perform well; a lower weight is assigned to means of groups expected to perform less well). A Pearson *r* is then computed that correlates the observed means with the weights. The Pearson *r* is then squared and called "r^2-alerting." The MPC-*F* is computed by multiplying the numerator degrees of freedom from the reported *F* by that *F*. Then the r^2-alerting is multiplied by the MPC-*F* to find the *F*-contrast. An *r*-contrast is computed from that as from any other *F* (see formula above). The *r*-contrast is then transformed to *r* by dividing it by the square root of $[(1-r^2) + r^2]/r^2$-alerting.
t-method	$r = \sqrt{t^2 / t^2 + df}$	When *t* values are provided or could be calculated and *df* are provided or could be calculated from *N*.
Chi-square (χ^2). method	Phi (ϕ) = $\sqrt{\chi^2 / N}$	When outcomes are dichotomous, we created tables of counts which yielded a chi-square (χ^2). The effect size for the χ^2 is phi (ϕ), which is a Pearson *r* on dichotomous data.
Glass's Delta (Δ) converted to *r*	$\Delta = (M_1 - M_2)/S$ computed from control group $\Delta = 2r/\sqrt{1-r^2}$	When means and standard deviations of control group are provided, with no *F* or *t* values reported; Ds then converted to *r*'s using formula 2.19, p. 19, Rosenthal, 1991. In the case of unequal *n*, we used the harmonic mean, necessary only when using the Δ-method.

Method	Formula	When Used and Notes on Computations and Formulas
p Method (Significance Method)	$r=Z/\sqrt{N}$	When only *p*'s associated with a significance level and *N*'s are reported, the *Z* corresponding to *p* was computed. When *p*'s were reported as "less than" X, the value of X was entered for *p*. When *p*'s are listed as, e.g., .00, the nearest possible higher value was entered (in this case, $p=.009$). When no *p* was given but results were reported as "nonsignificant," an effect size *r* of 0 was entered.
From the R^2 reported from a regression.	$\sqrt{R^2}$	When a regression was performed, with one factor regressed, and an R^2 was reported, we computed an effect size *r* by taking the square root of the R^2. This method was used only once, for Dwinell & Hogrebe (1984), who used ability to predict participation in the arts.
Pretest *r* subtracted from posttest *r*.	*r* posttest – *r* pretest	When means and standard deviations for pretests and posttests were supplied, but the authors did not control for pretest differences, *r*'s were computed for both pretest and posttest, and then the pretest *r* was subtracted from the posttest *r*. We were guided in this method by Cook & Campbell (1979), pp. 103-108 and Wortman (1994) p. 105-6, who recommend such a method when initial differences at pretest are not controlled.

achievement in high school and college-age students; one assessed students ranging in age from the elementary through high school years.

The meta-analysis was based on five studies. In all cases, these studies reported levels of academic achievement in students who voluntarily chose to study the arts. All of these studies demonstrated higher academic achievement among students who chose to study the arts than among students who did not choose to study the arts.

Study Characteristics

Study characteristics are shown in Table 4 and summarized below.

Sample Size. The total sample size of the studies analyzed was n= 3,408,635. Sample sizes ranged from n=200 to n=3,367,000 with a mean of n=681,727 and a median of n=17,143. The difference between the mean and the median indicates that a few large studies are inflating the mean. However, all studies were based on quite large samples: one study was between n=100-1,000; one was between n=1,000-10,000; two were between n=10,000-100,000; and one was between n=100,000-500,000.

Duration. These studies assessed academic achievement in high school seniors who had been intensively involved in the arts since 8th grade,[27] in high school seniors who had taken over 3 high school credits in the arts,[28] in college freshmen who had taken 4 semesters of arts in high school,[29] in high school seniors who had participated in band, orchestra, chorus, or dance in or out of school,[30] and in students anywhere between 8-20 years of age who had studied arts in after-school arts organizations for nine hours a week for at least one year.[31] Four studies thus assessed programs lasting a year or more and were coded as long; one study assessed an arts program lasting under a year and was coded as short.[32]

Integrated/Separate. All of the studies involved the arts taught as separate disciplines rather than as integrated into the academic curriculum.

Participant Characteristics. Three studies assessed achievement in average SES students; two assessed achievement in low SES students who can be considered academically at risk.[33] None of the studies assessed participants selected for high or low ability of any kind.

Age of Participants. Three studies assessed high school seniors, one assessed college freshmen, and one assessed students outside of school ranging in age from 8-20.

Outlet. Two studies were published in non-peer-reviewed journals. The rest were in the form of a doctoral dissertation (1), a conference presentation (1), and published raw data (1).

Results

Table 4 lists each study along with the sample size (N), mean effect size (r), Z associated with that effect size, and the p associated with the Z. (For each

Table 4. 5 Correlational Studies with Composite Outcomes: Meta-analysis 1

Study	N	r	$Z(p)$* (*<.0001)	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size a Method	Outlet
Catterall, Chapleau, & Iwanaga (1999) [a]	2,813	.08	4.42*	2 (Arts involvement from 8^{th} –12^{th} grade)	No	Composite scores	Low SES	12^{th} grade	Phi (ϕ) method	Non-peer reviewed publication
Dwinell & Hogrebe (1984)	21,479	.05	7.33*	1 (Any participation in senior year of high school)	No	Composite scores	Average SES	12th grade	Square root of R^2	Conference presentation
Heath (1998)	17,143	.04	4.88*	2 (9 hrs. week for at least 1 year after school arts)	No	Academic awards (self-report)	Low SES	Arts group: 8-20 years; Control group:10th grade	Phi (ϕ) method	Non-peer reviewed journal
National Center for Education Statistics (1984)	3,367,000	.05	97.48*	2 (Arts concentration in high school)	No	Composite scores	Average SES	High school	Phi (ϕ) method	Unpublished tabulated data
Whitener (1974) [b]	200	.04	-.31 (p=.38)	2 (At least 4 semesters high school art)	No	Composite grades	Average SES	College freshmen	F-method	Doctoral dissertation

Notes:

a. We calculated an effect size from the low SES sample of Catterall, Chapleau, & Iwanaga (1999) in order to render the results as comparable as possible to Heath (1998). Had we used the all SES 12th grade sample, our effect size would have been r=.18, considerably higher.

b. Whitener (1974) also reported results of an ANCOVA in which ability level as measured by SAT was controlled. We did not use the F from this analysis to calculate the effect size because we did not think it made sense to control for ability level at the end of 12th grade, after exposure to the arts, when the outcome measured was ability level in college (as measured by grades). Thus we used the t method to calculate effect sizes, as a t was reported for males and again for females. Because the sample sizes of males and females were identical, we were then able to average the effect size of males and females (averaging the Zrs, then reconverting these back to rs, as recommended by Rosenthal (1991).

meta-analysis reported here there is a similarly structured table.) Effect sizes are shown in a stem and leaf display in Table 5. They ranged from r=.04 – r=.08. A mean effect size of r=.05 (d=.10)[34] was found, and this remained unaltered when weighted by size of study. As recommended by Rosenthal,[35] we then combined the Zs from our studies to yield a Stouffer's Z=50.89, p<.001. This indicates that we can generalize our findings to other subjects who might have been included in these particular studies.

A more conservative test, the t-test of the mean Zr, indicates whether we can generalize our findings to new studies on this research question. This test yielded a value of 5.97, significant at p = .004. We can thus reject the null hypothesis and conclude that there is indeed a relationship between arts education and composite measures of academic achievement, and this relationship can be generalized both to new subjects who might have been selected for these studies and to future research studies on this question.

The 95% confidence interval indicates that, given another sample of five similar studies, the mean effect size is likely to fall within the range of r=.03 to r=.08. Because this interval does not span zero, we can be confident that, given another sample of five similar studies, the average effect size would be positive.

The standard deviation of the mean effect size was only .02. We tested whether this standard deviation was significant using the χ^2 statistic.[36] This analysis yielded a nonsignificant χ^2=7.18, df=4, p=.13. Thus, our effect sizes were not significantly different from one another.

A possible objection to any meta-analytic finding is that the researcher might have failed to uncover unpublished (hence "fugitive" studies) lurking in researchers' file drawers. Since unpublished studies are more likely to have null results than published ones, such a failure would result in a positivity bias. However, we conducted a comprehensive and exhaustive search and were able to locate many unpublished studies. Indeed, as seen in Table 2, almost all of the studies across all five meta-analyses were unpublished. Second, we performed a "file drawer analysis" to determine the number of studies which would have to be found hidden in researchers' file drawers and averaging null results (i.e., mean probability level of p=.50) in

Table 5. Stem and Leaf Display of 5 Effect Size *r*s from Meta-Analysis of Correlational Composite Studies

Stem	Leaf
+.2	
+.1	
.0	8, 5, 5, 4, 4
-.0	
-.1	
-.2	

order to bring the probability level of our Stouffer's Z down to the barely significant level of p=.05. In this case, 4,781 such studies would be required. Thus, we can be confident that even if we have failed to find some studies, entering them into our analysis would be highly unlikely to render our results nonsignificant (since it is rather unlikely that we have overlooked over four thousand studies averaging null results!)

Correlational Arts-Verbal Outcomes

A meta-analysis was next performed on 11 correlational studies assessing verbal outcomes, listed in Table 6.[37] Ten of these were studies released yearly (between 1987-1998, excluding 1993) by the College Board comparing verbal SAT scores of students with zero vs. four years of arts courses in high school. The 11th was a study of 12th graders who had been highly involved in arts classes since 8th grade.[38]

Study Characteristics

Study characteristics are shown in Table 6 and summarized below.

Sample Size. The total sample size of the studies analyzed was n=3,210,921. Sample sizes ranged from n=7,440 to n=367,314, with a mean of n=291,902 and a median of n=318,392. Of these, ten were between n=100,000-500,000, and one was between n=1,000-10,000.

Duration. All studies assessed the effects of arts instruction lasting at least one year. Ten of the studies were the College Board studies comparing students with four or more years of high school arts courses to students with none. One was the study by James Catterall assessing students who had had high arts involvement from the 8th to 12th grades. Thus, all of the studies assessed the verbal achievement of students with a nontrivial amount of arts involvement.

Integrated/Separate. All studies assessed the verbal performance of students in arts courses taught as separate disciplines.

Participant Characteristics. In all of the studies, participants were of average SES and no other special characteristics were noted.

Age of Participants. All participants were high school seniors.

Outlet. One of the studies appeared as a non-peer reviewed publication. Ten of the studies were from the College Board and were given to us in the form of unpublished tabulated data.

Results

Effect sizes, shown in a stem and leaf display in Table 7, ranged from .14 to .25, with a weighted and unweighted mean of r=.19 (d=.39). This mean effect size was shown to be significant both by the Stouffer's Z=333.43,

Table 6. 11 Correlational Studies with Verbal Outcomes: Meta-analysis 2

Study	*N*	*r*	*Z*(*p*)* (*<.0001)	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
Catterall, Chapleau, & Iwanaga (1999) [a]	7440	.19	16.24*	2 (Arts involvement from 8^{th}-12^{th} grade)	No	Reading scores	Average SES	12^{th} gr	Phi (ϕ) method	Non-peer reviewed publication
College Board (1988) [b]	353,679	.14	80.64*	2 (4 years high school art courses)	No	SAT verbal scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1989)	296,189	.15	80.42*	2 (4 years high school art courses)	No	SAT verbal scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1990)	274,168	.16	81.92*	2 (4 years high school art courses)	No	SAT verbal scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1991)	273,034	.17	86.95*	2 (4 years high school art courses)	No	SAT verbal scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1992)	269,453	.18	95.54*	2 (4 years high school art courses)	No	SAT verbal scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1994)	352,824	.20	121.48*	2 (4 years or more high school art courses)	No	SAT verbal scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data

Study	N	r	Z(p)* (*<.0001)	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
College Board (1995)	360,911	.21	129.10*	2 (4 years or more high school art courses)	No	SAT verbal scores	Average SES	12th gr	*t*-method	Unpublished tabulated data
College Board (1996)	367,314	.23	137.09*	2 (4 years or more high school art courses)	No	SAT verbal scores (recentered scale)	Average SES	12th gr	*t*-method	Unpublished tabulated data
College Board (1997)	337,517	.25	146.31	2 (4 years high school art courses)	No	SAT verbal scores (recentered scale)	Average SES	12th gr	*t*-method	Unpublished tabulated data
College Board (1998)	318,392	.23	130.18	2 (4 years high school art courses)	No	SAT verbal scores (recentered scale)	Average SES	12th gr	*t*-method	Unpublished tabulated data

Notes:

a. We used the all SES sample from Catterall, Chapleau, & Iwanaga (1999) to render the results as comparable as possible to the College Board results. Had we used the low SES 12th grade sample, our effect size would have been r=.10, considerably smaller.

b. For the College Board data, we compared students with no arts to those with four years of arts in order to test the effects of many years of arts. We did not compare no arts students to those with *over* four years because this group was a smaller sample than those with four years. We could not use 1993 College Board data because the standard deviations for scores from this year were not available. For 1994-96, students with 4 years arts are not separated from those with over 4 years. Scores from 1996-8 were recentered by the College Board, and hence were higher than earlier years. However, since we were not comparing across year, but rather within year across levels of arts courses, this was not a concern. Finally, *n*s were taken from raw data sent to us by the College Board. These *n*s are slightly different from the *n*s published in the College Bound Seniors Profile of SAT and Achievement Test Takers because the raw data separates 0 arts from 1/2 year of arts, while the published booklets combine these.

Table 7. Stem and Leaf Display of 11 Effect Size rs from Meta-Analysis of Correlational Verbal Studies

Stem	Leaf
+.3	
+.2	1, 2, 2, 3, 5
+.1	4, 5, 6, 7, 9, 9
.0	

$p<.0001$, and the *t*-test of the mean $Zr=16.52$, $p<.0001$. Thus we can generalize these findings both to other subjects and to other studies on this research question. The 95% confidence interval showed that mean unweighted effect size of another analysis of studies like these is likely to fall within the range of $r=.17$ to $r=.22$. A file drawer analysis revealed that 451,924 more studies averaging null effects would be needed to yield a nonsignificant Stouffer's Z. Thus our finding is highly resistant to the file drawer effect.

