BEYOND THE SOUNDBITE:

Arts Education and Academic Outcomes

Conference Proceedings from

Beyond the Soundbite:

What the Research Actually

Shows About Arts Education and

Academic Outcomes

The Getty Center

Los Angeles, California

August 24–26, 2000

Ellen Winner and Lois Hetland

Editors

Beyond the Soundbite: Arts Education and Academic Outcomes

Conference Proceedings from

Beyond the Soundbite: What the Research Actually Shows About Arts Education and Academic Outcomes

The Getty Center • Los Angeles, California • August 24–26, 2000

Ellen Winner and Lois Hetland
Editors

Beyond the Soundbite: Arts Education and Academic Outcomes

1200 Getty Center Drive
Suite 400
Los Angeles, California 90049-1681
www.getty.edu/

Publisher
Christopher Hudson
Managing Editor
Mark Greenberg

Editors
Ellen Winner and Lois Hetland
Project Coordinator
Josephine Ramirez
Manuscript Editor
Mollie Holtman
Production Coordinator
Elizabeth Burke Kahn
Designer
Hespenheide Design

Table of Contents

Preface

As part of its ongoing work in education and the arts, the Getty Trust joined with Harvard's Project Zero to convene a conference at the Getty Center from August 24 to 26, 2000, entitled "Beyond the Soundbite: What the Research Actually Shows About Arts Education and Academic Outcomes." The goal of the conference was to examine, in all their complex implications, the many answers that have been given by researchers to the question of whether and how the study of the arts can improve a student's academic performance.

The focus of the conference was a three-year study, Reviewing Education and the Arts Project, also known as REAP. The REAP study was directed by Ellen Winner and Lois Hetland of Project Zero, funded by the Bauman Foundation, and presented first as a series of papers in a special Fall/Winter issue of the *Journal of Aesthetic Education*. REAP set out to evaluate the hundreds of studies examining the relationship between some form of arts study and some form of academic outcome. The project examined several key relationships, from broader issues, such as the relationship between the study of the arts in general and a student's academic achievement as measured by standardized tests and grades, and the relationship between arts study and creativity as measured by various creativity tests; to more particular effects, such as the relationship of music and spatial reasoning, and the relationship between visual arts and reading.

The conference took up these same issues, but provided the opportunity for scholars from widely divergent perspectives to discuss critically and reflect upon the REAP findings. What follows are the proceedings of this conference, including the talks, commentaries, and suggestions for further research. Although the points of view on the REAP findings varied greatly

among the thirty-six conference participants, virtually all of the participants came to agree that the answers to the central question about the relationship between arts study and academic achievement cannot be characterized as simple or straightforward. In some cases, it was clear that positive relationships were demonstrated and in others they were not. Often, it appeared that more research was necessary before any definitive conclusions could be drawn or, in still other cases, that different questions really needed to be asked.

Unfortunately, however, the desire, particularly on the part of the media and policy makers, for a single, definitive answer to the art-and-academic-achievement question is a powerful one. We can only hope that the conference, and this present volume of its proceedings, will act as a counterweight by taking the issues beyond soundbite simplicity and by helping to convey the true complexity of the issues involved.

On the Getty Trust's behalf, I would like to take this opportunity to thank Ellen Winner and Lois Hetland for all their hard work in helping to organize and chair the conference, and to thank the conference participants for their thoughtful contributions. This conference represents the first in what we hope will be a series of convenings at the Getty Center that will examine central questions related to research in arts education, and we look forward to the opportunity to work with a wide range of scholars in this field.

Jack Meyers
Deputy Director
Getty Grant Program
The J. Paul Getty Trust

Acknowledgments

We would like to thank John Landrum Bryant, codirector of the Bauman Foundation, for funding the Reviewing Education and the Arts Project (REAP) work and Ralph Smith, former editor of the *Journal of Aesthetic Education*, for publishing the collected REAP papers in a special issue. We express our special thanks as well to the REAP advisors: Howard Gardner, John H. and Elisabeth A. Hobbs Professor of Cognition and Education, Harvard Graduate School of Education; Richard Light, Professor, Harvard Graduate School of Education and John F. Kennedy School of Government; David N. Perkins, Professor, Harvard Graduate School of Education; Robert Rosenthal, Edgar Pierce Professor of Psychology, Emeritus, Harvard University, Distinguished Professor, University of California, Riverside; and Judith D. Singer, Academic Dean and Professor of Education, Harvard Graduate School of Education.

We are most grateful that the President and Chief Executive Officer of the J. Paul Getty Trust, Barry Munitz, was intrigued enough by our work to offer to host a conference on the topic of arts education research. We are especially grateful to Jack Meyers, Deputy Director of the Getty Grant Program, for his guidance in helping to plan this conference. We thank Josephine Ramirez, Program Officer, and Jenny Goto, Senior Staff Assistant, of the Getty Grant Program for their help in planning and running the conference on which this volume is based and for their assistance in preparing this publication.

We would like to acknowledge the other REAP researchers whose papers were drawn upon in the summary of the REAP findings: Kristin Burger, Ron Butzlaff, Monica Cooper, Mia Keinanen, Erik Moga, Ann Podlozny, and Kathryn Vaughn. Additionally, we thank Ron Butzlaff for serving as the statistical advisor to the project.

We also want to thank the research assistants who helped us at various times during the project: Kristin Burger, Lisa French, Kimberlee Garris, Nandita Ghosh, Maxwell Gomez-Trochez, Jessica Gordon, Joanna Holtzman, Jenny Martin, Elisabeth Moriarty-Ambrozaitis, Brian Moss, Melissa Mueller, Leah Okimoto, Nina Salzman, and Daniel Schneider.

And finally, for their willingness to answer all our questions about details of their work, we thank the researchers whose work we reviewed.

Ellen Winner and Lois Hetland

I.

Meta-Analysis: Its Use and Value in Arts Education Research

ROBERT ROSENTHAL
University of California, Riverside

LOIS HETLAND
Project Zero, Harvard University

The research conducted by the Reviewing Education and the Arts Project (REAP), summarized in this volume, is a set of meta-analyses that synthesized empirical evidence about transfer effects from arts to non-arts cognition and learning.[1] Because meta-analysis is a relatively unfamiliar tool in the field of arts education research, a description of the nature and value of this quantitative technique for synthesizing empirical research should help readers better understand the summaries that follow.

The fundamental goal of meta-analytic procedures is the cumulation of evidence for the purpose of furthering understanding of past research and guiding future inquiry. Meta-analytic methods are similar to those of other quantitative research procedures. First, meta-analysts define a research domain. Then the search for studies (the "subjects" of a meta-analysis) commences in a comprehensive and systematic manner so as to achieve a sample as unbiased as possible to represent the population of studies that explores the question. Next, studies are coded for qualities whose variation may be associated with the size of effect. Average effects and combined significance levels are computed through standard statistical procedures. Studies can then be grouped and compared to explain whether effects vary systematically by identified "moderator" variables. Finally, results are clearly reported according to standard criteria for empirical research and interpreted in relation to the body of literature and to the field at large.

There has long been underlying pessimism in the younger social, behavioral, educational, and biomedical sciences that our progress has been exceedingly slow and less orderly than we would like. Work in physics and

chemistry, for example, seems to build directly upon the older work of these sciences, whereas the social, behavioral, educational, and biomedical sciences seem often to be starting from scratch. In the arts, the issue of cumulation may be especially controversial, since some maintain that arts or aesthetics cannot even be studied empirically, much less quantitatively summarized. Debate on this question has persisted for at least half a century, despite serious attention to subtle human behaviors that are relevant to artistic experience (e.g., exploration, pleasure, flow, preference) and the development of methods suited to the study of aesthetic and artistic outcomes.[2–6] Measures still may be too blunt, treatments ill-defined, or interpretation flat-footed, but all these elements can be improved through the reflective process of ongoing research, synthesis, and redefinition of the questions of interest. Methods only improve through creative thinkers' efforts to solve interesting puzzles.

Not surprisingly, the REAP researchers found nuanced, conflicting results in studies of variable quality, often incompletely reported. Although it may be cold comfort, such problems are not specific to research in the arts but characterize a great deal of social science and biomedical research. While meta-analysis cannot completely solve these problems of cumulation, it can help fields to address them by bringing methodological and theoretical issues to light and by specifying what has and has not been accomplished by research conducted to date.

The problems of cumulation do not seem to be due to lack of replication, or to the failure to recognize the need for replication. Indeed, there are many areas of the social, behavioral, educational, and biomedical sciences—including the arts, as demonstrated through REAP's efforts—for which the results of many studies, all addressing essentially the same question, are available. However, reviews of research in education have not been nearly as informative as they might have been, either with respect to summarized significance levels (i.e., how confidently we can assert that there is some non-zero effect of an intervention) or with respect to summarized effect sizes (i.e., how strongly any intervention is related to the outcome as measured). Even the best reviews of research by the most sophisticated scholars have been primarily qualitative narratives and have rarely told us much more about each study in a set of studies than the direction of the relationship between the variables investigated (e.g., smaller classes improve student learning), and whether or not a given significance level was attained (i.e., usually whether an effect found in a given study showed $p \leq .05$). If traditional narrative reviews only summarize researchers' conclusions rather than synthesizing and interpreting study data, the actual findings may be misinterpreted.