The standard deviation of the mean effect size was .04, and this was significant, $\chi^2=4963$, $df=10$, $p<.0001$. Thus, our effect sizes were significantly heterogeneous. We considered the possibility that heterogeneity was due to the fact that one of our studies came from a different "lab" (all but the study by Catterall, Chapleau, and Iwanaga came from the College Board). When we analyzed the College Board data only, heterogeneity remained high, $\chi^2=4963$, df=9, $p<.0001$. As shown in Table 2, effect sizes for the College Board data rose fairly steadily over the 11-year period examined from 1988-1998. A linear contrast test on year of data revealed that contrast weights associated with year of data correlated with effect sizes at $r = .98$, $Z= 68.63$, $p< .0001$. The heterogeneity should not be due to the fact that the last three years were based on a recentered scale, since students with no arts as well as those with 4 years of arts were both scored on the same recentered scale. However, to be certain that the recentering was not the cause of the heterogeneity, we eliminated these years from the analysis; heterogeneity still remained high, $\chi^2=1903$, df=6, $p<.0001$.

Correlational Arts-Math Outcomes

Next, a meta-analysis was performed on 11 identified correlational studies assessing math outcomes, summarized in Table 8. As with the correlational verbal studies, ten of these studies were released yearly (between 1988-1998, excluding 1993) by the College Board comparing math SAT scores of students with zero vs. four years of arts courses in high school. The 11th study was one in which 8th graders who had taken two or more years of arts at least four times a week were compared to those with no such arts involvement in terms of their performance on math problems.

Table 8. 11 Correlational Studies with Math Outcomes: Meta-analysis 3

Study	*N*	*r*	$Z(p)$* (*<.0001)	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	SES	Age	Effect Size Method	Outlet
College Board (1988)[a]	353,679	.05	31.66*	2 (4 years high school art courses)	No	SAT math scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1989)	296,189	.06	31.99*	2 (4 years high school art courses)	No	SAT math scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1990)	274,168	.08	40.95*	2 (4 years high school art courses)	No	SAT math scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1991)	273,034	.08	43.17*	2 (4 years high school art courses)	No	SAT math scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1992)	269,453	.10	49.34*	2 (4 years high school art courses)	No	SAT math scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data
College Board (1994)	352,824	.12	71.98*	2 (4 or more years high school art courses)	No	SAT math scores	Average SES	12^{th} gr	*t*-method	Unpublished tabulated data

Study	*N*	*r*	*Z*(*p*)* (*<.0001)	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	SES	Age	Effect Size Method	Outlet
College Board (1995)	360,911	.15	87.51*	2 (4 or more years high school art courses)	No	SAT math scores	Average SES	12th gr	*t*-method	Unpublished tabulated data
College Board (1996)	367,314	.15	91.28*	2 (4 or more years high school art courses)	No	SAT math scores (recentered scale)	Average SES	12th gr	*t*-method	Unpublished tabulated data
College Board (1997)	337,517	.18	100.93*	2 (4 years high school art courses)	No	SAT math scores (recentered scale)	Average SES	12th gr	*t*-method	Unpublished tabulated data
College Board (1998)	318,392	.14	80.43*	2 (4 years high school art courses)	No	SAT math scores (recentered scale)	Average SES	12th gr	*t*-method	Unpublished tabulated data
Demeter (1986)	128	.00	.00 (p=.50)	2 (2 or more years art or music at least 4 x wk.)	No	Multi-step math problems	Average SES; at or above grade level in math	8th gr	Signif. Method	Doctoral dissertation

Note:
a. See note b, Table 6

Study Characteristics

Study characteristics are shown in Table 8 and summarized below.

Sample Size. The total sample size of the studies analyzed was n= 3,203,609. Sample sizes ranged from n=128 to n=367,314, with a mean of n=291,237 and a median of n=318,392. Of these, ten studies (from the College Board) had sample sizes between n=100,000-500,000 and one (Demeter) had a far smaller sample size of n=128.

Duration. Ten studies assessed students with four or more years of arts study (the College Board data); one assessed students with two or more years of arts study four times a week (the study by Demeter). Thus we can see that the students in these studies all had nontrivial involvement in the arts.

Integrated/Separate. All studies assessed the math performance of students exposed to arts courses taught as separate disciplines.

Participant Characteristics. All 11 studies assessed students of average SES. One study assessed students at or above grade level in math (the study by Demeter).

Age. The ten College Board studies assessed high school seniors. The study by Demeter assessed 8th graders.

Outlet. The ten College Board studies were in the form of unpublished tabulated data; the study by Demeter appeared as a doctoral dissertation. Thus, again, none of the studies were in published, peer-reviewed form.

Results

Effect sizes are shown in Table 9 in the form of a stem and leaf display. Effect sizes ranged from r=.00 to r=.17, with a mean unweighted effect size of r=.10 (d=.20), and a mean weighted effect size of r=.11 (d=.22) (Stouffer's Z=189.73, p<.0001; t-test of the mean Zr=6.36, p<.0001). The 95% confidence ranged from r=.07 to r=.14. The file drawer analysis tells us that 146,313 more studies averaging a null effect would be needed to yield a barely significant Stouffer's Z of p=.05.

Table 9. Stem and Leaf Display of 11 Effect Size rs from Meta-Analysis of Correlational Math Studies

Stem	Leaf
+.3	
+.2	
+.1	2, 4, 4, 5, 7
.0	0, 5, 6, 8, 8, 9
-.0	
-.1	

The standard deviation of the mean effect size was .05, and this was shown to be significant, χ^2=5399, *df*=10, *p*<.0001. Thus, our effect sizes were significantly heterogeneous. As can be seen in Table 7, ten of the effect sizes were positive (all from the College Board data), and one was at zero (the study by Demeter). When we analyzed the College Board studies alone, effect sizes remained significantly heterogeneous, χ^2=5396, *df*=9, *p*<.0001. As with the verbal effect sizes, the effect sizes for the College Board math data rose fairly steadily over the 11-year period from 1988-1998. A linear contrast test on year of data revealed that contrast weights associated with year of data correlated with effect sizes at *r* = .96, Z=70.06, *p*< .0001. Again, this heterogeneity was not related to the fact that the last three years were based on a recentered scale because when we eliminated these years from the analysis, heterogeneity remained high, χ^2=2325, *df*=6, *p*<.0001.

Discussion

These three meta-analyses reveal a positive and significant relationship between arts education and academic outcomes, a relationship that can be generalized to new studies that might be carried out on this research question. The claim that involvement in the arts improves verbal and math achievement is consistent with—but not proven by—the positive effect sizes found here. Because the effect sizes are based on correlational studies, they do not allow us to conclude that arts education *causes* academic skills to improve. It is certainly possible that studying the arts leads to the development of cognitive skills that in turn lead to heightened achievement in academic areas. It is also possible that studying the arts leads to greater engagement in school, which in turn leads to greater academic achievement. But these studies do not allow us to rule out a causal relationship in the opposite direction: high academic achievers may choose to study the arts.

There are various plausible reasons why students who are high academic achievers to begin with might choose to study the arts: they may believe that they can afford to take fewer academic courses; they may believe that arts courses will help them to get into a selective college; they may come from families that value both academic achievement and the arts; or they may attend schools with high level arts facilities (since academically excellent schools tend to have far better arts programs than do our poorest schools). These explanations are typically ones that might be associated with socioeconomic status (SES). However, the studies by Catterall and Heath demonstrate that SES cannot be the only explanation, since in both cases the same link was found among students from low SES backgrounds. Thus, low SES students who choose to become involved in the arts may do so because their families, like high SES families, also value the arts as well as academic achievement. Or such students may attend better schools than those attended by students who are not involved in the arts.

And, in fact, a recent critical analysis of the College Board data by Elliot Eisner reveals the fallacy of assuming a special link between arts study and SAT scores.[39] Eisner compared the differential in SAT scores in 1998 between students taking only one year of arts and those taking four years of arts to the differential between students taking one vs. four years of various academic subjects: English, math, science, history/social studies, and foreign languages. What he found was striking. In all cases, students who took more of specific kinds of courses achieved higher scores. And in all cases, the score differential between one vs. four years of study was higher for the academic areas than for the arts. For example, the difference in verbal SAT scores between those with one vs. four years of arts was 40 points, while that between those with one vs. four years of English was 55 points. And the difference in math SAT scores between those with one vs. four years of arts was 23 points, while that between those with one vs. four years of science was 57 points. Thus, it appears that what is really going on is that high achievers focus and take more nonrequired courses in a particular area. As a result, we see an association between number of nonrequired courses taken in a particular area (including the arts) and SAT scores.

Eisner's comparative observations caution us against assuming a specific link between arts and SAT scores. Another piece of evidence should also caution us against assuming a specific link between arts involvement and academic success. In the study by Heath, included in our first meta-analysis, students involved in arts after school organizations (n=143) were compared to both a national sample as well as to students involved in sports after school organizations (n=31). We noted that 83% of those in the arts group and 81% of those in the sports group received some academic award and found this difference to be not significant, χ^2=.12, df=1, p>.90.[40] In contrast, the data entered into our meta-analysis were based on a comparison of those in the arts group vs. those in the national sample: those in the arts group were significantly more likely to win an award than those in the national sample, χ^2=23.87, df=1, p<.001.

The fact that both arts- and sports-involved students did better academically than the national sample underscores the problems in drawing causal conclusions from any correlation between involvement in the arts and academic achievement. While a correlation is consistent with a causal relationship, it does not demonstrate such a relationship. The evidence in this case suggests that students who are intensively involved in after-school arts are ones who are highly motivated to begin with, just as are those who are intensively involved in after-school sports. Thus, the students in Heath's sample who spend time in after school arts organizations do well in school not because of their exposure to art, but because they are the type of student who have the drive and determination to spend at least nine hours a week mastering a skill, whether this be arts or sports.

Evidence from a related study by the sociologist Paul DiMaggio[41] provides further support, we believe, for the possibility that students who are high in academic motivation to begin with choose to study the arts. DiMaggio analyzed a sample of data from about 3,000 11th graders in the U.S. in 1960. Students indicated how interested they were in a range of arts activities (including reading literature); they indicated the degree to which their self-image was one of being "cultured" (e.g., by responding to sentences such as "I am a cultured person" or "I like beautiful things"); and they rated the extent to which they were involved in the arts as makers or audience members (and here they were asked to exclude arts activities in school, with the exception of school trips). Together these ratings defined the students' degree of "cultural capital," a concept taken from the French sociologist, Pierre Bourdieu.[42] Students also reported their grades in English, History/ Social Studies, and Mathematics, and a composite measure of grades was created from these self-reports.

A regression analysis, controlling for ability level (as measured by a vocabulary test) and social class (as measured by father's education) showed that "cultural capital" was positively related to grades in each of the four subjects and to composite grades. In addition, the impact of arts involvement on grades was as great as the impact of underlying ability, with the exception of grades in mathematics. The finding that involvement in the arts predicted grades independent of social class is entirely consistent with the findings of Catterall and Heath. However, the finding that involvement in the arts predicted *grades independent of underlying ability* is distinct: none of the correlational studies that we identified controlled for ability. To the extent that we can accept a vocabulary score to be an index of academic ability (and we believe such a score to be limited because it is biased toward verbal rather than mathematical or scientific ability), we must conclude that the correlation between arts involvement and academic achievement cannot be explained by high-ability students' choosing to study the arts. However, this correlation can still be explained by an underlying dimension of academic *motivation*. Students who are motivated to succeed in school may be high-energy students who have a wide range of interests that include the arts.

DiMaggio's interpretation of these findings differs from our interpretation. He argues that teachers give students perceived to have cultural capital more attention and perceive them as more intelligent than students without such capital. DiMaggio argues that the measure of cultural capital is not simply tapping a general dimension of academic achievement motivation because the impact of cultural capital was stronger for grades in the humanities and social sciences than mathematics. However, we argue that it is just as likely that students with a high interest in the arts have high achievement motivation in general. Achievement motivation may be more likely to lead to actual higher grades in subjects with somewhat subjective grading criteria (e.g., English, History/Social Studies) than in subjects with more objective grading criteria such as mathematics. Support for this argument comes from a study cited by DiMaggio showing that achievement in Mathematics is more strongly related to years of math study than is achievement in English and Civics (suggesting that grades in math are more reflective of what students have learned than are grades in English or Civics).[43]

One fascinating finding in DiMaggio's study points up the flaw in concluding that such a correlation demonstrates that involvement in the arts causes grades to improve by some kind of learning mechanism. DiMaggio also had a measure of learning in the arts—assessing the amount of information that students had acquired about literature, music, and the visual arts. Scores on this test did not predict grades. Thus, if we can assume that the test measured what students actually learned in arts courses, we can conclude that what they learned in art did not contribute to their academic success. However, this finding is consistent with our interpretation that the kind of student who has arts involvement (or cultural capital) is also the kind of student who works hard and strives to achieve. This finding also does not rule out the possibility that success in arts courses may lead students to become more motivated in school, including in academic subjects.

Because all of these studies are correlational in design and have a self-selected sample of students who chose (for whatever reason) to immerse themselves in the arts, we cannot draw from these studies any inferences about the causal power of the arts to elevate academic achievement. The association found could be a product of self-selection. However, it is possible that both self-selection and causal factors are at work. Might it not be the case that high-achieving students self-select into the arts and then go on to develop higher academic achievement as a direct consequence of their involvement in the arts?

Fortunately, the study by Catterall allows a test of this hypothesis because it measured self-selected students over 3 points in time—8th, 10th, and 12th grade.[44] Support for this hypothesis would be gained if effect sizes increased over time. We calculated effect sizes for the all SES sample of students at these three points in time (with the outcome of composite academic achievement score) and we found no change: the effect size remained at r=.18 (d=.37) for all three years.[45] We might have expected to find a rise in effect sizes as a result of the attrition that occurred over the years (possibly the students not benefiting from the arts ceased their involvement in the arts and were no longer in the high arts sample). However, despite the fact of attrition, effect sizes did not budge. This is one piece of evidence that studying the arts does not causally affect academic achievement even for those students who initially self-selected into the arts.[46]

A comparison between the verbal and math outcome studies shows that mean weighted effect sizes were almost twice as high for verbal than for math outcomes. This finding shows that verbally achieving students are more likely to choose to study the arts than are mathematically achieving ones (or that students who study the arts are more likely to achieve in the humanities than in math).

For a test of the hypothesis that studying the arts causes academic outcomes to improve, we turn to the experimental studies.

Experimental Studies

From the experimental studies, we calculated 24 effect sizes based on verbal outcomes and 15 based on math outcomes. Some of the studies reported results separately by age group, and we thus had to calculate a separate effect size for each age group. Some of the studies reported both verbal and math outcomes, based on the same participants. Because the separate effect sizes in a meta-analysis must not be based on the same participants, we analyzed the math and verbal outcomes in two separate meta-analyses.

Experimental Studies with Verbal Outcomes

The characteristics of the experimental studies with verbal outcomes are shown in Table 10 and are summarized below:

Study Characteristics

Sample Size. The total sample size of the studies analyzed was n=19,277. Sample sizes ranged from n=11 to n=13,338, with a mean of n=803 and a median of n=197.

Design. Eleven effect sizes came from experimental-unmatched studies, 13 from experimental-matched studies. Of the matched studies, only two were true experiments in which students were randomly assigned to arts vs. control groups at the individual level.[47]

Duration. Nineteen of the effect sizes came from studies that assessed the verbal achievement of students who had studied the arts for at least one year (and of these, 2 effect sizes came from studies that assessed students who had studied the arts for two years); and 5 from studies that assessed students with less than one year's study of the arts. Thus, again, most of the effect sizes came from studies that assessed students with a nontrivial amount of arts study, but the duration of arts instruction was far briefer in the experimental than in the correlational studies. This difference in duration means that we cannot rule out the possibility that any weaker effects shown by the experimental studies are due to the fact that students in these

Table 10. 24 Experimental Studies with Verbal Outcomes: Meta-analysis 4

Study	N	r	$Z(p)$* (*<.0001)	Design	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
Aschbacher & Herman (1991)[a]	11 classes; 520 students. Classes served as the unit.	.00	0 (p=.5)	Unmatched	2 (Over a year of Humanitas program)	Yes	CTBS Language scores	Average SES	11th gr	Signif. method (p entered as 0 bec. reported ns)	Technical report
Baum & Owen (1997)	132	.12	1.40 (p=.08)	Matched	2 (2 years arts integration)	Yes	Reading scores	Average SES & Academically at risk (all arts talented)	4-6th gr	Maximum possible contrast F	Conference presentation
Brock (1991a)[b]	308	.02	.33 (p=.37)	Unmatched	1 (probably under 1 year)	Yes	ITBS Reading and Language scores	Average SES	K	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991a)	354	.02	.39 (p=.35)	Unmatched	2 (probably over 1 year)	Yes	ITBS Reading and Language scores	Average SES	1st gr	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991a)	438	.11	2.37 (p=.009)	Unmatched	2 (probably over 1 year)	Yes	ITBS Reading and Language scores	Average SES	2nd gr	Signif. Method; Reading and Language rs averaged	Technical report

Study	*N*	*r*	*Z(p)** (*<.0001)	Design	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
Brock (1991a)	438	-.07	-1.55 (p=.06)	Matched	2 (probably over 1 year)	Yes	ITBS Reading and Language scores	Average SES	3rd gr	Signif. Method; Reading and Language *rs* averaged	Technical report
Brock (1991a)	516	-.10	-2.37 (p=.009)	Matched	2 (probably over 1 year)	Yes	ITBS Reading and Language scores	Average SES	4th gr	Signif. Method; Reading and Language *rs* averaged	Technical report
Brock (1991a)	392	-.12	-2.37 (p=.009)	Matched	2 (probably over 1 year)	Yes	ITBS Reading and Language scores	Average SES	5th gr	Signif. Method; Reading and Language *rs* averaged	Technical report
Brock (1991b)[c]	384	.05	.92 (p=.18)	Matched	2 (probably over 1 year)	Yes	ITBS Reading and Language scores	Average SES	6th gr	Signif. Method; Reading and Language *rs* averaged	Technical report
Brock (1991b)	316	.00	.08 (p=.47)	Matched	2 (probably over 1 year)	Yes	ITBS Reading and Language scores	Average SES	7th gr	Signif. Method; Reading and Language *rs* averaged	Technical report
Brock (1991b)	352	.00	.05 (p=.48)	Matched	2 (probably over 1 year)	Yes	ITBS Reading and Language scores	Average SES	8th gr	Signif. Method; Reading and Language *rs* averaged	Technical report

Study	N	r	$Z(p)$* (*<.0001)	Design	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
Catterall and Waldorf (1999) [d]	13,338	.02	2.33 (p=.01)	Unmatched	2 (Probably over 1 year)	Yes	ITBS Reading scores	Low SES	6th	Signif Method	Technical report
Coakley (1995) [f]	63	.16	1.27 (p=.10)	Matched	2 (Two years arts twice a wk)	No	Peabody Picture Vocabulary scores	Low SES	Pre-school	Pretest r minus posttest r	Technical report
Dillard (1982)[g]	97	.03	.34 (p=.37)	Matched (Random assignment)	1 (7 mos, 1hr. per wk.)	No	Metropolitan Readiness scores	Average and below SES; Academically Gifted	K-3rd gr.	F-method	Doctoral dissertation
Gardiner et al. (1996)[h]	80	.10	.89 (p=.19)	Unmatched	1 (7 mos special art and music)	No	MAT Language Arts scores	Average SES	1st gr	Posttest r minus pretest r	Peer-reviewed journal
Glismann (1967)	149	-.03	-.31 (p=.38)	Matched	1 (1 yr. experimental arts and crafts)	No	Grades in English	Average SES; Slow learners	9th gr	Phi (ϕ) method	Doctoral dissertation
Hudspeth (1986)	32	.66	3.72*	Matched	1 (6 mos of curriculum integrated with music/poetry)	No	CAT Language scores	Average SES; Low achievers	4th gr	Signif. Method	Doctoral dissertation
Jackson (1979)	245	.02	-.60 (p=.27)	Matched (Random assignment)	2 (1 yr. arts integration)	Yes	McGraw-Hill CTBS Reading Scores	Average SES; Academically at risk	3rd-5th gr	t-method	Doctoral dissertation

Study	N	r	$Z(p)$* (*<.0001)	Design	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
Marston (1997)	40	-.25	2.23 (p=.01)	Unmatched	2 (Probably over 1 yr. bec. prob. not students' 1st yr.)	Yes	Stanford Achievement Test Reading scores	Low SES	3rd gr	t-method	Masters thesis
Norman (1987)	1,444	-.10	4.69*	Unmatched	2 (At least a year of arts-integrated curriculum)	Yes	ITBS Language scores	Average SES	4th gr	t-method	Doctoral dissertation
Tunks (1997)[i]	32	.29	1.64 (p=.05)	Unmatched	2 (At least a yr. of arts integrated curriculum)	Yes	ITBS Reading	Average SES	K	Signif.Method	Technical report
Tunks (1997)	32	.29	1.64 (p=.05)	Unmatched	2 (At least a yr. of arts integrated curriculum)	Yes	ITBS Reading	Average SES	1st gr	Signif.Method	Technical report
Tunks (1997)	39	.26	1.64 (p=.05)	Unmatched	2 (At least a yr of arts integrated curriculum)	Ys	ITBS Reading	Average SES	2nd gr	Signif. Method: p from Tukey test used	Technical report
Tunks (1997)	45	.00	0 (p=.50)	Matched	2 (At least a yr of arts integrated curriculum)	Yes	TAAS Reading	Average SES	6th gr	Signif. Method; Tukey reported ns; p of 0 entered.	Technical report

Notes:

a. Sample sizes for Aschbacher & Herman (1991) were estimated from their report of 11 classes with 20 students per class.

b. Sample sizes for Brock (1991a) were estimated based on number of enrolled children at each age. Sample sizes of control groups at each age were estimated as identical to arts groups because the report stated that the two groups were "approximately equal" (p.23) and because the total *N* across ages was 1144 for arts and 1097 for control participants. We assumed an arts exposure duration of under a year for kindergarten, since students had probably not been in the school a full year before being tested. We assumed an exposure of over a year for all other grades since most students tested in these grades had probably been in the school for at least a year.

c. Sample size of control groups in Brock (1991b) was estimated to be identical to arts groups because in Brock (1991a) she reported almost equal arts and control groups, and these two evaluation studies appeared to be very similar in design. Brock (1991a) was coded as unmatched for K, 1st, and 2nd grades because scores were not corrected for pretest differences, and as matched for 3rd-5th grades because scores were analyzed by ANCOVA controlling for pretest differences.

d. Catterall and Waldorf (1999) evaluated the Chicago Arts Partnerships in Education (CAPE), and we used their data. CAPE had previously been evaluated by North Central Regional Educational Laboratory (1998). We used Catterall's data because he compared CAPE schools to non-CAPE schools matched in SES.

e. We used the all SES sample from Catterall, Chapleau, & Iwanaga (1999) to render the results as comparable as possible to the College Board results. Had we used the low SES 12th grade sample, our effect size would have been r=.10, considerably smaller.

f. Although an Ancova was reported for Coakley (1995), we were not convinced that a main effect of Group controlled for pretest differences. We felt we would derive a more accurate effect size if we used the method of subtracting pretest r from posttest r.

g. Dillard (1982) reported results separately for 4 age groups. Since the report supplied the raw scores, we recomputed a repeated measures ANOVA (Group x Test [pretest, posttest]), and calculated the r from the F given for the interaction between Group and Test, showing that the arts group improved somewhat more than the non-arts group.

h. For Gardiner et al. (1996) we computed a pretest and posttest r, and then subtracted the pretest r from the posttest r. We did this because the authors had not taken pretest differences into account in reporting their findings.

i. In calculating effect sizes for Tunks (1997), we compared only the arts group (which received arts integrated into the curriculum) to the full control group. We did not use the modified control group (which received arts but not integrated into the curriculum) so as to make these results comparable to our other studies, most of which compared an arts group to a no treatment control group. The sample size for this study was estimated from *df* within groups; *n*s were assumed to be as equal as the *N* permitted; where a choice between 2 *N*s differing by 1 was needed, the lower *N* was entered. Finally, only the 6th grade from Tunks (1997) was coded as matched in design because this was the only sample for which pretest differences were controlled.

studies were exposed to the arts for less time than were students in the correlational studies.

Integration/Separate. Nineteen (79%) of the effect sizes came from studies that assessed the effects of studying the arts integrated with academics, five (21%) from studies that assessed the effects of studying the arts as separate disciplines. Thus, here is another way in which the experimental studies differed from the correlational ones: none of the correlational studies assessed the effects of integrated arts and academics. Thus we might expect the experimental studies to have stronger effect sizes than the correlational ones, assuming that when the arts are integrated with academics they are more likely to lead to improved academic performance than when they are taught as separate disciplines.

Participant Characteristics. Three effect sizes came from studies that assessed students of low SES, 20 were from studies that assessed students of average SES, and one assessed a mixture of both. Four effect sizes came from studies that assessed academically at-risk students, one from a study of academically gifted students, and one from a study of students with talent in art; the rest came from studies assessing students not selected for any kind of academic ability. Thus we can see that students of widely different ability levels in academics were included; but in only one study were students selected for artistic talent (the study by Baum).

Age of Participants. Ages ranged from preschool to 11th grade, but most of the effect sizes came from studies of students at the elementary school level.

Outcome. In two of our verbal outcome studies, two relevant outcomes for the same participants were reported—a reading and a language score.[48] In this case, we calculated the effect size for each, and then averaged the standardized *Zr* associated with each to yield one composite effect size.[49] A few studies also reported writing outcomes, but these were not included because there were so few.

Outlet. Only one of the studies appeared in published form, and it appeared in a peer-reviewed journal. (This was the study by Gardiner and his colleagues, which appeared as a letter in *Nature*.) The remaining studies appeared as technical reports written by evaluators contracted to evaluate a particular program, doctoral dissertations, and conference presentations.

Results

Effect sizes, shown in Table 11, in the form of a stem and leaf display, ranged from -.25 to .66. The average unweighted effect size was r=.07 (d=.14); weighting by effect size reduced the effect to r=.01 (d=.02). The Stouffer's Z= 3.82 was significant, p<.0001, allowing us to generalize the results to new subjects in the same studies. However, the more important test, the t-test of the mean Zr, yielded 1.66, which was not significant, p=.11. The lack

Table 11. Stem and Leaf Display of 24 Effect Size *r*s from Meta-Analysis of Experimental Verbal Studies

Stem	Leaf
+.6	6
+.5	
+.4	
+.3	
+.2	6, 9, 9
+.1	0, 1, 2, 6
.0	0, 0, 0, 0, 2, 2, 2, 2, 3, 5
-.0	3, 7
-.1	0, 0, 2
-.2	5

of significance of this test shows that we cannot confidently generalize our findings to new though similar studies on this research question. In addition, the 95% confidence interval for unweighted effect sizes spanned zero: the interval ranged from $r= -.01$ to $r=.14$. Thus, the mean effect size of a new set of similar studies might well be at zero. A file drawer analysis revealed that 106 more studies averaging null effects would be needed to reduce the Stouffer's Z to $p=.05$.

Two studies were assigned effect sizes of 0 because the authors had reported only that the results were not significant. Because this is a conservative estimate, likely to underestimate effect size, we performed the meta-analysis again without these two studies. The mean weighted *r* remained unchanged.

The mean effect size found here of $r =.07$ is very small; the weighted mean effect size of $r = .02$ is essentially a zero effect size. In addition, we cannot generalize these findings to new studies, and the mean effect size of a new set of studies could well be zero. Clearly we must conclude that the studies identified here offer no evidence for any educationally significant impact of arts education on verbal achievement.

The standard deviation of the mean effect size was .18; effect sizes were significantly heterogeneous: $\chi^2=75.42$, $df=23$, $p<.0001$. We thus performed linear contrast tests in order to try to account for the heterogeneity.

Matched vs. Unmatched Studies

We performed a linear contrast test to assess the hypothesis that effect sizes from the unmatched experimental studies are higher than those from matched studies (because subjects in the control group might have been higher in academic achievement to begin with). Contrast weights correlated with effect sizes in the opposite of the predicted direction at $r = -.14$, Z =

-1.08, which was not significant, p=.54. Thus, matching of groups at pretest was not associated with size of effect.