This state of affairs is beginning to change, however. More and more reviews of the literature in the social sciences are moving from the traditional literary approach toward quantitative approaches to research synthesis described in an increasing number of textbooks of meta-analysis.[7–14] Because meta-analytic procedures and methods are codified and described explicitly, meta-analytic reviews are more replicable than traditional reviews and allow reviewer bias to be revealed over time through the scientific process.

Meta-analytic syntheses do not intend to be the final word on a research area, but rather to clarify what has been learned thus far from the studies conducted and to determine what remains to be learned. The REAP research summarized here assesses what we know to date from quantitative studies about cognitive transfer to non-arts learning that results from arts education. Its results can be used to guide future studies on this complex question.

Defining Research Results

Before we can consider various issues and procedures in the cumulation of research results, we must become quite explicit about the meaning of the concept "results of a study." It is easiest to begin with what we do not mean. We do not mean the prose conclusion drawn by the investigator and reported in the abstract, results, or discussion sections of a research report. We also do not mean the results of an omnibus F test with $df > 1$ in the numerator or an omnibus χ^2 test with $df > 1$. Such statistics are commonly reported when research designs employ more than two conditions (i.e., a "treatment" and one "control"), but they do not answer precise research questions and can obscure important trends that data might reveal when more focused methods of analysis (i.e., contrasts) are used.

What we do mean by results is the answer to the question: What is the relationship between any variable X (e.g., arts instruction) and any variable Y (e.g., test scores or grades)? The variables X and Y are chosen with only the constraint that their relationship be of interest to us. In the case of the REAP analyses, X is some form of arts intervention (e.g., dance, drama, music, visual art, or a combination of these areas that the REAP researchers call "multi-arts"), and Y is a score on some measure of cognition or learning in a non-arts area (e.g., grades, mathematics or verbal achievement test scores, or scores on intelligence subtests, such as spatial tests). The answer to the question of a relationship should normally come in two parts: (a) the estimate of the magnitude of the relationship (the effect size), and (b) an indication of the accuracy or reliability of the estimated effect size (e.g., as indexed by a confidence interval placed around the effect size estimate). An alternative to the second part of the answer is one not intrinsically more useful, but one more consistent with

the existing practices of researchers: the significance level of the difference between the obtained effect size and the effect size expected under the null hypothesis (usually an effect size of zero).

The REAP analyses employed all these methods for assessing accuracy and reliability of results. Descriptive statistics report the range of effects (largest and smallest), the quartiles (twenty-fifth, fiftieth—also called the median—and seventy-fifth percentiles), and the percentage of effects greater than zero. They also report 95% confidence intervals around the average effect size. When these intervals span zero, confidence that a positive effect actually exists is typically reduced. The REAP researchers also combined significance tests from individual studies by using two models. The fixed-effects model (indexed by Stouffer's *Z*) employs as its sample size the total number of participants in all the studies and tends to find more significant results.[15,16] The generalization from fixed-effects models, however, is limited to other participants in the same studies, but often it is the best analysts can do. However, because the REAP searches found a substantial number of studies, the REAP researchers were also able to compute significance levels using the more generalizable random-effects model (the *t*-test of the mean *Zr*, which tells us whether the average obtained effect size is significantly different from zero.[16,17] Unlike the situation for fixed-effects analyses, results from random-effects models can be generalized to other studies on the research question, and not merely to other participants in the *same* studies.

Because a complete reporting of the results of a study requires reporting both effect size and level of statistical significance, it is useful to make explicit the relationship between these quantities. The general relationship is given by the following formula:

$$\text{Significance Test} = \text{Effect Size} \times \text{Study Size}$$

The point is this: the same small, moderate, or large effect size could be significant or not *depending only on the size of the sample*. In other words, the larger the study in terms of the number of sampling units, the more significant the results will be. This is true unless the size of the effect is truly zero, in which case a larger study will not produce a result that is any more significant than a smaller study. Effect sizes of exactly zero, however, are rarely encountered.

Meta-Analysis: A Brief Historical Note

Many are inclined to think of meta-analysis as a recent development, but it is older than the *t*-test, which dates back to 1908![18] In 1904, Karl Pearson[19] collected correlation coefficients; there were six of them with values of .58, .58, .60, .63, .66, and .77. The weighted mean of these six correlation coefficients was .64, the unweighted mean was .63, and the median was .61. Pearson was

collecting correlation coefficients because he wanted to know the degree to which inoculation against smallpox saved lives. His own rough and ready summary of his meta-analysis of six studies was that there was a .6 correlation between inoculation and survival—a truly huge effect.

When Karl Pearson quantitatively summarized six studies of the effects of smallpox inoculation, a meta-analysis was an unusual thing to do. Recently, however, there has been an explosion of meta-analytic research syntheses across the social sciences such that a rapidly increasing proportion of all reviews of the literature are in the form of quantitative reviews, i.e., meta-analyses. The REAP analyses are recent examples of this trend: although far from the first meta-analyses in arts education, the technique is unusual in this field. Despite its increasing frequency in the literature, however, meta-analysis is not without controversy and criticism.

Criticism of Meta-Analysis

> *No one pretends that democracy is perfect or all-wise. Indeed, it has been said that democracy is the worst form of government except all those other forms that have been tried from time to time.*
>
> SIR WINSTON CHURCHILL
> Speech in the House of
> Commons, November 11, 1947

Churchill's statement about democracy perhaps offers a useful analogy when considering the value of meta-analysis. Compared to other methods of research synthesis, even taking into account any imperfections of the technique, it is the method least likely to lead to error in synthesis of complex research domains. Although some researchers see the enormous increase in the number of meta-analytic literature reviews as representing a giant stride forward in the development of the behavioral, educational, and social sciences generally, others see it as signaling a lemming-like flight to disaster. When some three dozen scholars in psychology were invited to respond to a meta-analysis of studies of interpersonal expectancy effects,[20] much of the commentary spontaneously addressed the methodological aspects of meta-analytic procedures. Judging from such reactions to the meta-analytic enterprise, it would be surprising not to find at least some arts education researchers who take the more pessimistic view. The next section briefly summarizes six features of meta-analysis and criticisms related to them. More detailed explications are available elsewhere.[14,21]

1. Sampling Bias

Meta-analyses synthesize research that has been conducted. Consequently, identifying as many of the studies that exist through a comprehensive, redundant, and systematic search is critical to insuring an unbiased sample and a valid analysis. A criticism of meta-analysis holds that the studies retrieved do not reflect the population of studies conducted. One version of this criticism is that the probability of publication is increased by the statistical significance of the results, so that published studies may not be representative of the studies conducted. This is a well-taken criticism, although it applies equally to more traditional narrative reviews of the literature. Another version of this criticism relates to the exclusion of qualitative results from the meta-analytic review.

An exhaustive, systematic search with redundant channels is the best way to minimize sampling bias. The REAP researchers, for example, used seven computer databases, reviewed reference lists of acquired articles, contacted authors in the field, and hand-searched fifty years of journals in the fields of interest (see the methods sections of the individual articles in Winner and Hetland[1]). Thus, the reviews identified and included a great deal of the "fugitive" literature.

Just as publication bias was addressed by this exhaustive search, the issue of the contribution of non-quantitative research is also partially addressed, since most qualitative research on the questions was certainly identified. But since the meta-analyses combine quantitative findings, do they truly represent an unbiased sample of the research conducted on questions of arts transfer? It is true that unless qualitative results can be represented as a numerical value (e.g., as they were in Shirley Brice Heath's qualitative study,[22] which was included in the Winner and Cooper[23] analysis summarized in this volume), they are not directly represented in the average effect sizes derived through meta-analysis. However, that neither reduces the contribution of qualitative studies nor the value of the meta-analysis.

We see qualitative research as having a fundamentally different purpose than quantitative research does, in the ongoing investigation of psychological and behavioral phenomena. Its importance in review and summary of the field is more interpretive than cumulative. Qualitative research does not attempt to calculate the size of a relationship, but rather seeks to inform us through triangulated examination of phenomena in lived contexts, "thick" description, and categorical analysis of the nature and dimensions of relationships. Such studies may inform interpretations of numerical results and contribute hypotheses that can be tested quantitatively—as part of the qualitative study or in later quantitative study. The relationship between qualitative and

quantitative research is often iterative, focusing on problems first in one way, then the other, as befits particular puzzles emerging at any time. Thus, the results of any qualitative study located by a search, as well as those studies rejected for other reasons, sometimes help to explain why the numerical findings turned out as they did, and suggest avenues and techniques for future investigation. (Quantitative methods such as contrasts explore similar questions through data-analytic procedures when specific directional hypotheses can be advanced. See below.)