We were unfortunately not able to compare true experimental studies (with random assignment) to quasi-experimental ones (in which intact classrooms are compared) because there were only two true-experimental studies. Random assignment is most likely to lead to equivalent groups and provides the most reliable kind of test of an experimental hypothesis. It is perhaps noteworthy that these two studies yielded negligible effect sizes of r=.03 and r=.02.[50]

Age of Student

We tested the possibility that age of student predicted size of effect, with studies using younger students producing higher effect sizes. A significant finding would demonstrate that the arts are most effective at younger ages. However, contrast weights correlated with effect sizes at r = -.24. This result, in the opposite of the predicted direction, was not significant, Z=-1.03, p=.15. Thus there was no evidence that younger children are more helped (in terms of verbal scores) by the arts than are older children.

Arts-Integrated vs. Arts Separate Instruction

All but four of the effect sizes came from studies that assessed the effects of an arts-integrated curriculum (i.e., the arts were integrated with academics as well as taught separately). We performed a linear contrast in order to test the prediction that effect sizes from studies in which the arts were integrated with academics would be higher than those from studies in which the arts were taught only as separate disciplines. Contrary to prediction, results were not significant. Contrast weights correlated with effect sizes at r = -.01, Z=.01, p =.48.

In one study, the experimental group was compared to a second experimental group that received the same arts exposure but without the arts being integrated into the academics.[51,52] In the 1st and 6th grades, there was no difference between these two groups; in the 3rd grade, the arts-integrated group performed significantly better. However, the integration component was confounded with three other relevant factors: in the integrated (but not the nonintegrated) group, the teachers had volunteered to participate, had chosen the art programs to which the students would be exposed, and were aware that they were in an experimental art program. Thus, we cannot conclude that the advantage associated with the integrated group in 3rd grade is due to the integrated nature of the arts programming.

We have no evidence from these studies that verbal academic improvement is more likely to occur when the arts are integrated with academics rather than taught separately.

Experimental Studies with Math Outcomes

The characteristics of the experimental studies with math outcomes are shown in Table 12 and summarized below:

Study Characteristics

Sample Size. The total sample size of the studies analyzed was n=18,736. Sample sizes ranged from n=80 to n=13,338, with a mean of n=1249 and a median of n =354.

Design. Ten of the effect sizes came from studies that compared matched treatment and control groups; 5 were from unmatched comparisons.

Duration. Twelve effect sizes came from studies that assessed students who had had at least one year of arts study (of these, one was from a study that assessed students with two years of arts study); three came from studies that assessed students with under a year of arts study. Thus, as with the verbal experimental studies, students in these studies usually have a nontrivial amount of exposure to arts classes but have far less than in the correlational studies.

Integration/Separate. Thirteen effect sizes came from studies that assessed the effects of studying the arts when integrated with academics; two from studies that assessed the effects of studying the arts as separate disciplines. Thus, as with the verbal experimental studies, we might predict higher effect sizes here compared to the correlational studies if the arts are more likely to lead to transfer when integrated with an academic subject.

Participant Characteristics. All but one of the effect sizes came from studies that assessed students classified as average in SES; one assessed low SES students. In addition, one effect size came from a study that assessed students classified as slow learners, and one from a study that contained a group of students classified as academically at-risk.

Age of Participants. Ages ranged from kindergartners to 9^{th} graders, with 80% of the studies assessing students in grades K-6.

Outlet. Only one of the studies appeared in published form, and this study appeared in a peer-reviewed journal. The remaining studies appeared as technical reports written by evaluators contracted to evaluate a particular program, doctoral dissertations, or conference presentations.

Results

Effect sizes, shown in a stem and leaf display in Table 13, ranged from r=-.14 to r=.34. The average unweighted effect size was r=.06 (d=.12); the mean weighted effect size was r=.02 (d=.04). The Stouffer's Z= 3.10 was significant, p=.001. However, the more conservative t-test of the mean Zr yielded a value of 1.63 which was not significant, p=.13. Thus, while we can

Table 12. 15 Experimental Studies with Math Outcomes: Meta-analysis 5

Study	N	r	$Z(p)$* (*<.0001)	Design	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
Baum & Owen (1997)	90	-.14	-.87 (p=.19)	Matched	2 (2 yrs. arts integration)	Yes	Math scores	Grp 1: Average SES; art talented Grp 2: Academically at risk; art talented	4-6th gr	Maximum possible contrast F	Conference presentation
Brock (1991a)	308	.13	2.37 (p=.009)	Unmatched	1 (probably under 1 yr.)	Yes	ITBS Math scores	Average SES	K	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991a)	354	.11	2.05 (p=.02)	Unmatched	2 (probably over 1 yr.)	Yes	ITBS Math scores	Average SES	1st gr	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991a)	438	.11	2.37 (p=.009)	Unmatched	2 (probably under 1 yr.)	Yes	ITBS Math scores	Average SES	2nd gr	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991a)	438	-.02	-.44 (p=.33)	Matched	2 (probably over 1 yr.)	Yes	ITBS Math scores	Average SES	3rd gr	Signif. Method; Reading and Language rs averaged	Technical report

Study	N	r	$Z(p)$* (*<.0001)	Design	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
Brock (1991a)	516	-.06	-1.34 (p=.09)	Matched	2 (probably over 1 yr.)	Yes	ITBS Math scores	Average SES	4th gr	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991a)	392	-.09	-1.75 (p=.04)	Matched	2 (probably over 1 yr.)	Yes	ITBS Math scores	Average SES	5th gr	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991b)	384	-.05	-.92 (p=.18)	Matched	2 (probably over 1 yr.)	Yes	ITBS Math scores	Average SES	6th gr	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991b)	316	.13	2.37 (p=.009)	Matched	2 (probably over 1 yr.)	Yes	ITBS Math scores	Average SES	7th gr	Signif. Method; Reading and Language rs averaged	Technical report
Brock (1991b)	352	.11	2.05 (p=.02)	Matched	2 (probably over 1 yr.)	Yes	ITBS Math scores	Average SES	8th gr	Signif. Method; Reading and Language rs averaged	Technical report
Catterall and Waldorf (1999)	13,338	.02	-2.33 (p=.01)	Unmatched	2 (probably over 1 yr.)	Yes	ITBS Math scores	Low SES	6th gr	Signif. Method	Technical report
Gardiner et al. (1996) [a]	80	.34	3.03 (p=.001)	Matched	1 (7 mos. special art and music)	No	MAT Language Arts scores	Average SES	1st gr	Posttest r minus pretest r	Peer-reviewed journal

Study	N	r	$Z(p)$* (*<.0001)	Design	Duration of Arts Exposure: 1=under a year; 2=at least a year.	Integ	Outcome	Participant Characteristics	Age	Effect Size Method	Outlet
Glismann (1967)	149	.33	4.01*	Matched	2 (1 yr experimental arts and crafts)	No	Grades in Math	Average SES; slow learners	9th gr	Phi (ϕ) method	Doctoral dissertation
Luftig (1993)[b]	137	.00	0 (p=.5)	Matched	1 (1 yr. arts integrated SPECTRA curriculum)	Yes	Math scores	Average SES	5th gr	Signif. Method; p of 0 entered because study reports ns.	Technical report
Norman (1987)	1,444	-.05	1.41 (p=.08)	Unmatched	2 (at least a year of arts-integrated curriculum)	Yes	ITBS Math scores	Average SES	4th gr	t-method	Doctoral dissertation

Notes:

a.. Sample sizes for Gardiner et al. (1996) were estimated from their report that pretests were available for 83% of the total sample of 96 (i.e., 80 students); we assumed equal n in arts and control groups because each group was composed of 4 classrooms. It is likely that the ns were not perfectly equal but were fairly close.

b. In calculating the effect size for math from Luftig (1993), we compared the arts group to the no-treatment control group. We did not use the modified control group, which received a progressive education curriculum unrelated to the arts, so that this study would be comparable to the others in this meta-analysis, all of which compared an arts group to a no-treatment control. We were not able to use the second year study, Luftig (1994), because sufficient data from this study was not made available to us. Luftig (1993) included measures of both math and reading. However, reading outcomes could not be used because the control group declined from pretest to posttest, while the art group remained the same. Thus the reported significant effect was due to the control group declining rather than to the arts group improving. This violates the principle that control groups should not decline in ability. It is worthy of note that this study is widely cited as demonstrating the power of the arts to elevate academic achievement. However, the reading data do not allow this conclusion, for the reason just given, and the math data yield an effect size of 0. The effect size of 0 was calculated from the report that no differences were found in Total Math between the SPECTRA+ and full control group in improvement from pre- to post-test.

Table 13. Stem and Leaf Display of 15 Effect Size rs from Meta-Analysis of Experimental Math Studies

Stem	Leaf
+.6	
+.5	
+.4	
+.3	3, 4
+.2	
+.1	1, 1, 1, 3, 3
.0	0, 2
-.0	2, 5, 5, 6, 9
-.1	4
-.2	

generalize our findings to new subjects in the same studies, we cannot generalize our findings to a similar set of new studies on this research question. In addition, the 95% confidence interval spanned zero and ranged from r=-.02 to r=.14. Thus, the mean effect size of a new set of similar studies might well be at zero. A file drawer analysis revealed that these results are only minimally resistant to a file drawer effect: if 38 more studies averaging null effects were found, the Stouffer's Z would be reduced to the barely significant level of $p = .05$.

One of the effect sizes was estimated to be zero because the authors had reported finding no significant results.[53] This is a conservative estimate (since a nonsignificant result may still result in a positive effect size), and may have resulted in too low an overall effect size. We thus repeated the analysis with this effect size excluded. However, this analysis had no effect on the weighted mean effect size (and altered the mean effect size only by .01, yielding r=.07).

Our mean effect size of .06 and our weighted mean effect size of .02 are very small effects. Thus far we must conclude that our studies provide no evidence for any educationally significant impact of arts education on math achievement.

The standard deviation of the mean effect size was quite large (.14); not surprisingly, effect sizes were significantly heterogeneous, χ^2=59.95, df=14, p<.0001. We therefore conducted linear contrast tests to try to account for the heterogeneity.

Matched vs. Unmatched

We conducted a linear contrast analysis to determine whether studies with groups matched at pretest produced smaller effect sizes than those without such matching. The contrast weights correlated with effect sizes at $r = .04$. This was not significant, Z=.28, p=.50. Thus, unmatched studies did not produce larger effect sizes than the more rigorously designed matched studies.

This suggests that the unmatched experimental studies did a good job of matching without using pretests, by locating comparison groups from similar schools and SES.

Age of Student

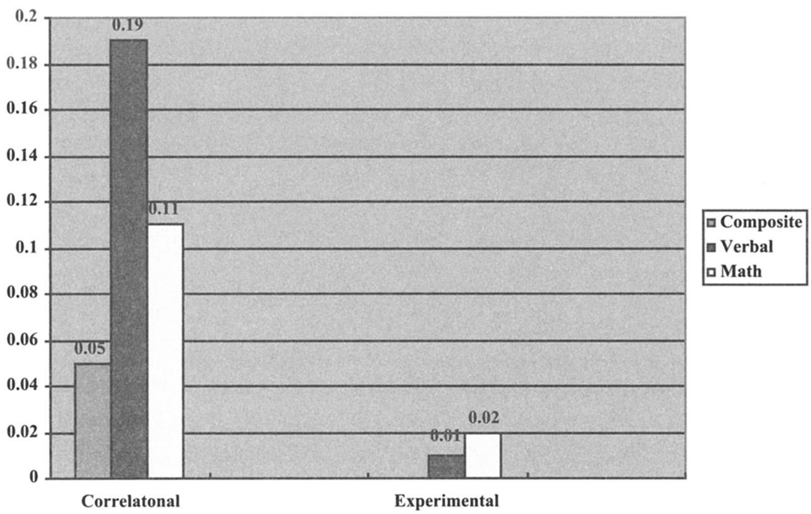
We again tested the possibility that age of student predicted size of effect, with studies using younger students producing higher effect sizes. However, contrast weights correlated with effect sizes in the opposite of the predicted direction, at $r = -.05$. This was not significant, $Z = -.34$, $p=.32$. Thus, younger students were not helped academically more than older students by the arts.

We could not test the hypothesis that effect size was related to arts integration, since all but two of the studies assessed the effects of an arts-integrated curriculum. We were unable to find a factor or factors that accounted for the heterogeneity in this group of studies.

Discussion

Figure 1 displays comparative findings across the five meta-analyses. This figure shows graphically the higher weighted effect size *r*s for the correlational than the experimental studies, and also shows that the strongest correlation was found for verbal outcomes.

Mean Weighted Effect Size *rs* in Correlational vs. Experimental Studies, by Outcome



Note:
There were no experimental studies with composite outcomes.

A comparison between our correlational and experimental findings shows that it is the correlational, self-selection studies that underlie the oft-reported relationship between studying the arts and academic achievement. However, a word of caution is necessary. Many of our correlational studies were ones in which students were exposed to the arts for four years,[54] while most of the experimental ones assessed exposures ranging from 6 months to 2 years. Further studies are needed if we are to disentangle the effects of at least 4 years of arts exposure from the effects of correlational, self-selected designs. It is possible that there is a threshold effect: students may need a certain amount of arts education for transfer to occur; with less than that criterial amount, no change may occur.

Relevant Studies Not Included in Meta-analyses

We identified 27 studies (summarized in Table 14) that could not be included in the meta-analysis because an effect size r could not be calculated from the data provided, or because the study lacked a control group. Most of these studies were correlational ones showing that students who attend a school in which the arts are central are high academic achievers. Without a demonstration that these students improved from a pretest prior to entering this school, to a posttest after having been at the school, we cannot determine whether these students do well because they are a self-selected sample, or whether they do well because of the type of education they have received.

Despite the problems with these studies, we have classified each one as demonstrating a positive association between academic achievement and an arts program, or as failing to demonstrate such an association, either because the results were inconsistent or null. These studies are listed by direction of outcome and summarized in Table 12.[55] Seventeen of these studies reported a positive outcome; ten reported null or mixed findings. Thus we can conclude that the majority of such studies show that arts education is associated with higher academic achievement. A binomial probability test demonstrates that the chances of finding 17 positive studies out of a total of 27 is associated with a probability level of p=.12, and thus does not achieve statistical significance at p<.05.

Two of the studies reporting negative findings deserve special comment, as they highlight the likely role that culture plays in interpreting a correlational finding between studying arts and academic achievement. A study recently completed in the Netherlands demonstrated that Dutch students who take arts electives in high school to prepare for a national examination that includes the arts attain the same ultimate educational level as those who take no arts electives.[56] This analysis controlled for the effects of the home environment, parents' educational level, and participation in cultural

Table 14: Relevant Studies Not Included in Meta- Analyses Classified by Direction of Outcome

	Outcome		Reason for Exclusion	Outlet
Study	Positive	Mixed, Null, or Negative		
Annenberg Challenge (1998)	Many schools in which arts are integrated into the curriculum report improvements in test scores in reading, writing, and math.		No control group Insufficient data	Non peer-reviewed journal
Brock (1991c); Newbill (1992)		Students at Paseo Academy of Fine and Performing Arts Magnet High School scored above district but not state norms. In math, magnet students performed equivalent to district and below state norms. Because no pretests were given we cannot determine whether test performance is a function of the type of student attending, or a function of the program.	No control group Insufficient data	Technical report
Catterall (1995)		After 2 years in "Different Ways of Knowing" program with arts and humanities integrated, students scored higher in language arts than those with 1 year, who in turn scored higher than those with 0 years. Program participation had no effect on reading or math scores. Because students received additional training besides the arts, including critical thinking, arts intervention was confounded with other academic interventions.	Insufficient data	Technical report
Chapman (1998)	After exposure to Discipline Based Arts Education, reading, writing, and math scores rose at Shady Brook Elementary School in Bedford, Texas. The rise in scores could be due to the fact that the school district had realigned its curriculum with the content of state testing.		No control group Insufficient data	Secondary source

	Outcome		Reason for Exclusion	Outlet
Study	Positive	Mixed, Null, or Negative		
Dept. of Test Development and Administration, Prince George's County Public Schools (1997); see also Maryland Alliance for Arts Education (1995)	After instituting an arts-integrated curriculum, three schools in Prince George's County, Maryland, reported rise in test scores.		No control group Insufficient data	Unpublished tabulated data
DiMaggio (1982)	A positive relationship was found between high self-reported interest in the arts and grades in English, History/Social Studies and Mathematics for US high school students assessed in 1960.		Type of data analysis did not allow us to compute an effect size *r*.	Peer-reviewed journal
Fowler (1979b)	After 10 years of arts integration, reading scores in all elementary schools in Mesa, Arizona, doubled.		No control group Insufficient data	Secondary source
Fowler (1979b)	Reading scores for grades 4-6 rose at Mosswood Mini School, Oakland, California, after arts integration.		No control group Insufficient data	Secondary source
Getz (1984), Hoffa (1979) Lawton (1987)		At five sites across the country, arts were integrated into the curriculum as part of Project IMPACT. Some schools reported gains in scores; others reported declines.	No control group Insufficient data	Secondary source
Goldberg (1998)	Reading, math, and language scores on Metropolitan Achievement Test rose after one year in SUAVE classes in which arts were integrated, in comparison to classes without SUAVE.		Insufficient data	Technical report

	Outcome		Reason for Exclusion	Outlet
Study	Positive	Mixed, Null, or Negative		
Haanstra (1999)		Students in Holland who take arts electives in high school to prepare for a national examination that includes the arts attain the same ultimate educational level as those who take no arts electives. This analysis controlled for the effects of the home environment, parents' educational level, and participation in cultural activities prior to taking the electives.	Type of data analysis did not allow us to compute an effect size *r*.	Unpublished paper
Harland, Kinder, Haynes, & Schagen, (1998).		A negative relationship was found between degree of concentration in the arts and performance on national English examinations for high school students in the UK. Authors explain by suggesting that in UK poor students are counseled into the arts.	Insufficient data to compute effect size *r*.	Technical report
Lardo (1982)		After a year of taking many arts courses at the Rogers Magnet Middle School for the Creative and Performing Arts in Pittsburgh, no improvement in 7^{th}-8^{th} grade scores was found. However, pretesting was administered after students had been in the magnet school from 6 months to 2.5 years. Thus study could not determine whether the program resulted in higher test scores prior to the pretest.	No control group	Doctoral dissertation
McGuire (1982)		Third graders receiving an arts integrated curriculum did not achieve higher verbal scores than either mean third grade score in district or third graders from previous year.	Insufficient data	Masters thesis
Missouri Arts Education Task Force (1990), as cited in Murfee (1993)	At the Guggenheim Elementary School, Chicago, scores in reading and math rose after arts were integrated into curriculum.		No control group Insufficient data	Secondary source

	Outcome		Reason for Exclusion	Outlet
Study	Positive	Mixed, Null, or Negative		
Redcliffe School Performance Profile (1990-1994); see also Edmunds (1991; Slay & Pendergast (1993); and Welch & Greene (1995)	After institution of an arts integrated curriculum, scores rose at the Redcliffe School in South Carolina. Unfortunately the pretest was given a year after the program began. Thus we could not know whether students entered with high performance or whether their performance is a result of the program.		No control group	Unpublished tabulated data
Redfield (1990)	After 8-16 week artist-in-residence program, grades of elementary students improved, $p<.01$. Students gained in all areas of effort ($p<.01$) except math and gained in work habits in many areas.		No control group	Technical report
Rombokas, Heritage, & West (1995)		Students with some extra-curricular arts receive same grades as students in sports, as well as students in the total sample	Students in arts may have overlapped with students in sports; students in arts were compared to the total sample, and the total sample thus included students in the arts.	Technical report
Ross (1990)		K-9th graders exposed to Arts Partners program (arts integrated curriculum) did not improve in oral language (grammar and expression) relative to students in same schools not in program.	Pretest given after students had been in program for at least a year. Thus no valid pretest.	Technical report
Spectra Rhode Island (1998)	Students at arts integrated Eliot Elementary School, Needham, Massachusetts, rank in top 5% of states elementary schools. But we cannot know whether this is due to type of student who attends or to type of program offered.		No control group Insufficient data	Unpublished secondary source

	Outcome		Reason for Exclusion	Outlet
Study	Positive	Mixed, Null, or Negative		
Spilke (1991)	Students in K-11 in Fine Arts Core Education (F.A.C.E.) school in Montreal, Canada, were more likely to pass Canadian Ministry exams than other students in Montreal, and also score above national norms on college aptitude tests. However, we cannot know whether this is due to the type of student who attends, or whether it is a function of the role of the arts at the school.		No control group Insufficient data	Technical report
Spilke (1991)	Scores on math and reading tests rose after arts were integrated into curriculum at Roosevelt Middle School, Milwaukee.		No control group Insufficient data	Secondary source
Spilke (1991)	Students at the Elm Creative Arts Elementary School in Milwaukee score above city and national averages. However, without pre and posttest measures we cannot know whether scores rose in conjunction with the integration of the arts, or whether students who attend are simply high performing students.		No control group Insufficient data	Secondary source
Spilke (1991)	Students at arts-integrated Aiken Elementary School, Aiken, South Carolina, score above district and state norms.		No control group Insufficient data	Secondary source
Spilke (1991)	Students at arts magnet Ashley River Elementary School, Charleston, South Carolina, score above district and state norms. However, students who enter first grade are already performing above average; thus we cannot attribute their high scores to the arts.		No control group Insufficient data	Secondary source
Spilke (1991)		Three years after the Karl Kellogg elementary school in Chula Vista, California, became an arts magnet school, scores rose in grades 3-5 but not in grade 6.	No control group Insufficient data	Secondary source
Walker (1995)	After 18 weeks of arts-oriented after school, 45% of students increased their grade point average (most were at risk academically).		No control group	Conference presentation

activities prior to taking the arts electives. Thus, studying the arts proved unrelated to academic success. In Britain, a similar study showed that studying the arts predicted poorer academic performance.[57] Researchers studied 27,607 British high school students who concentrated in the arts. They found a significant *negative* correlation between number of arts courses taken and performance on national exams taken at the end of secondary school.

These two results stand in direct contrast to those of the College Board (where students who take four years of arts score higher on their SATs than students who take none), as well as to the results of James Catterall and his colleagues [58] and those of Shirley Brice Heath,[59] both of whom found that high arts involvement in the United States is associated with high academic achievement. Why might this be? The authors of the British study suggest that in Britain, students poor in academics are often steered into the arts. This would account for the negative association found. Perhaps in the U.S. we steer poor students into remedial academic work rather than into the arts. Such cultural differences could account for the contrasting U.S. and U.K. findings.

Similarly, the author of the study conducted in Holland notes that the general feeling in Holland is that studying the arts influences school achievement in a negative way (by taking time away from more "academic" subjects).[60] And Dutch universities take the view that since universities do not teach the arts, students should not be obligated to study the arts in high school. Since American universities (especially the most selective) often expect students to have well-rounded application materials that include excellence in something besides academics (such as arts or sports or leadership), American students who want to get in to the best universities may choose to study the arts. This could account for the correlation found between arts study and academic achievement in high school. European students may feel their best chances for admission to university lie in an exclusive focus on traditional academic subjects.

The negative findings from Britain and the neutral findings from Holland caution against drawing causal conclusions from correlations. In the U.S., we are all too apt to assume that it is the taking of the arts courses that causes the high SAT scores. Perhaps if the U.S. findings had been negative, American researchers would have leaned toward a correlational explanation, as the British researchers attempted to do.[61]

Studies with Motivational Outcomes Not Included in Meta-Analysis

If studying the arts can lead to academic improvement by means of a motivational mechanism, the kinds of outcomes one might look for would be improved attitude toward school or improved academic self-concept. We found a number of studies that assessed the effects of arts study on such motivational outcomes. Unfortunately, these studies did not also examine academic outcomes. Ideally studies should examine two links: that between arts and motivation, and then a link between the motivation found and academic improvement.

We identified 23 results assessing such motivational indicators, indicators that might be associated with grades and scores: academic self-concept, attendance, boredom in school, college aspirations, involvement in community service, dropping out of school, election to class office, participation in math/science fairs, reading, and television watching. (Note that these outcomes have not yet been demonstrated to be associated with improved academic achievement because the relevant studies have not been done. It is certainly reasonable to hypothesize, however, that they are associated with academic achievement.) Most of these studies did not provide enough information for meta-analysis. In addition, there were too few studies of each outcome to be meta-analyzed confidently.

As can be seen in Table 15, 21 of these studies reported positive results; only two reported null results (one for the outcome of attendance, one for dropping out of college).[62] All other studies reported that these indirect indicators improved in association with studying the arts. Thus, the overwhelming majority of studies showing indirect academic effects are in the positive direction. These kinds of indirect effects are those on which a motivational explanation of academic improvement from arts study might rest.

General Discussion

Our five meta-anlyses converge to demonstrate that a positive relationship between studying the arts and academic achievement does exist. However, thus far we have no evidence to suggest a causal relationship between arts study and verbal or math achievement. The weighted mean effect size for the verbal outcome experimental studies was only r=.01; the weighted mean effect size for the math outcome experimental studies was only r=.02.

The finding that students in the U.S. who self-select into the arts are high academic achievers was shown by our five meta-analyses to be an extremely reliable finding.[63] Two of our studies showed that even among poor, at-risk students, those who choose to study the arts do far better in school than those who do not choose to study the arts.[64] This demonstrated relationship, however, was tempered by the study from Britain showing the reverse relationship: while positive links were found between participation in individual art forms and performance on specific kinds of exams, this study reported that students who took a great deal of arts courses (arts concentrators) actually performed worse on their academic exams than those who did not

Table 15: Studies Assessing Relationship between Arts and Motivational Indicators of Academic Achievement

Study	Motivational Indicator	Outcome: Positive	Outcome: Null or Negative	Confounds/ Limitations	Outlet
Burton, Horowitz, & Abeles (1999)	Academic self-concept	41% of high arts students scored in top quartile of academic self-concept, compared to 18% in low arts group, no p value reported.			Technical report
Heath (1998a,b)	Academic self-concept	High arts students more likely to feel they can do things as well as others than do students in a national sample (89% vs. 76%).		Self-selected sample (i.e., students self-selected into arts) Self-report	Non-peer reviewed journal
Aschbacher & Herman (1991)	Attendance	Arts involved students (in Humanitas curriculum linking social studies, literature, and arts) had higher attendance than control group, p=.07.		Self-selected sample	Technical report
Glissman (1967)	Attendance		9th grade slow learners did not improve attendance when given arts classes, p=n.s.		Doctoral dissertation
Fowler (1979b)	Attendance	Attendance rose at Mosswood Mini School, Oakland, Calif, after school incorporated arts.		No data given	Secondary source
Heath (1998a,b)	Attendance	High arts students 3 times more likely to win award for attendance, compared to national sample.		Self-selected sample Self-report	Non-peer reviewed journal
Kantrowitz (1997)	Attendance	Attendance rose at Charles R. Bugg Elementary School after school incorporated arts.		No data given	Secondary source
Murfee (1993) Missouri Arts Education Task Force (1990)	Attendance	Attendance rose at Guggenheim Elementary School, Chicago, after school incorporated arts.		No data given	Secondary source

Study	Motivational Indicator	Outcome Positive	Null or Negative	Confounds/ Limitations	Outlet
Spilke (1991)	Attendance	Attendance rose at Roosevelt Middle School for the Arts, Milwaukee, after school incorporated arts.		No data given	Secondary source
Catterall (1998)	Boredom in school	48.9% low arts 8th graders (all SES) reported being bored in school over half or most of the time, compared to 42.2% high arts 8th graders. For lowest SES 8th graders, 46% low arts vs. 41% high arts reported boredom.		Self-selected sample Self-report	Nonpeer-reviewed journal
Aschbacher & Herman (1991)	College aspirations	Humanitas students more likely to plan to attend 4 year college than control group (71% vs. 55%), and less likely to plan to attend 2 year college (17% vs. 25%) or no college (7% vs. 16%), $p<.05$.		Self-selected sample Self-report	Technical report
Heath (1998, personal communication)	College aspirations	Students in afterschool arts organizations more likely to plan to go to college compared to national sample (86% vs. 65%)		Self-selected sample Self-report	Nonpeer-reviewed journal
Spady (1971)	College aspirations	Boys with high school arts 9.4% more likely to have college aspirations than those without arts.		Self-selected sample. Self report. Arts group included students involved in student newspaper or yearbook. Any advantage of arts group could be due to members also involved in publications.	Peer-reviewed journal