2. Loss of Information

The intent of meta-analyses is to cumulate and explain research, and as a result, they necessarily reduce the complexity and nuances found in individual studies. One criticism has been that summarizing a research domain by a single value, such as a mean effect size, loses valuable information. However, comparing studies through contrasts, which is an effort to understand differences between study results, is as much a part of meta-analytic procedures as is summarizing the average results of the set of studies. We should also note that all models—qualitative as well as quantitative—reduce the complexity of natural objects in order to perceive their essential elements more clearly. Even within a single study, psychologists have historically found it quite helpful to compute the mean of the experimental and control groups, despite the fact that computing a mean always involves a "loss of information." A certain reduction of complexity, if it correctly identifies what is essential, is valuable. Thus, eschewing quantitative study or synthesis is not a reasonable solution for understanding the arts. Rather, arts researchers need to embrace scientific rigor in all research paradigms, in an effort to identify correctly which elements are essential, how they can best be measured, and what they might imply for learning and development in the arts or in areas to which the arts might transfer, such as are synthesized by the REAP analyses.

The REAP analyses conducted extensive contrast analyses of variables that might systematically moderate effect sizes. Examining moderator variables across a body of studies offers information that no single study can provide and guides the directions of future research. For example, Podlozny[24] examined nine such variables in her analysis of classroom drama, and Hetland's "instruction" analysis[25] examined seventeen.

3. Heterogeneity of Method and Quality

Meta-analysts combine studies that operationalize independent and dependent variables differently and employ different types of sampling units

(e.g., people, animals, or classrooms). Critics sometimes question combining such diverse studies. Meta-analysts have nicknamed this issue the "apples and oranges" problem. Can results as diverse as, for example, number correct on paper and pencil tests, reaction time, and solving contextual problems in the real world be combined? That depends on whether the studies address a similar research question. One would not want to call an apple an orange, but it makes perfect sense to reflect on apples and oranges mixed in one bowl when one wishes to generalize to fruit.[26,14] Well-done meta-analyses take such differences into account by treating them as moderator variables.

Meta-analyses are also criticized for throwing together good and bad studies. Such a critique has merit: any empirical finding is only as informative as its methods are rigorous. However, as meta-analysts, we tend to favor erring on the side of inclusion, for three reasons. First, the alternative to meta-analytic summary of such flawed work is ignoring data. Early in the developmental stages of a research area, and particularly in complex naturalistic settings such as those in which educational research is usually conducted, studies may have more threats to reliability and validity than do studies conducted under laboratory conditions or with the benefit of earlier studies that inform us about which variables systematically affect results. Early literatures *are the evidence* for the case so far, and they deserve summary and interpretation. Far better to include weak results in an analysis than to exclude single studies, which, without synthesis, may be interpreted as more conclusive than they are. Current claims that justify the arts as generically beneficial to academic test scores are a case in point: these claims are based on a few prominent studies that are cited repeatedly. Second, it is difficult to define "bad studies"—the studies of my "enemies," as Glass et al.[10] have put it. Third, exclusion of weak studies is not necessary, because analysts can deal with the problem quite simply by weighting studies by quality. Such weighting includes a weight of zero for the truly terrible study.[14,27]

The REAP analyses assessed quality in a variety of ways and found that effect sizes of both high- and low-quality studies were similar, so that weighting was unnecessary.

4. Problems of Independence

Sometimes the same subjects generate multiple effect sizes within the same study, often creating a problem for significance testing in particular. Technical procedures are available for adjusting for such nonindependence, and the REAP reviewers used several of these.[13,27] More subtle problems of possible nonindependence arise because different studies conducted in one labora-

tory may yield results that are more correlated with each other than with different studies conducted in another laboratory. In other words, there may be "laboratory effects."[29–33] These can be handled by treating laboratory effects as moderator variables, and by analyzing research domains by laboratory as well as by study.[32,14]

5. Exaggeration of Significance Levels

As explained earlier, there are a variety of ways to assess confidence about the average effects found in meta-analyses. A criticism related to establishing confidence through significance testing is based entirely on a misunderstanding of the fundamental equation of data analysis, i. e., *Significance Test* = *Effect Size* × *Study Size*. The criticism is that, as more and more studies are added to a meta-analysis, the results are more and more significant. That is certainly true, but it is difficult to perceive as a negative feature or as anything other than a mathematical fact. Certainly one would expect that, as the number of studies combined increases, our ability to generalize the results to other studies should be more reliable.

6. Small Effects

There are a host of statistics by which the strength of relationship between "treatment" and "outcome" can be indexed. Two of the most common are *d* and *r*. *d* is a standardized difference between means that is reported in standard deviation units (σ). Psychometricians often employ *d*, because they can interpret it readily as showing how far and in what direction a given treatment "pushes" performance along the normal distribution curve. The *d* also readily translates to IQ points, where the mean centers on one hundred and the SD is known for common tests.[34] Thus, a *d* of .50 is half a standard deviation unit, or about 8 points on the Stanford-Binet: Fourth Edition.[35]

The effect size *r*, used in the REAP analyses, indexes the correlation between treatment and outcome and ranges from −1.0 to +1.0, with the 1.0s showing a perfect correlation between two variables. An *r* can be used whenever *d* can, but not the reverse—*d* is meaningless for experiments employing three or more conditions, because a difference between means cannot be computed for more than two groups. In addition, *r* is more readily interpreted by the lay reader, since it can easily be transformed into a Binomial Effect Size Display (BESD).[14, 36–39]

Because algebraic transformations can readily transform one index to another (e.g., an *r* at low values such as those found in the REAP analyses can often be doubled to get a rough estimate of an equivalent *d*), the choice of effect-size statistic is mainly based on ease of interpretation (which varies by

what is familiar to particular audiences) and on the ability to compute the chosen statistic for the body of primary studies being reviewed.

The "small effects" criticism of meta-analysis holds that the results of socially important syntheses show only "small effects" because the r^2s obtained are small. This criticism has been addressed in detail elsewhere, where it has been shown that r^2s of nearly zero can save thirty-four lives per thousand, e.g., in the Physicians' Aspirin Study.[40,41] Thus, the practical value of an effect of any size must be considered in relation to the importance of the outcome compared to the cost (in effort and dollars) of the intervention required to produce it. If a small effect represents an average increase of a few points on a standardized test, as in the Winner and Cooper analysis summarized in this volume, that may not be of practical consequence. However, if it represents the effect of even a few students staying in school as opposed to dropping out as a result of a program easily implemented in many classrooms, that is of great consequence. Even small effects can be important in education, where progress is most often incremental.[13]

Benefits of Meta-Analysis

There are several fairly obvious benefits of meta-analysis. Our quantitative summaries of research domains using meta-analytic procedures are likely to be more complete, more explicit, and more powerful than qualitative reviews, in the sense of decreasing "type II errors," that is, errors of overlooking real effects. Moderator variables are more easily spotted and evaluated in a context of a quantitative research summary, thereby aiding theory development and increasing empirical richness. For example, Podlozny's[24] investigation of program duration suggests that more research comparing programs with different session lengths and number would inform decisions about how best to design classroom drama to increase verbal learning. For all these reasons, meta-analyses assist in the process of cumulation. There are also some less obvious benefits.

1. Decreased Overemphasis on Single Studies

As research literatures grow ever larger, science must cumulate efficiently, so that next studies can build upon previous ones and so that policy and practice can rely on research to help address pressing puzzles and promises identified anecdotally, intuitively, or empirically. One less obvious benefit of meta-analytic methods that will accrue to the science of educational psychology is the gradual decrease in the overemphasis on the results of a single study. Single studies cannot tell us enough about behavioral phenomena to generalize responsibly to other settings and populations, because they are

necessarily conducted in one setting with one sample of subjects. Their external validity—that is, their meaning in relation to predicting effects for a broader population of interest—cannot be as great as a group of similar studies carried out on a single research question. Thus, synthesis of research literatures is critical for understanding to advance.

There are good sociological grounds for our preoccupation with the results of a single study. Those grounds have to do with the reward system of science in which recognition, promotion, and reputation depend on the results of the single study, the smallest unit of academic currency. The study is "good," "valuable," and, above all, "publishable" when $p \leq .05$ and not when $p > .05$. Our disciplines would be further ahead if we adopted a more cumulative view focused on results of syntheses of entire research literatures. With such a view, the impact of any one study would be evaluated less on the basis of p levels, which vary by the size of the samples employed, and more on the basis of effect size. Such a view would lead us to evaluate any new study by its impact on and comparison to the average effect size synthesized from a body of literature. For example, new studies on the "Mozart effect" can now be compared to $r = .25$, the average effect of the experiments synthesized by Hetland.[29]

2. Identifying Continuing Trends in Research

Related to the problem of overemphasis on single studies is the problem of "differentiation drive," a motivational state (and possibly even a trait) sometimes found among scientists in all fields. This is the drive to be more different, more ahead, more right, and more unique than others. Another reflection is the occurrence of "renomination," the mechanism by which a well-known process is given a new name in hopes of effecting "concept capture"—the ownership of a concept claimed by virtue of renaming it. Differentiation drive keeps us from seeing the similarity of our work to the work of others and labors against the responsible cumulation of scientific evidence. Thus, researchers begin again from scratch with each new iteration of a question, and our collective memories of previous research trends fade. Meta-analysis is an antidote, because it seeks the similarities along with the distinctions in bodies of research literature.