Study	Motivational Indicator	Outcome Positive	Null or Negative	Confounds/ Limitations	Outlet
Catterall (1998)	Community service	Among average 10^{th} graders, 65.2% of high arts students rarely performed community service, vs. 86% low arts. Among low SES 8^{th} graders, 74.5% high arts rarely/never served, vs. 83.2% low arts. Among low SES 10^{th} graders, 65.2% of high arts 10^{th} graders rarely served, vs. 86% low arts.		Self-selected sample Self-report	Nonpeer-reviewed journal
Heath (1998a,b)	Community service	Students in afterschool arts organizations more likely perform community service than students in a national sample.(30% vs. 6%).		Self-selected sample Self-report After-school arts organizations attended by high arts students all stressed community service.	Nonpeer-reviewed journal
Aschbacher & Herman (1991)	Dropping out of high school	Academically at-risk students in Humanitas program less likely to drop out than those not involved, 1% vs. 7%, $p<.05$.		Self-selected sample	Technical report
Center for Music Research, Florida Dept. of Education (1990)	Dropping out of high school	30 out of 36 academically at risk students said arts courses affected their decision to stay in school.			Technical report
Mahoney & Cairns (1997)	Dropping out of middle or high school	Academically and SES at-risk middle school students more likely to drop out if no involvement in arts (40% vs. 35%, $p>.10$, n.s.). Academically and SES at-risk high school students more likely to drop out if no involvement in arts (27% vs. 7%, $p=.08$).		Self-selected sample. Involvement in athletics or vocational training far more predictive of school retention than involvement in arts.	Peer-reviewed journal

		Outcome			
Study	Motivational Indicator	Positive	Null or Negative	Confounds/ Limitations	Outlet
Spady (1971)	Dropping out of college		Average students involved in arts in high school 9% less likely to remain in college over a year than students without high school arts. When college aspirations controlled, arts involved students 4.3% less likely to remain in college over a year than students without high school arts.	Self-selected sample	Peer-reviewed journal
Heath (1998a,b)	Election to class office	Students in afterschool arts organizations 3 times more likely to win election to class office than students in a national sample.		Self-selected sample Self-report	Nonpeer-reviewed journal
Heath (1998a,b)	Participation in math/ science fairs	Students in afterschool arts organizations four times more likely participate in a math/science fair compared to a national sample.		Self-selected sample Self-report	Nonpeer-reviewed journal
Heath (1998a,b), personal communication	Read for pleasure	Students in afterschool arts organizations more likely to read for pleasure than national sample (57% vs. 35%).		Self-selected sample Self-report	Nonpeer-reviewed journal
Catterall (1998)	Television watching	Among 10th graders, 28.2% high arts students watch 1 hr. or less TV per day vs.15.5% low arts students, who watch more. Among low SES 10th graders, 16.4% high arts students watch 1 hr or less, vs. 13.3% low arts students.		Self-selected sample Self-report	Nonpeer-reviewed journal

concentrate in the arts.[65] This finding suggests that cultural factors are at work here: in the U.S., we guide poor students into remedial courses, not the arts, but in the U.K. , nonacademically strong students may be guided into the arts. In addition, in the U.S., the most selective colleges expect students to have a wide range of accomplishments, and accomplishments in an art form improve a student's chances of admission to the most selective colleges. Perhaps the reason that the relationship between studying arts and College Board scores has risen steadily since 1988 is that students perceive colleges to have become even more selective, and hence feel increasing pressure to demonstrate a skill outside of the traditional academic skills in order to achieve a "competitive edge." The perception that one's chances for admission to the best universities would be helped by demonstrating mastery of an art form seems to be less the case in the U.K. Thus, academically strong students in Britain may not spend time on the arts, but instead may focus all their energies on academic coursework.[66] The same issues are raised by the Dutch study which found no relationship between arts courses and academic performance.[67] In any case, the contrast between the American findings, on the one hand, and the British and Dutch findings, on the other hand, make it clear that no causal conclusions can be drawn from any association between arts and achievement.

Why have the experimental studies been unable thus far to demonstrate that studying the arts leads to improved academic performance? We suggest two possible reasons. First, most of the experimental studies examine the effects of a year or so of arts study. In contrast, most of the correlational studies examine the effects of at least several years of studying the arts. For instance, the College Board findings show that studying the arts for four years in high school is associated with higher SAT scores, compared to students who study no arts. And Catterall and his colleagues showed high academic performance in students who had remained immersed in the arts from 8^{th} to 12^{th} grade. One can hardly expect a brief exposure to the arts, lasting a year or less, to have strong academic transfer effects.

A second reason for the failure to demonstrate a causal relationship has to do with the kinds of outcomes that almost all of our studies measured: multiple-choice test scores. The arts are messy, they do not point to clear answers, and they call for multiple and conflicting interpretations. They are fundamentally divergent. Standardized tests call for right and wrong answers and convergent thinking. These are not the measures where we should expect the arts to transfer.

Yet a third reason is that for transfer to occur, teachers must teach explicitly for transfer.[68] For instance, to demonstrate transfer of learning from one domain to another, teachers must explicitly lay the groundwork and help students to see how learning in area X may be similar to learning in area Y. Teachers need to help students develop strategies in one domain that can be applied to another domain, without at the same time simply teaching students a superficial rule or technique. In short, to find transfer, teachers need to teach not only for deep understanding of the learning domain (in this case, an art form), but also for transfer to the new domain (in this case, an academic area). None of the studies we found examined the effects of programs in which teachers taught for transfer.

It is noteworthy that our search turned up only two studies with a true-experimental design, in which students were randomly assigned to arts vs. control programs. These studies had extremely low effect sizes (r=.03 and r=.02).[69] Random assignment provides the best guarantee that the groups being compared are identical except for their exposure to arts education (the experimental treatment). Studies that do not randomly assign students to groups cannot rule out the possibility that differences between groups are due to preexisting differences. This is most clearly true of our correlational studies, but it is true as well of all of our experimental studies that compared intact classrooms rather than randomly assigning individuals to groups. Even though researchers controlled statistically for preexisting differences (in the studies we coded as matched-experimental), such procedures do not allow us to rule out preexisting differences as certainly as does random assignment.[70] This is because researchers can match only for some characteristics. Other unidentified preexisting differences may contribute to differences in scores.

Studies attempting to demonstrate transfer should include a control group in which students are given another form of treatment besides the arts. Otherwise, any positive effect of the arts may be due to the Hawthorne effect, in which any new kind of program at first leads to academic improvement. We were able to find only one experimental study that included both a no-treatment control group and a control group receiving some other form of treatment besides the arts. This study, by Richard Luftig, entitled "The Schooled Mind: Do the Arts Make a Difference?" is often cited by policymakers as having demonstrated that the arts cause academic improvement. In this study, some of the students receiving an arts-integrated program (the SPECTRA+ program) were compared not only to a full control group receiving no new treatment, but also to a "modified" control group. This modified control group was introduced to a new educational program involving cooperative learning, flexible ability grouping, and parental involvement. When all three groups were compared in terms of degree of improvement (from pre- to posttest) on reading and math, no differences were found among the three groups after either one or two years of the program.[71] (The justification for this conclusion, which differs strikingly from the way in which this study is usually described, can be found in the endnotes.)

The Need for Stronger Experimental Studies

The conclusions of the meta-analyses presented here refer only to studies assessing the impact of the arts in general, rather than those assessing the impact of specific art forms. Most of the remaining papers in this issue examine the evidence for a causal link between specific art forms and specific forms of non-arts cognition.

We conclude that we have as yet no evidence that studying the arts has a causal effect on academic achievement. We cannot draw any inferences about transfer from the correlational studies that we have reviewed and that are so often cited in the press, since correlational studies do not prove causation. For an investigation of causality, we must turn to the experimental studies. And the experimental studies revealed essentially no impact of the arts on academic outcomes.

The existing research is limited in its exclusive focus on outcome rather than process. That is, studies have examined the outcome of academic achievement and have simply speculated on how such an outcome might be achieved. We found only one study that assessed not only the desired outcome (e.g., increase in test scores) but also a plausible mechanism for such an outcome.[72] Below we suggest several types of study that could tell us not only whether the arts transfer, but if so, how they might do so.

Can study of the arts promote a way of thinking that then can be applied in an academic area in a way that improves performance in that academic area? For example, can an art class teach students the skill of close observation which could then be used in a science class? Or could an art class teach students the ability to think "outside the box" and make unusual connections, a skill that might then be used in an English literature or science class? Any study seeking to test such a hypothesis would need to assess two outcomes: the desired form of thinking must be demonstrated to occur in the art class; a correlation must then be shown between the mastery of this form of thinking in art class and the use of this form of thought in another class. In addition, such questions should be asked only of arts classes in which there is explicit teaching for transfer, since, as already stated, distant transfer is unlikely to occur unless teachers make explicit links from the learning to the transfer domain.

Can study of the arts promote a way of working that can then be applied in an academic area to the benefit of that area? For instance, do students learn perseverance when they are asked to work hard over long periods of time on an art project, and do they come away from such an experience with high standards about what they expect of themselves? And do they carry this learning with them into their academic subjects? A study seeking to test this hypothesis would need to assess whether perseverance and high standards develop in the art class, and then whether students who demonstrate these characteristics show these habits in an academic class. Again, such transfer, if it has any chance of occurring, would need to be explicitly encouraged by teachers urging students to take their new-found habits of learning and apply them to other areas of the curriculum. Note that the same question (with the same research design) could be asked about attitudinal mechanisms such as self-esteem which may derive from arts study.

Are the arts the best entry point into an academic area for all students, or is this true only for some students who are academically at risk? If researchers seek an answer to these questions, they must examine the effectiveness of teaching a subject matter directly vs. teaching that same subject "through the arts" (e.g., teaching history by getting students to study the art of the period, make art in the style of the period, enact dramas about the period, etc.). This comparison must be made for both at-risk and for non-at-risk students. If it can be demonstrated that both groups learn better when the subject is integrated with the arts, we can conclude that all students learn best when academic subjects are tied to the arts. If only some at-risk students learn best when the subject is integrated with the arts, we can conclude that students who do not learn easily can be helped to learn by integrating the subject with the arts. And of course, if neither group learns best through the arts, we will have found support for neither hypothesis.[73]

So often schools that make the arts central report that students learn better. We must not discount these claims. Rather, we suggest that researchers look closely and ethnographically at what happens to schools that grant the arts a central role in the curriculum. As mentioned at the outset of this article, schools that choose to integrate the arts with academic subjects may make other educational reforms at the same time. They may make learning more project-based, they may encourage more open-ended inquiry, or they may attract more enthusiastic, motivated teachers and/or students. If such changes occurred in association with bringing in the arts, we could account for the fact that so many schools that have infused the arts have enthusiastically reported that they have become more exciting places of learning.

It is our hope that the critique we have provided here of the extant literature will lead to stronger, more theory-driven research as illustrated by the examples we have provided above.

Policy Implications

The failure to find evidence of a causal relationship between arts study and academic achievement should never be used as a justification for cutting arts programs. The arts deserve a justification on their own grounds, and advocates should refrain from making utilitarian arguments in favor of the arts. Such arguments betray a misunderstanding of the inherent value of the arts.

As soon as we justify arts by their power to affect learning in an academic area, we make the arts vulnerable. Rarely has learning in any subject matter transferred to learning in another area, and we should not require more of the arts than we do of other subjects.[74] Were we to test whether math learning transfers to other subject areas, we would most likely find that it does not. But no one would use such a finding as a reason to cut mathematics from the curriculum. Stronger experimental research might ultimately reveal some causal by-products of the arts on academic achievement. And these findings may prove of educational value. But these by-products must never be used as utilitarian justifications for the presence of the arts in the curriculum.[75]

Any evaluation of the educational outcomes of arts education should be based on learning in the arts. We evaluate the outcomes of the study of math by determining the most important kinds of math understanding that we want our children to possess. Similarly, we should evaluate the outcomes of the study of the arts by determining the most important kinds of arts understanding that we wish to instill.

The arts are at a distinct disadvantage compared to academic areas when it comes to evaluating learning outcomes because, while the arts teach measurable skills, they also teach experiences and outcomes that are inherently difficult to measure and quantify. When we engage in the arts, we are likely to experience states of joy, appreciation, engagement, and flow. These are important positive experiences that enrich our lives. But they are not easily assessed by standard measures. We might profit from the work of Mihalyi Csikszentmihalyi, who has developed ways to assess such states.[76]

When we study the arts, we also learn new ways of self-expression and of communication. And we master symbol systems as complex and cognitively demanding as those of language and science.[77] The arts are important human ways of understanding and knowing, no less important than the sciences. Studying the arts should thus never be a frill, but should be a basic part of what we expect our children to learn. If they can be shown to aid learning in another domain, fine. But this should never be their primary purpose.