3. Emphasis on the Practical Interpretation of Results

Despite the growing awareness of the importance of estimating effect sizes, there is a problem in evaluating various effect-size estimators from the point of view of practical usefulness.[41] Rosenthal and Rubin[39,42] found that neither experienced behavioral researchers nor experienced statisticians had a good intuitive feel for the practical meaning of common effect-size estimators

and that this was particularly true for such squared indices as r^2, *omega*^{2}, *epsilon*^{2}, and similar estimates.

The Binomial Effect Size Display (BESD) is a way of showing the practical importance of any effect indexed by a correlation coefficient (that is, an *r*). The correlation is shown to be a simple difference in outcome rates between the experimental and the control groups in this standard table, which always adds up to column totals of one hundred and row totals of one hundred.[39] We obtain the BESD from any obtained effect size *r* by computing the treatment condition success rate as .50 (a level equivalent to chance) plus $r/2$. Similarly, the control condition success rate is .50 minus $r/2$. Thus an *r* of .20 yields a treatment success rate of $.50 + .20/2 = .60$ and a control success rate of $.50 - .20/2 = .40$. A BESD of these results looks like this:

	Success	Failure	Σ
Treatment	60	40	100
Control	40	60	100
Σ	100	100	200

Had we been given the BESD to examine before knowing *r* we could easily have calculated it mentally for ourselves; *r* is simply the difference between the success rates of the experimental versus the control group ($.60 - .40 = .20$).

As social and educational researchers, we are not used to thinking of small *r*s as reflecting effect sizes of practical importance. The results of a famous meta-analysis of psychotherapy outcome studies reported by Smith, Glass, and Miller[43] were dismissed by many psychologists because, at $r = .32$, the results explained only 10% of the variance! Interestingly for behavioral and educational researchers, the magnitude of the effect of psychotherapy was substantially greater than the effects of a good many breakthrough medical interventions (e.g., the *r* for the Physicians' Aspirin Study mentioned earlier was .034). One desirable result of comparing biomedical effect size estimates to those achieved in the social sciences is to make those of us working in the "softer" sciences less pessimistic about the magnitude and importance of our research results.[44,45] We need to reevaluate our interpretation of small effects. Usually, changes in education occur incrementally and not as dramatic sea-changes. We need to understand the results of our research accordingly.

4. The Demise of Dichotomous Significance Testing

Far more than is good for us, social, behavioral, and educational scientists operate under a dichotomous null hypothesis decision procedure in which evidence is interpreted as anti-null if $p \leq .05$ and pro-null if $p > .05$. If our dissertation *p* is $< .05$ it means joy, a Ph.D., and a tenure-track position at a major

university. If our $p > .05$ it means ruin, despair, and our advisor suddenly thinking of a new control condition that should be run. That attitude is not helpful. When researchers report only whether $p \leq .05$ rather than reporting exact *p*s, "likelihood" is assessed as a cliff. Such reporting of results equates as equally likely probabilities of, for example, $p = .06$ and $p = .50$ (one-tailed), when they are not equivalent. A $p = .06$, one-tailed, indicates that if the null hypothesis were true we would find *t* so large in the predicted direction only 6% of the time. A $p = .50$, one-tailed, however, indicates that if the null hypothesis were true we would find a *t* so large in the predicted direction 50% of the time. Clearly, God loves the .06 nearly as much as the .05. As a matter of fact, two .06 results are much stronger evidence against the null than one .05 result, and 10 *p*s of .10 are stronger evidence against the null than 5 *p*s of .05. Indeed, there is good information that God views the strength of evidence for or against the null as a fairly continuous function of the magnitude of *p*. We refer those unwilling to accept our theological arguments to the Task Force on Statistical Inference of the American Psychological Association, which recommends reporting exact *p*s so that distinctions can be assessed along a continuum, rather than at an arbitrarily defined cutoff between "true" and "false."[44]

5. The Increased Recognition of Contrast Analysis

Meta-analytic questions are basically contrast questions. *F* tests with $df > 1$ in the numerator or x^2 with $df > 1$ are useless in meta-analytic work. That leads to the following additional scientific benefit: Meta-analytic questions require precise formulation of questions, and contrasts are procedures for obtaining answers to such questions. Although most statistics textbooks describe the logic and the machinery of contrast analyses, one still sees contrasts employed all too rarely. That is a real pity given the precision of thought and theory that they encourage, and (especially relevant to these times of publication pressure) given the boost in power conferred with the resulting increase in .05 asterisks.[34–36]

6. Meta-Analytic Procedures Are Applicable Beyond Meta-Analyses

Many of the techniques of contrast analyses among effect sizes, for example, can be used within a single study.[34] Computing a single effect size from correlated dependent variables, or comparing treatment effects on two or more dependent variables serve as illustrations.[27]

7. The Decrease in the Splendid Detachment of the Full Professor

Meta-analytic work requires careful reading of research and moderate data analytic skills. We cannot send an undergraduate research assistant to the

computer or the library with a stack of 5 x 8-inch cards to bring us back "the results," which seems often to have been done with narrative reviews. With meta-analysis, the reviewer must get involved with the actual data, and that is all to the good of science.

Conclusion

The benefits of meta-analysis are to be found in virtually all areas of scientific study, from Pearson's 1904 meta-analysis of vaccination benefits to REAP's meta-analytic studies of arts education summarized and commented upon in these pages. Whether samples are huge or tiny in size, whether there are hundreds of studies or only two, meta-analytic procedures are available, when wisely used, to assist in the development of science, medicine, education, and human welfare. The arts must participate in the process of use and refinement of such methods or be left at the periphery of those phenomena judged important in human cultures. While quantitative study and synthesis are not the only means by which arts can be understood and validated, they are an important one. The REAP meta-analyses have shown just where causal conclusions can be drawn about the effects of arts on other forms of learning, and they also caution against drawing facile causal inferences that have not yet been supported by rigorous empirical study. Such careful analyses advance our understanding of the problems of transfer from arts to non-arts domains and, perhaps even more importantly, demonstrate that the arts are more than a sideshow. They need to be studied seriously through all available means, just as are other human inventions and behaviors that, together, constitute our humanness.

Endnotes

1. Winner, E. and Hetland, L. (eds.). (2000). The arts and academic achievement: What the evidence shows. *Journal of Aesthetic Education* 34 (3/4).
2. Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York: McGraw Hill.
3. Berlyne, D. E. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation*. New York: John Wiley and Sons.
4. Berlyne, D. E. and Madsen, K. B. (1973). *Pleasure, reward, preference: Their nature, determinants, and role in behavior*. New York: Academic Press.
5. Chalmers, B. (1976). Musicians and mathematical measurement: Changing attitudes. *The Australian Journal of Music Education* 19, 13–16.
6. Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper Perennial.
7. Cooper, H. M. (1989;). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.
8. Cooper, H. and Hedges, L. V. (eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage.
9. Cooper, H. and Hedges, L. V. (1994). Potentials and limitations for research synthesis. In H. Cooper

and L. V. Hedges (eds.). *The handbook of research synthesis*. New York: Russell Sage, 521–529.

10. Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
11. Hedges, L. V. and Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
12. Hunter, J. E. and Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
13. Light, R. J. and Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
14. Rosenthal R. (1991a). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage Publications.
15. Hedges, L. V. (1994). Fixed effects models. In H. Cooper and L. V. Hedges (eds.). *The handbook of research synthesis*, 285–299). New York: Russell Sage.
16. Rosenthal, R. (1995b). Writing meta-analytic reviews. *Psychological Bulletin* 118 (2), 183–192.
17. Raudenbush, S. W. (1994). Random effects models (pp. 301–321). In H. Cooper and L. V. Hedges (eds.). *The handbook of research synthesis*. New York: Russell Sage.
18. Gosset, W. E. (Student) (1908). The probable error of a mean. *Biometrika* 6, 1–26.
19. Pearson, K. (1933). Appendix to Dr. Elderton's paper The Lanarkshire milk experiment, *Annals of Eugenics* 5, 337–338.
20. Rosenthal, R., and Rubin, D. B. (1978a). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences* 3, 377–386.
21. Rosenthal, R., and Rubin, D. B. (1978b). Issues in summarizing the first 345 studies of interpersonal expectancy effects. *The Behavioral and Brain Sciences* 3, 410–415.
22. Heath, S. B. (1998). Youth development and the arts in nonschool hours. Grant-makers in the Arts 9 (1), 9–16, 32.
23. Winner, E. and Cooper, M. (2000). Mute those claims: No evidence (yet) for a causal link between arts study and academic achievement. *Journal of Aesthetic Education* 34 (3/4), 11–75.
24. Podlozny, A. (2000). Strengthening verbal skills through the use of classroom drama: A clear link. *Journal of Aesthetic Education* 34 (3/4), 239–275.
25. Hetland, L. (2000b). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education* 34 (3/4), 179–238.
26. Glass, G. V. (1978). In defense of generalization. *The Behavioral and Brain Sciences* 3, 394–395.
27. Rosenthal, R. (1991b). Quality-weighting of studies in meta-analytic research. *Psychotherapy Research* 1, 25–28.
28. Rosenthal, R. and Rubin, D. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin* 99, 400–406.
29. Hetland, L. (2000a). Listening to music enhances spatial-temporal reasoning: Evidence for the "Mozart effect." *Journal of Aesthetic Education* 34 (3/4), 105–148.
30. Jung, J. (1978). Self-negating functions of self-fulfilling prophecies. *The Behavioral and Brain Sciences* 3, 397–398.
31. Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
32. Rosenthal, R. (1969). Interpersonal expectations. In R. Rosenthal and R. L. Rosnow (eds.). *Artifact in behavioral*

research. New York: Academic Press, 181–277.

33. Rosenthal, R. (1976). Experimenter effects in behavioral research (enlarged ed.). New York: Halsted Press.
34. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
35. Spreen, O., and Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford.
36. Rosenthal, R., and Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge: Cambridge University Press.
37. Rosenthal, R., and Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.
38. Rosenthal, R., Rosnow, R. L., and Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, United Kingdom: Cambridge University Press.
39. Rosenthal, R., and Rubin, D. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology* 74, 166–169.
40. Steering Committee of the Physicians Health Study Research Group. (1988). Preliminary report: Findings from the aspirin component of the ongoing physicians' health study. *New England Journal of Medicine* 318, 262–264.
41. Cooper, H. M. (1981). On the significance of effects and the effects of significance. *Journal of Personality and Social Psychology* 41, 1013–1018.
42. Rosenthal, R., and Rubin, D. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology* 9, 395–396.
43. Smith, M. L., Glass, G. V., and Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
44. Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist* 45, 775–777.
45. Rosenthal, R. (1995a). Progress in clinical psychology: Is there any? *Clinical Psychology: Science and Practice* 2, 133–150.
46. Wilkinson, L., and the Task Force on Statistical Inference (1999, August). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* 54 (8), 594–604.

The Relationship Between Arts and Academic Achievement: No Evidence (Yet) for a Causal Relationship. A Summary of a Meta-Analytic Study

ELLEN WINNER

In today's educational climate, basic academic skills are valued and the arts are considered frills. When budgets are tight, as they always seem to be, the arts are typically the first subjects to be cut or trimmed down. In reaction to the marginalization of the arts, arts educators and other arts advocates have argued with increasing vociferousness that the arts are a means to improved basic academic skills. For example, according to a 1995 report by the President's Committee on the Arts and Humanities, "teaching the arts has a significant effect on overall success in school."[1] The report justifies this claim by noting that both math and verbal SAT scores are higher for high school students who take the arts compared to those who take none. And in 1999, testimony was presented to the United States House of Representatives that "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."[2] Senator Alan Simpson stated as a given, "A love of arts helps the learning process . . . all studies tell us that."[3]

The arts are also said to have social payoffs. The arts advocacy group Americans for the Arts has stated that the arts keep children off the streets and improve their self-concepts. A letter to the *New York Times* in August 2000, by the manager of an organization that presents concerts in inner-city Los Angeles schools, stated that "the principal benefit of instruction in music is that it increases self-esteem and self-worth, especially for . . . underprivileged and bilingual children. . . . It is also well established that the child who has a good

sense of self-esteem is not likely to commit acts of violence or use drugs."[4] The logic here is clear: we go from music to self-esteem to lowered crime.

At Project Zero, at the Harvard Graduate School of Education, Lois Hetland and I looked systematically at the evidence for the claims that the arts improve academic skills. In this project, which we called Reviewing Education and the Arts Project (or REAP), we conducted a comprehensive search for all studies since 1950, published and unpublished, that examined the relationship between arts study and academic achievement. Note that we did not examine the evidence for the claimed social outcomes of the arts such as improved academic motivation, higher school attendance, etc. We looked only for studies that actually measured some kind of non-arts, cognitive outcome; thus, we did not include studies in which teachers said they *believed* students' cognitive skills were boosted by the arts. And we looked only for studies with a control group; thus, we did not include studies that showed improvement over time with arts study but lacked a control group with no arts study measured over the same period of time. Almost two hundred studies met our criteria for inclusion.

We then classified the studies by art form and non-arts outcome (e.g., visual arts and reading, music and math, etc.). For each group of comparable studies, we conducted a series of meta-analyses. Meta-analysis is a statistical technique that makes it possible to synthesize, quantitatively, a body of studies (see "Meta-Analysis: Its Use and Value in Arts Education," pp. 1–16, this volume). Our meta-analyses are in no way intended to be presented as the final answer on the question of whether arts transfers to academic outcomes. Rather, our goal was to take stock of what we know so far and to provide guidance for future research on this question.

The heart of a meta-analysis is the calculation of an "effect size" for each study; effect sizes are then combined and compared across studies. Effect sizes show the strength of the relationship between two variables; in this case, between some form of arts study and some kind of academic outcome. There are several different kinds of statistics one can use to show an effect size, and we used the statistic called *r*, as recommended by Robert Rosenthal.[5] This effect size ranges from −1.0 to +1.0. An effect size of .10 is considered small, .24 is considered medium, and .37 is considered large.[6] A meta-analysis can do several things that a traditional narrative review of the literature cannot do, and these benefits are spelled out in more detail in the article by Rosenthal and Hetland just mentioned. For one thing, a meta-analysis reveals an average effect size over a group of studies. Thus it can specify the average relationship between arts and academic outcomes over a body of studies rather than in just one study. Second, a meta-analysis tells us the strength and reliability of this

average effect size. A meta-analysis also allows one to test hypotheses. One can code studies for important variables (e.g., length of arts study or whether arts were taught separately or integrated into the curriculum), and then one can test whether any one of these variables influences the size of the effect. Thus, far more than a summary, a meta-analysis makes it possible to explore hypotheses using previously conducted studies as the unit of analysis.

Three motivations guided our research. Our first motivation was theoretical. Is the kind of learning and thinking stimulated by the arts the same as (and hence generalizable to) learning and thinking in non-aesthetic domains? Our second motivation was educational and applied. Can the arts be used as a tool to help children learn better in other school subjects? Our third motivation concerned policy. We wanted to examine the relationship between advocacy (the strong claims made about the power of the arts to transform academic achievement) and science (what the research evidence actually shows).

Our findings revealed a mixed picture. We found three cases of clear causal links between studying an art form and some kind of non-arts academic outcome: music listening and spatial reasoning, music instruction and spatial reasoning, and classroom drama and verbal skills. The music findings are spelled out in more detail in the paper by Lois Hetland (this volume). The drama findings can be found in the paper by Ann Podlozny (this volume). In what follows, I present the areas in which we were forced to conclude that there is, at least as yet, no strong evidence for a causal relationship between arts study and academic achievement.

Does Studying the Arts Enhance Academic Achievement?

Our first meta-analytic study synthesized studies that examined the relationship between academic achievement and what we called "multi-arts" study (the study of some combination of visual arts, music, drama, and dance).[7] In the studies synthesized, students were either exposed to the arts as separate disciplines, or they received such exposure but were also given an arts-integrated academic curriculum. Unfortunately, few of the studies explained in much detail anything about the nature and quality of the arts instruction, or about what it really meant to study an academic subject with arts integration. Academic achievement in these studies was measured primarily in the form of test scores (composite verbal and math scores, or verbal and math scores separated) but also sometimes in the form of overall academic grades or the receipt of academic awards. Here is what we found.

We first examined the correlational studies—studies that compared the academic profile of students who do and do not study the arts either in

school or in after-school programs. For example, we included James Catterall's study in the analysis, in which he demonstrated that students who are highly involved in the arts in middle and high school outperform those who are not involved in the arts on a multitude of academic indicators; this relationship holds even for students in the lowest SES quartile of the United States.[8] These students earned higher grades and test scores than those not arts-involved. The high arts students were also less likely to drop out of high school, and they watched fewer hours of television than did the low-arts students. We included Shirley Brice Heath's study showing that at-risk students who participate in after-school arts organizations for at least nine hours a week over the course of at least a year are ahead of a random national sample of students on a wide range of academic indicators: their school attendance is higher, they read more, and they win more academic awards.[9] And we included data from the college board revealing that the average SAT scores of students with four years of high school arts were higher than the scores of those who took no arts courses at all in high school.[10]

Three meta-analyses synthesizing the correlational studies were performed, each on a different academic outcome (composite verbal and math outcomes summed; verbal outcomes; math outcomes). All three showed a clear relationship between academic achievement and studying the arts. When we examined the five studies examining composite outcomes (verbal and mathematics achievement indicators summed), we found a small but highly significant relationship ($r = .05$). When we examined the eleven studies examining verbal outcomes (and this included ten years of the College Board data), we found a small-to-medium relationship ($r = .19$) that was also highly significant. And when we examined the eleven studies examining math outcomes (and this included ten years of the College Board data), we again found a small-to-medium relationship ($r = .10$) that was highly significant. All three effect sizes were significantly different from zero, as shown by a *t*-test.

These three meta-analyses show that students in the United States who choose to study the arts are students who are also high academic achievers. But because the studies on which these meta-analyses were based were correlational in design, they allow no causal inferences. One plausible interpretation of the findings is that high academic achievers (no matter what their SES) may be more likely to choose to study the arts than low academic achievers. This could occur for several reasons. High academic achievers may attend schools strong in both academics and the arts; they may come from families that value both academics and the arts; or they may have high energy and thus have time for and interest in both academics and the arts. One piece of evidence for the drive hypothesis comes from the study by Shirley Brice Heath.