NOTES

1. Charles Leonhard, *The Status of Arts Education in American Public Schools* (Urbana: National Arts Education Research Center, University of Illinois, 1991).
2. Louis Harris, *Americans and the Arts VII: Full Report* (New York, N.Y.: Americans for the Arts, 1996); Nancy Welch, *Schools, Communities, and the Arts: A Research Compendium* (Washington, D.C.: National Endowment for the Arts, 1995).
3. As reported by *The New York Times* (November 11, 1998): A28.
4. For narrative reviews of research evaluating the claim that arts education transfers to academic learning, see Jaye T. Darby and James S. Catterall, "The Fourth

R: The Arts and Learning," *Teachers College Record* 96, no. 2 (1994): 299-328; Karen A. Hamblen, "Theories and Research that Support Art Instruction for Instrumental Outcomes," *Theory Into Practice* 32, no. 4 (1993): 191-98; James Hanshumaker, "The Effect of Arts Education on Intellectual and Social Development: A Review of Selected Research," *Bulletin, Council for Research in Music Education* 61 (1980): 10-28; David McCarthur and Sally A. Law, *The Arts and Public Safety Impact Study: A Review of Current Programs and Literature* (Santa Monica, Calif.: Rand Institute on Education and Training, 1996); Caroline Sharp, *The Effects of Teaching and Learning in the Arts: A Review of the Research* (London: National Foundation for Educational Research, 1998); Welch, *Schools, Communities, and the Arts.*

5. From testimony by the head of a cable TV station along with Wynton Marsalis presented to the U.S. House of Representatives Education Caucus, July, 1999.
6. From a letter to *The New York Times* by violinist Isaac Stern on August 9, 1998.
7. From Edward Fiske, ed., *Champions of Change: The Impact of the Arts on Learning* (Washington, D.C.: Arts Education Partnership and President's Committee on the Arts and Humanities, 1999): p.vi.
8. From Senator Alan K. Simpson's 1997 Nancy Hanks Lecture at the John F. Kennedy Center for the Performing Arts in Washington, D.C.
9. Judith H. Weitz, *Coming up Taller: Arts and Humanities Programs for Children and Youth at Risk* (Washington, D.C.: President's Committee on the Arts and Humanities, 1996).
10. For how the arts might transfer, see David N. Perkins, *The Intelligent Eye: Learning to Think by Looking at Art. Occasional Paper 4* (Los Angeles: Getty Center for Education in the Arts, 1994); and David N. Perkins, "Art as Understanding," in *Art, Mind, and Education*, ed. Howard Gardner and David Perkins (Urbana-Champaign: University of Illinois Press, 1989), pp. 111-31.
11. For a discussion of "teaching for transfer," see Gavriel Salomon and David N. Perkins, "Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon," *Educational Psychologist* 24, no. 2 (1989): 113-42.
12. Bruce Boston, *Connections: The Arts and the Integration of the High School Curriculum* (New York: College Board Publications, 1996); Liora Bresler, *General Issues across Sites: End of Year Report for the Getty/College Board Project, The Role of the Arts in Unifying the High School Curriculum* (Los Angeles: The J. Paul Getty Trust, 1997).
13. Joseph L. Mahoney and Robert B. Cairns, "Do Extracurricular Activities Protect against Early School Dropout?" *Developmental Psychology* 33, no. 2 (1997): 241-53.
14. For this argument, see McCarthur and Law, *The Arts and Public Safety Impact Study*, and Weitz, *Coming up Taller: Arts and Humanities Programs for Children and Youth at Risk.*
15. For relevant arguments and evidence, see Annenberg Challenge, "How the Arts Transform Schools: A Challenge for All to Share," *Challenge Journal* (1998): 1-4; Bresler, *General Issues Across Sites*, and Judith Burton, *Learning in and through the Arts* (New York: Center for Arts Education Research, Teachers College, Columbia University, 1999); Charles Fowler, *Strong Arts, Strong Schools: The Promising Potential and Shortsighted Disregard of the Arts in American Schooling* (New York: Oxford University Press, 1996); John Harland, Kay Kinder, Jo Haynes, and Ian Schagen, *The Effects and Effectiveness of Arts Education in Schools. Interim Report 1* (United Kingdom: National Foundation for Educational Research, 1998); Jane Remer, *Changing Schools through the Arts* (New York, N.Y.: Americans for the Arts, 1990); Bruce Wilson, Dick Corbett, Aimee Adkins, and George Noblit, *North Carolina A+ Schools Program. Report 2: 1995-1996* (Chapel Hill: Wilson-Corbett Associates and University of North Carolina, 1996).
16. For problems with including one group pre-posttest studies, see Paul M. Wortman, "Judging Research Quality," in *The Handbook of Research Synthesis*, ed. Harris Cooper and Larry V. Hedges (New York: Russell Sage Foundation, 1994): pp. 97-109.

17. Thomas D. Cook and Donald T. Campbell, *Quasi-experimentation: Design and Analysis Issues for Field Settings* (Boston: Houghton-Mifflin, 1979).
18. In two cases researchers provided pretest and posttest results, but in one case did not statistically compare pre-posttest difference scores for individuals: Martin G. Gardiner, Alan Fox, Faith Knowles, and Donna Jeffrey, "Learning Improved by Arts Training," *Nature* 381 (1996): 284. In the second case, the researchers did not make it clear whether they used gain or posttest scores: Mary C. Coakley, *Kaleidoscope Preschool Arts Enrichment Program: Making a Difference in the Lives of Children* (Philadelphia: Arts Programming and Research Services, Settlement Music School, 1995). For these two studies, we computed effect sizes for the pretest and posttest and then subtracted the pretest effect size from the posttest effect size. We classified these studies as experimental-matched even though this is not the ideal way, statistically, to determine differences in amount of gain from pre- to posttest between groups.
19. Unless the students were specified as low SES, we coded them as average SES. Thus, for example, urban public school students were coded as average, even though the students may come from below-average SES, since there was no way to know their SES.
20. The shortest program which we considered to have lasted for one academic year lasted from October through April.
21. Harland, *The Effects and Effectiveness of Arts Education.*
22. Robert Rosenthal, *Meta-Analytic Procedures for Social Research* (Newbury Park, Calif.: Sage Publications, 1991).
23. Ibid.
24. Ibid, pp. 17-18, for a discussion of why r is to be preferred over d.
25. We thus had to exclude two otherwise relevant studies: Dorothy Mathison, "A Comparative Analysis of the Effectiveness of Language Arts Instruction with Special Emphasis on Aesthetic Values for Selected Kindergarten, First, and Second Grade Children" (Ed.D. diss., University of Sarasota, Florida, 1977), and Geneva Dillard, "The Effect of a Fine Arts Program on Intelligence, Achievement, Creativity and Personality Test Scores of Young Gifted and Talented Students" (Ph.D. diss., East Tennessee State University, 1982). However, when these two studies were included, the mean effect size reported below remained unchanged at $r = .05$.
26. The studies in this meta-analysis are listed in Table 2; full references for each study can be found in the list of studies, below. The preschool study dropped was Mary Coakley, *Kaleidoscope Preschool Arts Enrichment Program: Making a Difference in the Lives of Children* (Philadelphia Arts Programming and Research Services, Settlement Music School, 1995).
27. James Catterall, Richard Chapleau, and John Iwanaga, "Involvement in the Arts and Human Development: General Involvement and Intensive Involvement in Music and Theater Arts," in *Champions of Change: The Impact of the Arts on Learning*, ed. Edward Fiske (Washington, D.C.: Arts Education Partnership and The President's Committee on the Arts and Humanities, 1999).
28. National Center for Education Statistics, *Course Offerings and Enrollments in the Arts and the Humanities at the Secondary Level* (Washington, D.C.: US Government Printing Office, 1984).
29. Scott Whitener, "Patterns of High School Studies and College Achievement" (Ph.D. diss., Rutgers University, 1974).
30. Patricia L. Dwinnell and Mark C. Hogrebe, "Differences among Ability Groups in Participation in the Performing Arts at the High School Level" (Paper presented at the American Education Research Association, 1984).
31. Shirley Brice Heath, "Living the Arts through Language and Learning: A Report on Community-Based Youth Organizations," *Americans for the Arts Monographs* 2, no. 7 (1998); Shirley Brice Heath, "Youth Development and the Arts in Nonschool Hours," *Grantmakers in the Arts* 9, no. 1 (1998): 9-16, 32.
32. One study (Patricia Dwinell and Mark Hogrebe, "Differences among Ability Groups in Participation in the Performing Arts at the High School Level," coded

students as having participated in the arts if they had participated at any time in their senior year of high school. Thus students might have participated for a year or for under a year. To be conservative we coded this as a program of study lasting less than a year.

33. Catterall, Chapleau, and Iwanaga, "Involvement in the Arts and Human Development"; James Catterall, "Involvement in the Arts and Success in the Secondary School," *Americans for the Arts Monographs* 1, no. 9 (1998). We selected the low SES sample from Catterall's study, rather than the all SES sample, because the data from Heath came from a low SES sample. These two studies were similar in design, and we therefore wanted to render them as similar as possible in type of student assessed. In the two later meta-analyses, we selected the all SES sample from Catterall's study, this time in order to make this study as similar as possible to the College Board studies included in the later meta-analyses. We also selected 12th rather than 8th or 10th graders from this study in order to make the data as comparable as possible to the College Board data.
34. For the formula to convert an effect size *r* to an effect size *d*, see Rosenthal, *Meta-Analytic Procedures*, p. 20.
35. Ibid., pp. 68-69, Formula 2.21.
36. Ibid., pp. 72-74.
37. Full references for these studies can be found in the list of studies below.
38. Catterall et al., "Involvement in the Arts and Human Development."
39. Elliot Eisner "What Justifies Arts Education: What Research Does *Not* Say," in *Enlightened Advocacy: Implications of Research for Arts Education Policy Practice*, ed. M. McCarthy (The 1999 Charles Fowler Colloquium on Innovation in Arts Education, College Park: University of Maryland, 2000).
40. Calculations made from data which Shirley Brice Heath kindly made available to us.
41. Paul DiMaggio, "Cultural Capital and School Success: The Impact of Status Culture Participation on the Grades of U.S. High School Students," *American Sociological Review* 47 (1982): 189-201. We were not able to include this study in our meta-analysis because we could not compute an effect size from the multiple regression that was performed.
42. Pierre Bourdieu, "Cultural Reproduction and Social Reproduction," in *Power and Ideology in Education*, ed. Jerome Karabel and A. H. Halsey (New York: Oxford, 1977).
43. W. W. Welch, R. E. Anderson, and L. J. Harris, "The Effects of Schooling on Math Achievement" (Unpublished document, University of Minnesota, 1980), cited in DiMaggio, "Cultural Capital."
44. Catterall et al., "Involvement in the Arts and Human Development."
45. Note that this effect size differs from the ones in the meta-analyses because these were based on the low SES sample (meta-analysis 1) or on the all SES sample with either verbal or math outcomes (meta-analysis 2 and 3).
46. We thank Teighe Sheehan, a graduate student at Boston College, for suggesting that causal effects might occur only for students who self-select into the arts.
47. The two true-experimental studies were Dillard (1982) and Jackson (1979) whose full references can be found in the list of studies, below.
48. Sharon Brock, *The Visual and Performing Arts Magnet Elementary Schools: Summative Evaluation* (Kansas City School District, Kansas City, Mo., Program Evaluation Office, 1991); Sharon Brock, *Achievement and Enrollment Evaluation of the Kansas City Middle School of the Arts Magnet, 1990-1991* (Kansas City, Mo.: Kansas City School District Program Evaluation Office, 1991).
49. This is the procedure recommended by Rosenthal, *Meta-analytic Procedures.*
50. Dillard had the effect size of .03, and Jackson had the effect size of .02.
51. Jeanne Tunks, *Changing the Face of American Education: The Partnership Assessment Project* (Dallas: Partnership for Arts, Culture and Education, Inc., 1997).
52. We calculated an effect size comparing the arts-integrated experimental group to the control group that received no arts. We made this decision in order to make this study as comparable to the others as possible.

53. Richard Luftig's study, grade 5. (Full reference in the list of studies, below.)
54. Catterall et al., "Involvement in the Arts and Human Development;" College Board, 1987-1997 (for full reference, see list of studies, below).
55. Full references to these studies can be found in the list of studies, below.
56. Folkert Haanstra, "Arts Education and the Educational Careers of Dutch Students" (Unpublished document, University of Amsterdam, 1999).
57. Harland et al., *The Effects and Effectiveness of Arts Education*.
58. Catterall et al., "Involvement in the Arts and Human Development."
59. Heath, "Living the Arts," and "Youth Development."
60. Haanstra, "Arts Education."
61. Two studies could not be included in Table 14 because their results are not yet in. The North Carolina A+ Schools Program, run by the Thomas Kenan Institute for the Arts in Winston-Salem, North Carolina, is currently under evaluation. An evaluation is planned of Spectra Rhode Island, a new state-wide arts and education program currently under development by Ralph Burgard of Burgard Associates and Educational Development Center, Inc. of Newton, Massachusetts.
62. Full references for these studies can be found in the list of studies, below.
63. Some of the effect sizes from our experimental studies may have been derived from a self-selected sample. Brock analyzed academic performance of students in an arts magnet elementary and middle school, and some of these children (or their parents) may have selected these schools for the arts. However, we do not believe that the case for self-selection here is as strong as it is for students taking arts electives in high school, since parents typically decide where their children go to school, and may choose an arts magnet school on the basis of factors such as location and academic reputation as well as the presence of the arts. Ashbacher and Herman (full reference in the list of studies, below) also evaluated academic performance of students who volunteered to participate in an arts program. This was the only experimental study with a clearly self-selected sample. In contrast, all of the correlational studies had self-selected samples.
64. Catterall et al., "Involvement in the Arts and Human Development"; Heath, "Living the Arts" and "Youth Development."
65. Harland et al., "The Effects and Effectiveness of Arts Education."
66. John Harland, the author of the British study, has indicated to us that our suggestion about the minimal role of arts study for admission to selective universities in the United Kingdom is correct.
67. Haanstra, "Arts Education." Haanstra has also indicated to us that having studied the arts is not something that influences admission to university in the Netherlands. Admission to Dutch universities is based entirely on performance on the final exam.
68. For a discussion of "teaching for transfer," see Salomon and Perkins, "Rocky Roads to Transfer."
69. The two experimental studies with these effect sizes were, respectively, Dillard (1982) and Jackson (1979). (Full references are in the list of studies, below.)
70. J. P. Greene "Rescuing Education Research: A Rule of Thumb for Fending off the 'Nihilism' of Competing Claims," *Education Week* (April 29 1998): 52.
71. In the first year evaluation of the SPECTRA+ program, Luftig reports that for School District A (the only one which included the modified control group), no differences were found in Total Reading (p. 20); he also reports no main effect of condition for Total Math (p. 21). In the second year evaluation of the program, Luftig reports no differences over time among the three groups on Total Reading (p. 20). A difference was reported for Total Math, with SPECTRA+ students highest. However, all three groups showed the same rate of improvement from pretest to posttest at Year 1, and final posttest at Year 2 (p. 24). Thus, we cannot conclude from these data that SPECTRA+ had any effect on either reading or math outcome. If a positive effect had been found, we could then compare the effect of SPECTRA+ with that of the modified control in order to rule out a Hawthorne effect. For a similar critique of this study, → Elliot Eisner, "Getting

Down to Basics in Arts Education," *Journal of Aesthetic Education* 33, no. 4 (1999): 145-59.