Heath actually studied three groups of students, not only those in after-school arts organizations but also those in after-school sports or community service organizations. All three groups were intensively involved in their choice of organization. Heath allowed us access to her unpublished data, and we compared the likelihood of winning an academic award for the arts versus the sports students. While both groups were significantly more likely to win an academic award than a random national sample of students, there was no difference between these two groups. Eighty-three percent of the group of 143 arts-involved students and 81% of the sports-involved students won an academic award, compared to 64% of the national sample. This finding suggests that students who immerse themselves in after-school arts may do well in school not because of their exposure to arts but because they are the type of student with the drive to spend at least nine hours a week mastering a skill, whether this be arts or sports.

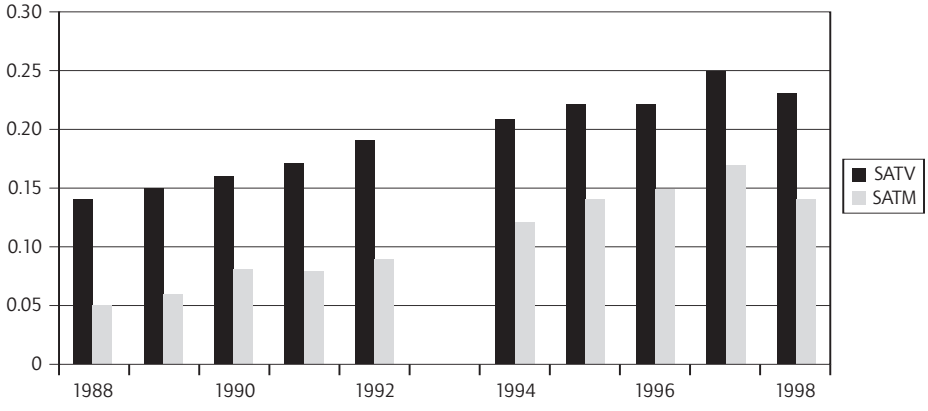
Further support for the drive hypothesis comes from a comparison pointed out by Elliott Eisner.[11] He compared the SAT advantage of students taking four years versus one year of arts to that of students taking four years versus one year of an elective academic subject such as science or a foreign language. Students who specialized in any subject, whether arts or an academic elective, all had higher SATs than those who had only one year in that subject. This supports the idea that students who specialize or focus have higher energy and/or motivation than those who do not, and that this higher energy or drive accounts for their higher academic achievement. However, he also found that the SAT advantage was considerably greater for students specializing in an academic subject than in the arts. For example, in 1998, while students with four years of arts had SATV scores that were forty points higher than those with only one year of arts, those with four years of a foreign language had SATV scores that were 121 points higher than those with only one year of foreign language. Similarly, while students with four years of arts had SATM scores that were twenty-three points higher than those with only one year of arts, those with four years of science had SATM scores that were fifty-seven points higher than those with only one year of science.[12]

Another, more cynical, possible explanation for the link between arts study and academic achievement comes to mind. Perhaps our highest achievers study the arts in order to enhance their chances of admission to selective colleges. One piece of evidence for this idea is that the academic profile of students choosing to take the arts has risen consistently over the last decade. When we plotted the relationship between SAT score and taking four years of arts in high school (compared to taking no arts), we found that this relationship grew stronger each year beginning with 1988, the first year in which

the data were available, and continuing through 1999 (the last year of data we examined). Rising effect sizes for the arts-SAT relationship are shown in Figure 1. Thus, the comparative SAT advantage for students with four years of arts grew greater each year. As our most selective colleges become more competitive each year, and a record of straight As and 800s on SATs is no longer distinctive enough to ensure admission, students may feel they need to build résumés showing strength in a non-academic area such as an art form.

Do we see the same link between self-selection into the arts and academic achievement in other countries? One study from the Netherlands and another from the United Kingdom show us that we do not. In the Netherlands, Folkert Haanstra found that students who take the arts in high school to prepare for a national exam that includes the arts attain the same educational level as those with no arts electives. This study, which controlled for students' SES, shows that in the Netherlands, taking the arts in high school does not predict ultimate educational level attained.[13] And in the United Kingdom, John Harland found that the greater the percentage of arts courses taken in high school, the poorer the performance on national exams at the end of secondary school.[14] Harland explained this finding by noting that in the United Kingdom, the only students who are permitted to prepare for more than one arts subject for their secondary school exams are those who are academically weak. This contrasts sharply with educational policy in the United States. We are likely to steer academically weak students into remedial academic courses, not into the arts. The comparison between the findings in the United States with those in the Netherlands and the United Kingdom suggests that the relationship between arts study and academic achievement is

Figure 1: Effect sizes showing relationship between four year arts study in high school and SAT scores. Effect sizes grow steadily larger from 1988–1998.



not a causal one but instead reflects different cultural values about who should study the arts.

But perhaps two factors are at work—both self-selection into the arts and then a subsequent causal relationship, in which studying the arts actually causes academic achievement to grow. We were able to test this hypothesis by examining the data in James Catterall's study, mentioned earlier. He reported longitudinal data on students who self-selected into the arts in eighth grade and remained highly involved in the arts through the twelfth grade. If both factors were at work, we would expect the effect sizes showing the strength of the relationship between arts involvement and academic performance to rise over the years. But we found no change. The effect size showing the relationship between studying the arts and academic achievement was $r = .18$ for students in the eighth grade, and this effect size remained unchanged in tenth and twelfth grade. Although these data come from only one study, they come from a very large-scale study: there were 3,720 students who were highly involved in the arts from the eighth through twelfth grades, and the same number who were not particularly involved in the arts over that time period. The data fail to support the view that the arts are what is causing the academic achievement of these students to be higher than that of students relatively uninvolved in the arts.

While the correlational studies, and the meta-analyses synthesizing them, do not permit causal inferences, studies with an experimental design do allow such inferences. We examined two bodies of experimental studies testing the causal claim that when students study the arts, their academic achievement rises. These studies compared academic performance before and after studying the arts. Typically these studies examined students at the elementary school level who had studied the arts for a year and who studied the arts both as separate disciplines and as integrated into the academic curriculum. The academic growth of these students was then compared to the growth of similar students not exposed to any special arts program.

We found twenty-four studies testing the hypothesis that verbal skills improve as a consequence of studying the arts, and fifteen studies testing the hypothesis that math skills improve. The meta-analysis performed on the verbal outcomes yielded an effect size r of .07. This effect size was not statistically significant. There was a 95% chance that given another body of similar studies, the average effect size might be zero. In addition, a t-test showed that the mean effect size found was not significantly different from zero. Moreover, the nineteen studies in which the arts were integrated into the curriculum yielded a mean effect size identical to that of the five studies in which the arts were only studied separately. Finally, when we weighted the importance

of each study by its sample size to arrive at a weighted mean effect size, the effect size was reduced to .01. Thus we had to conclude that we had found no evidence that studying the arts, including the arts integrated with academic subjects, resulted in enhanced verbal skills.

The meta-analysis performed on the math outcomes yielded an effect size of $r = .06$. Again we found a 95% chance that the average effect size from a new body of similar studies might be zero, and a *t*-test showed that the mean effect size was not significantly different from zero. In this case we could not statistically compare the studies with and without arts integration since all but two were based on an arts-integrated curriculum. Again, then, we had to conclude that we found no evidence that studying the arts, including the arts integrated with academic subjects, resulted in enhanced math achievement.

Thus we can see that studying the arts, and studying an academic curriculum in which the arts are somehow integrated, does not result in higher verbal and math achievement, at least as measured by test scores, grades, or winning academic awards. We should be very skeptical, then, when we hear advocacy claims that studying the arts improves academic performance. It may or it may not. Some studies find a causal link while others do not. When the studies are combined and synthesized in a meta-analysis, we must conclude that the evidence is just not there. Arts study has not yet been demonstrated to cause academic (verbal and math) growth.

Perhaps the problem lies with the kind of arts programs studied. In none of the studies could we find much of a description of the type of arts instruction received, and it is possible (indeed likely, given the state of arts education across our nation) that students in many of these studies received weak arts instruction. Thus, perhaps no transfer has been found because the arts have not been well enough taught. That is certainly a possibility, and researchers in the future need to describe very clearly the quality of the arts program with which they are testing the claim of transfer.

Perhaps the problem lies with the kind of outcome measures used. Almost all of the research has looked at standardized verbal and mathematics test scores. But is this really where we should be looking? Perhaps we should be looking for whether the arts help students to tolerate ambiguity and resist premature closure when solving a messy problem with no clear right or wrong answers. We found no research that examined this kind of "higher-order" thinking. Researchers would need to demonstrate first that this kind of thinking was actually promoted in an arts class and then test whether students who had first learned this kind of thinking in arts class went on to use it in an academic class with a very different kind of problem.

Finally, all too often the research has been atheoretical. That is, researchers have looked for a connection between arts learning and academic achievement without considering what underlying mechanism might make such a link possible. There are many different possible mechanisms. It is possible that the arts train certain kinds of cognitive skills (such as the higher-order thinking skill just described) and that these transfer. It is possible that the arts train certain kinds of working habits such as perseverance and high standards, and these then transfer. It is possible that the arts serve as an entry point into an academic subject, particularly for non-academically inclined students. And finally, it is possible that non-academically inclined students discover competence in an arts class, and this discovery leads to greater self-confidence and a more positive attitude about school, both of which could lead to greater academic achievement. It is difficult to justify an investigation of the possibility of transfer in the absence of a theory of the underlying mechanism that might be at work.

Does Studying the Arts Enhance Creative Thinking as Measured by Creativity Test Scores?

It has been claimed that studying the arts trains critical and creative thinking. As mentioned above, we found no studies testing this claim by anything other than standard paper and pencil creativity tests. We did find four studies comparing the creativity scores of students who took arts courses versus those who did not. With Erik Moga, Kristin Burger, and Lois Hetland, I carried out a meta-analysis of these studies and found a very small relationship between studying arts and verbal creativity test measures that was not statistically significant ($r = .05$).[15] We did find a small-to-medium-size relationship ($r = .19$) between studying arts and figural creativity tests (which themselves are visual tests), but again, this relationship was not significantly different from zero.

We suggest that standardized creativity tests are not the right kinds of outcomes to be using. Such tests may assess fluency and cleverness rather than deep creativity, such as the ability to find new problems to solve. A study by Jacob Getzels and Mihalyi Csikszentmihalyi published in 1976 demonstrated that art students whose work was judged creative achieved high scores on a measure of "problem-finding."[16] When given a still life to paint, these students spent time exploring and rearranging the object until they found a design problem that interested and challenged them. In contrast, those whose work was judged less creative were passive, and simply accepted the problem as given and drew the objects in the arrangement presented to them. Getzels and Csikszentmihalyi's study demonstrated a relationship between creativity and problem finding in the same domain—the visual arts. It should be possi-

ble to devise problem-finding measures in other domains and to then test whether students who become problem finders in an art form are more likely to be problem finders in a non-arts, academic area. Such a study, which has not, to our knowledge, been conducted, would put to a serious test the hypothesis that study in the arts trains creative thinking in the arts, which then transfers to other domains.

Does Studying Music Enhance Math Achievement?

In 1999, a study published in *Neurological Research* received considerable publicity. This study reported that piano keyboard training along with computer-based spatial training led to greater improvements in math than when spatial training was combined with computer-based English-language training.[17] We searched for other studies examining the power of music to stimulate mathematical thinking and found six. Kathryn Vaughn carried out a meta-analysis of these studies and found an average effect size of $r = .13$.[18] When Vaughn tested whether this average effect size was significantly different from zero, she found that the answer was yes at the probability level of ninety-four percent (just about but not quite at the conventional standard of ninety-five percent). We believe that these findings suggest that there may indeed be a causal link between some forms of music instruction and some forms of math outcomes. But we feel that no firm conclusions should be drawn at this point since the finding was based on only six results. Moreover, of these six results, only two yielded medium effect sizes ($r = .31, .20$), one yielded a small-to-medium effect size ($r = .17$), and the remaining three were below .10, the level considered to be small (and one of these was actually negative). Thus, more research on this question is needed before we can be sure about the result.

Can the Study of Either Visual Arts, Music, or Dance Be Used to Enhance Reading Achievement?

Claims have been made that an art form can be used as a way to enhance reading abilities in young children. The visual arts, music, and dance have each been asserted to be a tool for reading improvement. For instance, several programs in New York City have been set up to use the visual arts to improve the reading skills of remedial readers (The Guggenheim Museum's Learning to Read Through the Arts, Reading Improvement Through the Arts, and Children's Art Carnival). In these programs, remedial readers are given experience in the arts integrated with reading: they make art and read and write about what they make. The program claims that these children greatly improve their reading ability and that the arts are a way into reading for these children. But the evaluations of this program never included a control group

of children given the same amount of extra reading without the arts.[19] Thus, we cannot determine whether improvement in reading was due to the extra reading instruction, or to the fact that the reading was integrated with art.

We were able to find nine studies that had appropriate control groups.[20] In five of these, the arts were taught separately from reading, and for these we found a very small relationship between visual arts and reading ($r = .05$), which could not be generalized to new studies. Moreover, this effect was entirely due to reading readiness outcomes (which are themselves visual) and did not hold up for reading achievement outcomes.

We also found four reports in which visual arts were integrated with reading instruction, as in the New York programs, and here we found a medium-size relationship between integrated arts/reading instruction and reading outcomes ($r = .23$). However, this result could not be generalized to new studies. We concluded that we found only marginal support for the claim that integrating reading with the visual arts works better than teaching reading directly.

Thus, there is not much support for the claim that the visual arts are the entry point into reading, and it is likely that programs such as Learning to Read through the Arts work so well because the children receive intensive reading training outside of the regular curriculum. The appropriate comparison, if we wanted to find out about the effectiveness of the arts per se as an entry point, would be one between an arts and reading program versus a direct remedial reading program such as Reading Recovery. This kind of controlled comparison has not, to my knowledge, been carried out.

Music has also been claimed to be a way to improve reading skills, possibly because of the effect of learning music notation. Ron Butzlaff located six experimental studies testing music's effect on reading and performed a meta-analysis on these studies.[21] He found a mean effect size of $r = .18$. This average was based on quite varied effect sizes, and a statistical test showed that the effect size was not significantly different from zero. Thus, we had to conclude that there is as yet no evidence that music instruction serves as a way into reading.

It is difficult even to imagine a mechanism by which dance could help children with reading at the level of decoding, though perhaps enactment of stories though dance might improve the comprehension of these stories. There is a program in Chicago called Whirlwind that tries to improve children's basic reading skills through dance.[22] Among other things, the children in this program (first-graders) learn to put their bodies into the shapes of letters. The children were shown to improve in beginning reading skills relative to a control group. However, the control group did not get the same kind of letter training,

and so the comparison is not really a fair one. In addition, putting one's body in the shape of letters is not a very authentic use of dance. We searched for other studies and found four that tested the use of dance as a way into reading, this time with more appropriate control comparisons. Mia Keinanen, Lois Hetland, and I found only a small relationship between dance and reading ($r = .10$), and this effect size was not significantly different from zero.[23]

We had to conclude, thus, that there is no clear evidence that reading ability can be enhanced through teaching methods that incorporate the visual arts, music, or dance. While dance has not been shown to help reading, we did find three studies that together demonstrated that dance training enhances some forms of visual-spatial skill.[24] Such a relationship is of course plausible, since dance itself is a visual-spatial form of activity.

Instrumental Claims Are a Double-Edged Sword

We need to distinguish between core justifications for teaching the arts versus bonus justifications. Core justifications are the central reasons: they are about learning in the disciplines of the arts themselves. Bonus reasons are the side effects: enhanced learning in non-arts disciplines, which may or may not occur. It could be suicidal to justify the arts on the basis of bonus effects. If the arts are given a role in our schools because people believe the arts cause academic improvement, then the arts will quickly lose their position if academic improvement does not result, or if the arts are shown to be less effective than lengthy, focused instruction in the Three Rs in promoting literacy and numeracy. Instrumental claims for the arts are a double-edged sword. It is implausible to suppose that the arts can be as effective a means of teaching an academic subject as is direct teaching of that subject.

In addition, when bonus reasons become the primary justification for arts education, arts teachers may stray from teaching the arts in an authentic way and begin to teach the arts in a way that will enhance academic (rather than artistic) understanding. The Whirlwind dance program, used to teach reading by asking children to form their bodies into the shape of letters, is a case in point. We have seen researchers turn strings of music notations into multiplication problems and bill this as music education, the kind likely to improve math scores. And we have heard of teachers teaching the physics of sound in music class rather than the aesthetics of sound, or having students build musical instruments (because it may improve their spatial abilities) rather than learn to play these instruments.

It is high time to state the right arguments for the arts in our schools and to begin to gather the right kind of evidence for these arguments. The

best hope for the arts in our schools is to justify them by what the arts can do that other subjects cannot do as well, or cannot do at all.

The two most important reasons for studying the arts are to enable our children to be able to appreciate some of the greatest feats humans have ever achieved (e.g., a painting by Rembrandt, a play by Shakespeare, a dance choreographed by Ballanchine, a sonata by Mozart), and to give our children sufficient skill in an art form so that they can express themselves in this art form. The arts are the chief scholarly domains in which deep personal meanings can be recognized and expressed, often in nonverbal form.

In reaction to our work, arts advocates have said that we are just returning to "art-for-arts'-sake" arguments, and that these old arguments just won't wash. But this is an admission of defeat. If we can finally understand (as many other cultures have) that the arts are as important as the sciences, and that the purpose of education is to teach our children to appreciate the greatest of human creations, then the arts will have a strong hold in our schools. But if we let ourselves get brainwashed by today's testing mentality and come to believe that the arts are important only (or even primarily) because they buttress abilities considered more basic than the arts, we will unwittingly be writing the arts right out of the curriculum. We favor arts for the mind's sake, no less than science and math for the mind's sake.