72. Susan Baum and her colleagues (full reference in the list of studies, below) compared self-regulatory behavior during classes in which the arts were and were not integrated into the lesson. Self-regulation was higher in the arts-integrated classes, as were reading scores. Studies that assess both academic and motivational/behavioral outcomes can tell us whether improved motivational/behavioral outcomes lead to higher academic achievement. Unfortunately, Baum's study did not test for a correlation between academic and behavioral outcomes. This kind of test is needed if we are to determine whether the behavioral outcome actually was associated at the individual level with academic improvement.
73. Baum's study (full reference in the list of studies, below) provides us with some evidence that at-risk students benefit more than average students from an arts-integrated curriculum. This study compared the effect of arts integrated study on average vs. academically at-risk students and reported that for reading, the greatest gains were for the at-risk students. The average students gained no more than the control group. However, when the outcome was math, neither average nor at-risk students exposed to an arts-integrated curriculum improved more than the control group. This finding is not reflected in the effect sizes we calculated for our meta-analyses, however. We made the decision to treat both arts groups as equivalent, assigned them the same lambda weight, and computed a maximum possible contrast *F* in order to calculate an effect size. We did this because none of our other studies compared an average to an at-risk group, and we wanted this study to be comparable to the others. In order to test the hypothesis that the arts serve as an entry point particularly for academically at-risk students, a repeated measures design is preferable, in which both at-risk and average students are taught an academic subject with and without the arts. One can then determine whether the arts make a difference only for the at-risk group.
74. For a discussion of transfer, see David Perkins, *Outsmarting IQ: The Emerging Science of Learnable Intelligence* (New York: Free Press, 1995).
75. For similar arguments, see Jessica Davis, "Why Must We Justify Arts Learning in Terms of Other Disciplines?" *Education Week* (October 16 1996): 32. Elliot Eisner, "Does Experience in the Arts Boost Academic Achievement?" *Art Education* 51 no. 1 (1998): 7-15; and Constance B. Gee, "For You Dear—Anything! Omnipotence, Omnipresence, and Servitude 'Through the Arts', Part I: Implications of Arts Education as Public Service," *Arts Education Policy Review* 100, no. 4 (1999): 3-17.
76. Mihaly Csikszentmihalyi and Isabella Csikszentmihalyi, *Optimal Experience: Psychological Studies of Flow in Consciousness* (New York: Cambridge University Press, 1988).
77. For a discussion of the symbol systems of the arts, see Nelson Goodman, *Languages of Art* (Indianapolis: Hackett, 1976); and Susanne Langer, *Philosophy in a New Key* (New York: Scribner, 1942).

Studies Included in Meta-analyses and in Tables 14-15

*Citations preceded by an asterisk were included in one or more of the three meta-analyses. All other citations are those studies listed in Tables 14 and 15.

Annenberg Challenge, "How the Arts Transform Schools: A Challenge for All to Share," in *Challenge Journal*, Annenberg Challenge (1998): 1-4.

* Aschbacher, Pamela, and Joan Herman, *The Humanitas Program Evaluation 1990-91* (Center for Study of Evaluation, UCLA Graduate School of Education and Information Studies, 1991).

* Baum, Susan, and Steven Owen, "Using Art Processes to Enhance Academic Self Regulation" (Paper presented at the ArtsConnection National Symposium on

Learning and the Arts: New Strategies for Promoting Student Success, New York, 1997).

* Brock, Sharon, *The Visual and Performing Arts Magnet Elementary Schools: 1988-1989,1989-1990, 1990-1991. Summative Evaluation.* (Kansas City, Mo.: Kansas City School District, Program Evaluation Office, Desegregation Planning Department, 1991a).

* Brock, Sharon, *Achievement and Enrollment Evaluation of the Kansas City Middle School of the Arts Magnet, 1990-1991* (Kansas City, Mo.: Kansas City School District, Program Evaluation Office, Desegregation Planning Department, 1991b).

Brock, Sharon, *The Paseo Fine and Performing Arts Magnet High School, 1990-1991. Formative Evaluation* (Kansas City, Mo.: Kansas City School District, Program Evaluation Office, Desegregation Planning Department, 1991).

Burton, Judith, Robert Horowitz, and Hal Abeles, "Learning in and through the Arts," in *Champions of Change: The Impact of the Arts on Learning* ed. Edward Fiske (Washington, D.C.: The Arts Education Partnership and The President's Committee on the Arts and the Humanities, 1999): pp. 35-46.

* Catterall, James, and Lynn Waldorf, "Chicago Arts Partnerships in Education: Summary Evaluation," in *Champions of Change: The Impact of the Arts on Learning*, ed. Fiske (Washington, D.C.: The Arts Education Partnership and The President's Committee on the Arts and the Humanities, 1999): pp. 47-62.

Catterall, James, "Different Ways of Knowing 1991-1994. National Longitudinal Study Final Report. Program Effect on Students and Teachers" (Los Angeles: The Galef Institution, 1995).

* Catterall, James, "Involvement in the Arts and Success in the Secondary School." *Americans for the Arts Monographs* 1, no. 9 (1998).

* Catterall, James, Richard Chapleau, and John Iwanaga, "Involvement in the Arts and Human Development: General Involvement and Intensive Involvement in Music and Theater Arts," in *Champions of Change: The Impact of the Arts on Learning*, ed. Edward Fiske (Washington, D.C.: The Arts Education Partnership and The President's Committee on the Arts and the Humanities, 1999): pp. 1-18.

Center for Music Research, *The Role of the Fine and Performing Arts in High School Dropout Prevention* (Florida Department of Education, Division of Public Schools, 1990).

Chapman, Russ, "Improving Student Performance through the Arts," *Principal* 77, no. 4 (1998): 20-24.

* Coakley, Mary, *Kaleidoscope Preschool Arts Enrichment Program: Making a Difference in the Lives of Children. Final Report of a Four-Year Evaluation of an Arts-Based Early Intervention Program.* (Philadelphia: Settlement Music School, 1995).

* *College Bound Seniors Profile of SAT and Achievement Test Takers* (College Board, 1987-1997).

* Demeter, Myra, "An Investigation of the Problem-Solving Processes Used by Art/Music Talent Students to Solve Multistep Mathematical Problems" (Doctoral diss., New York University, 1986).

Department of Test Development and Administration, Prince George's County Public Schools, *Unpublished Data on Rockledge Elementary School and Thomas G. Pullen Creative and Performing Arts School* (1997).

* Dillard, Geneva, "The Effect of a Fine Arts Program on Intelligence, Achievement, Creativity, and Personality Test Scores of Young Gifted and Talented Students" (Doctoral diss., East Tennessee State University, 1982).

* Dwinnell, Patricia, and Mark Hogrebe, "Differences among Ability Groups in Participation in the Performing Arts at the High School Level" (Paper presented at the American Education Research Association 1984).

Edmunds, R. "A Livelier Way to Learn: The Greater Augusta Arts Council's Innovative Arts Infusion Program Uses the Arts to Bring New Skills to the Classroom," *Augusta Magazine* (1991): pp. 64-67.

Fowler, Charles, *Try a New Face* (Washington, D.C.: Office of Education, 1979).

* Gardiner, Martin, Alan Fox, Faith Knowles, and Donna Jeffrey, "Learning Improved by Arts Training," *Nature* 381, (1996): 284.

Getz, Russell, "Excellence and the Classroom," *Design for Arts in Education* 85, no. 4 (1984): 38-40.

* Glismann, Leonard, "The Effects of Special Arts and Crafts Activities on Attitudes Attendance, Citizenship, and Academic Achievement of Slow Learning Ninth Grade Pupils" (Ed.D. diss., Utah State University, 1967).

Goldberg, Merryl, "SUAVE: Integrated Arts and Academic Achievement" (Unpublished paper, California State University, 1998).

Haanstra, Folkert, "Arts Education and the Educational Careers of Dutch Students" (Unpublished document, University of Amsterdam, 1999).

Harland, John, Kay Kinder, Jo Haynes, and Ian Schagen, *The Effects and Effectiveness of Arts Education in Schools. Interim Report 1* (UK: National Foundation for Educational Research, 1998).

* Heath, Shirley Brice, "Living the Arts through Language and Learning: A report on Community-Based Youth Organizations," *Americans for the Arts Monographs*, 2, no. 7 (1998).

* Hudspeth, Carolyn, "The Cognitive and Behavioral Consequences of Using Music and Poetry in a Fourth Grade Language Arts Classroom" (Ed.D. diss., Texas Women's University, 1986).

* Jackson, Ernest, "The Impact of Arts Enrichment Instruction on Self-Concept, Attendance, Motivation, and Academic Performance" (Doctoral diss., Fordham University, 1979).

Lardo, Nancy,"Examining Changes in Self-Concept and Other Selected Variables in Seventh and Eighth Grade Students Participating in a Creative and Performing Arts Magnet Middle School Program" (Doctoral diss., University of Pittsburgh, 1982).

Lawton, Edward, "The Role of the Arts in Schools: Another Reminder," *Contemporary Education* 59, no. 1 (1987): 15-16.

* Luftig, Richard, *The Schooled Mind: Do the Arts Make a Difference? An Empirical Evaluation of the Hamilton Fairfield SPECTRA+Program, 1992-1993*. (Oxford, Ohio: Center for Human Development, Learning, and Teaching, Miami University, 1993).

Luftig, Richard, *The Schooled Mind: Do the Arts Make a Difference: An Empirical Evaluation of the Hamilton Fairfield SPECTRA+Program, 1993-1994* (Oxford, Ohio: Center for Human Development, Learning, and Teaching, Miami University, 1994).

Mahoney, Joseph, and Robert Cairns, "Do Extracurricular Activities Protect Against Early School Dropout?" *Developmental Psychology* 33, no. 2 (1997): 241-53.

* Martson, Janis, "Arts Integration: Evaluating Its Effectiveness at the Sherwood Heights School" (Masters thesis, University of Maine, 1997).

Maryland Alliance for Arts Education, "The Arts and Children: A Success Story" (1995).

Mathison, Dorothy, "A Comparative Analysis of the Effectiveness of Language Arts Instruction with Special Emphasis on Aesthetic Values for Selected Kindergarten, First, and Second Grade Children" (Ed.D diss., University of Sarasota, 1977).

McGuire, Gary,"Effects of the Arts on Academic Achievement" (Masters thesis, California State Polytechnic University, 1982).

McLaughlin, John, *Building a Case for Arts Education: An Annotated Bibliography of Major Research* (Lexington, Ky.: Kentucky Alliance for Arts Education, 1990).

Missouri Arts Education Task Force, "Final Report" (St. Louis, Mo, 1990).

Murfee, Elizabeth, *The Value of the Arts* (President's Committee on the Arts and the Humanities, National Endowment for the Arts, 1993).

* National Center for Education Statistics, *Course Offerings and Enrollments in the Arts and the Humanities at the Secondary Level* (Washington, D.C.: US Government Printing Office, 1984).

Newbill, Sharon, *The Paseo Fine and Performing Arts Magnet High School, 1989-1990, 1990-1991, 1991-1992. Summative Evaluation* (Kansas City, Mo.: Kansas City School District, Program Evaluation Office, Desegregation Planning Department, 1992).

* Norman, Janis, "The Effectiveness of an Integrated Arts Program on Students' Academic Achievement" (Doctoral diss., University of Kansas, 1987).

Redcliffe School Performance Profile 1990-1994.(Aiken, S.C.: Aiken County School District, 1994).

Redfield, Doris, *Evaluating the Broad Educational Impact of an Arts Education Program. The Case of the Music Center of Los Angeles County's Artist in Residence Program* (Center for the Study of Evaluation, UCLA Graduate School of Education, 1990).

Rombokas, Mary, Jeannette Heritage, and W. Beryl West, "High School Extracurricular Activities and College Grades" (Paper presented at the Southeastern Conference of Counseling Center Personnel and the Tennessee Counseling Association Convention, Jekyll Island, Ga., and Nashville, Tenn, October 25-27 and November 19-21, 1995).

Ross, Gerald, *Arts Partners Research Study: Final Report* (National Arts Education Research Center, New York University, 1990).

Slay, Jane, and Sherril Pendergast, "Infusing the Arts across the Curriculum," *School Administrator* 50, no. 5 (1993): 32-35.

Spady, William, "Status, Achievement, and Motivation in the American High School," *School Review* 29 (1971): 367-97.

Spectra Rhode Island, *Basic Questions* (Unpublished report, Burgard Associates, Box 697, Providence, R.I. 02091-0697, 1998).

Spilke, Gertrude, with Charles Fowler and Bernard McMullen, *Understanding How the Arts Contribute to Excellent Education. Study Summary* (Philadelphia: National Endowment for the Arts, 1991).

* Tunks, Jeanne, *Changing the Face of American Education: The Partnership Assessment Project*. (Dallas: Partnership for Arts, Culture and Education, Inc., 1997).

Walker, Doris, "Connecting Right Brain and Left Brain: Increasing Academic Performance of African American Students through the Arts" (Paper presented at the Annual Meeting of the National Alliance of Black School Educators, Dallas, 1995).

* Whitener, Scott, "Patterns of High School Studies and College Achievement" (Doctoral diss., Rutgers University, 1974).

Wilson, Bruce, Dick Corbett, Aimee Adkins, and George Noblit, *North Carolina A+ Schools Program. Report 2: 1995-1996* (Chapel Hill: Wilson-Corbett Associates and University of North Carolina, 1996).

Acknowledgments: This research was supported by the Bauman Foundation. We thank John Landrum Bryant of the Bauman Foundation for his support and guidance. We thank Ron Butzlaff for statistical advice and for providing us with the program with which to conduct these analyses. We thank the following individuals for their helpful comments about the ideas developed here: Howard Gardner, Lois Hetland, Elliot Eisner, James Catterall, Ali Banuazizi, Joseph Tecce, and Teighe Sheehan. We thank Kristin Burger for help in conducting the literature searches, Erik Moga for serving as second coder on all studies, and Elisabeth Moriarty-Ambrozaitis for help in acquiring and summarizing many of the studies used in these meta-analyses. Portions of this work were presented at the American Psychological Association, August, 1999; the Social Science Research Council, October, 1999; a Boston College Faculty Seminar, November, 1999; and the American Educational Research Association, March, 2000.
Requests for reprints should be sent to Ellen Winner, Department of Psychology, Boston College, Chestnut Hill, MA 02467; or by email to ewinner@mediaone.net.