An analogy from medicine may prove instructive. We all believe that vegetables are good for people's health. Until recently we believed that vegetables protected against colon cancer because of their high fiber content. But new research has failed to find support for this protective effect of high fiber. Does this mean that we should suppress the research because people will stop eating their beans? Of course not. It means we should go on to do more research to figure out just what eating vegetables really does that is of benefit. Does this new research provide us with an excuse to cut vegetables from our diet? Of course not. Vegetables can be delicious in their own right and probably provide us with vitamins that we cannot get from other foods. The challenge for researchers is to demonstrate the gains that come from eating vegetables that cannot be found from eating other kinds of foods. And so also the challenge for arts education researchers. We need to demonstrate the gains that come from studying the arts that our students do not get from studying math or science or English. We need to help people to understand that an education without the arts is as impoverished as an education without the sciences or the humanities.

Endnotes

1. Eloquent evidence: Arts at the core of learning. Report by the Presidents' Committee on the Arts and Humanities, 1995.
2. Testimony by the head of a cable television station and Wynton Marsalis to the United States House of Representatives Caucus, 1999.
3. Alan K. Simpson (1997). The 1997 Nancy Hanks Lecture. Washington, D.C.: Americans for the Arts, March 10.
4. Letter to the *New York Times* from Eugene Golden, Manager of the Music Guild, Los Angeles, August 12, 2000.
5. Rosenthal, R. (1994). Parametric measures of effect size. In Cooper, H. and Hedges, L. (eds.). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.
6. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
7. Winner, E., and Cooper, M. (2000). Mute Those Claims: No Evidence (Yet) for a Causal Link between Arts Study and Academic Achievement. *Journal of Aesthetic Education* 34 (3/4), 11–75.
8. Catterall, J., Chapleau, R., and Iwanaga, J. (1999). Involvement in the arts and human development: General involvement and intensive involvement in music and theater arts (pp. 1–18). In E. Fiske (ed.). *Champions of Change: The Impact of the Arts on Learning*, Washington, D.C.: The Arts Education Partnership and the President's Committee on the Arts and the Humanities.
9. Heath, S., with Roach, A. (1999). Imaginative actuality: Learning in the arts during the nonschool hours (pp. 20–34). In E. Fiske (ed.). *Champions of Change: The Impact of the Arts on Learning*. Op cit.
10. *College-Bound Seniors: A Profile of* SAT *and Achievement Test-Takers*. (1987–1992; 1994–1997). Princeton, NJ: The College Board.
11. Eisner, E. (2000). What justifies arts education: What research does *not* say. In McCarty, M. (ed.). *Enlightened Advocacy: Implications of Research for Arts Education Policy Practice. The 1999 Charles Fowler Colloquium on Innovation in Arts Education*. College Park: University of Maryland.
12. *College-Bound Seniors: A Profile of* SAT *and Achievement Test-Takers*. (1998). Princeton, NJ: The College Board.
13. Haanstra, F. (2000). Dutch studies of the effects of arts education on school success. *Studies in Art Education,* 19–33.
14. Harland, J., Kinder, K., Lord, P., Stott, A., Schagen, I., Haynes, J. (2000). *Arts Education in Secondary Schools: Effects and Effectiveness*. National Foundation for Educational Research. York, United Kingdom.
15. Moga, E., Burger, K., Hetland, L., and Winner, E. (2000). Does studying the arts engender creative thinking? Evidence for near but not far transfer. *Journal of Aesthetic Education* 34 (3/4), 91–104.
16. Getzels, J. and Csikszentmihalyi, M. (1976). *The Creative Vision: A Longitudinal Study of Problem-Finding in Art*. New York: Wiley
17. Graziano, A., Peterson, M., and Shaw, G. (1999). Enhanced learning of proportional math through music

training and spatial-temporal training. *Neurological Research* 21 (2), 139–152.

18. Vaughn, K. (2000). Music and mathematics: Modest support for the oft-claimed relationship. *Journal of Aesthetic Education* 34 (3/4), 149–166.
19. After we completed our work, we subsequently discovered a paper that made the same point: Jambro, T. (1978). The arts and reading: A review of the Learning to Read through the Arts Program. *Bulletin of the New York State Art Teachers Association* 28 (2), 4–9.
20. Burger, K., and Winner, E. (2000). Instruction in visual art: Can it help children learn to read? *Journal of Aesthetic Education* 34 (3/4), 277–293.
21. Butzlaff, R. (2000). Can music be used to teach reading? *Journal of Aesthetic Education* 34 (3/4), 167–178.
22. Rose, D. (1999). *The Impact of Whirlwind Basic Reading Through Dance Programs on First-Grade Students' Basic Reading Skills: Study* 2. Chicago.
23. Keinanen, M., Hetland, L., and Winner, E. (2000). Teaching cognitive skill through dance: Evidence for near but not far transfer. *Journal of Aesthetic Education* 34 (3/4), 295–306.
24. Keinanen et al. Teaching cognitive skill through dance.

COMMENTARY

Main Points in Response to "Mute those Claims: No Evidence (Yet) For a Causal Link Between the Arts and Academic Achievement"[1]

JAMES S. CATTERALL
Professor of Education, University of California, Los Angeles,
Graduate School of Education and Information Studies

I will keep my responses brief and aimed at two essential aspects of this chapter of the proceedings—the actual meta-analyses and the authors' assertions of meaning.

On the chapter's meta-analyses of arts-related studies and multi-arts programs: the Winner and Cooper meta-analyses of correlational and experimental arts and multi-arts programs across the United States, with one exception discussed below, are extensive and competent. Among the arts experiences included in this chapter are the arts and verbal outcomes and the arts and mathematics outcomes. The authors also report on the basic skills effects of multi-arts programs. This term refers mainly to whole-school and multi-school efforts to integrate multiple art-forms into academic teaching and learning— using drama in a history class for example, or music in a mathematics instructional unit.

In my own experience, Winner and Cooper were dogged in pursuing the technical details of our team's test score findings from a 1999 evaluation of Chicago Arts Partnerships in Education.[2] Winner called or e-mailed a dozen or more times to get things right. If Winner and Cooper pursued other studies with such diligence, they surely did good work in providing informed and correct reviews of the test-score effects of the arts.

The studies Winner and Cooper reviewed in separate meta-analyses traverse quite a range—arts and motivation, arts and math achievement outcomes, and arts and verbal outcomes, multi-arts programs. The meta-analyses generally point to small positive effects of the arts on standardized test scores—usually of language arts or mathematics skills. One conclusion of the chapter is that the arts unquestionably have been shown in research to be motivating to children.

REAP focuses on experimental studies because true experiments anchor claims of causal linkages more firmly than many other research designs. Winner and Cooper, to their credit, pay some homage to "correla-

tional" studies, a category into which they seem to relegate all non-experimental research. The studies they threw out in their culling generally show that arts' effects on most outcomes reported are considerably higher than those found in REAP.

Winner and Cooper generally dismiss the importance of non-experimental studies of the arts, but not completely. They assert that we have much evidence of the arts' effects on student motivation: this conclusion is derived from non-experimental studies examined but not included in the meta-analyses in the Winner and Cooper chapter.

This dismissal of most relevant research and the global "no-effect" claims made on the basis of so narrow a slice of our knowledge about the academic effects of the arts is curious and non-scientific. Research in general, and research in the arts as well, has long valued multiple ways of knowing. All of Project Zero, the research institute issuing the REAP review, is dedicated to multiple ways of knowing and expression. The Winner and Cooper chapter, and the REAP report as a whole, disenfranchise a great many research models and their expert practitioners—anthropologists, ethnographers, program evaluators, arts education researchers, and even teachers. These non-experimentalists enlist many ways of understanding the effects of arts programs. Proven methods such as observation, participant-observation, surveys of clients and providers, intensive interview research, triangulating research using multiple lenses, and examination of student work are only a few of the methods denied credence by REAP.

A substantial problem results for the Winner and Cooper chapter and for REAP as a whole. The issue is not that REAP's various authors did not perform competent meta-analyses. They did so. The primary flaw in Winner and Cooper, as well as in the whole of REAP, is that the lead author/editors claim that they have important all-encompassing things to say about the lack of academic effects of the arts—namely that we do not yet have evidence of any effects of the arts on academic success. In actuality, the REAP research reviews negate such a claim.

REAP itself presents plenty of evidence, in drama and language skills and in certain music experiences and vitally important spatial reasoning skills of humans. Beyond this, the Winner and Cooper chapter contradicts Ann Podlozny's excellent chapter on drama and verbal skills.

Witness Podlozny's REAP chapter title: "Strengthening Verbal Skills Through the Use of Classroom Drama: A Clear Link."[3] How, in the face of the largest body of evidence concerning verbal skills accumulated and analyzed by REAP, can Winner and Cooper come to a negative global conclusion about the effects of the arts on verbal skills?

Why would their meta-analysis of the "effects of the arts on verbal skills" not include the many studies found in REAP regarding drama and verbal skills that come to the opposite conclusion?

REAP author Ann Podlozny claims,[4] "Clearly drama is an effective tool for increasing achievement in story understanding, reading achievement, reading readiness, and writing." Just who should we listen to in this situation?

By excluding a huge body of evidence reaped by their own author Podlozny, Winner, and Cooper were quite selective when analyzing the effects of the arts on verbal skills in their chapter. They missed the most important evidence, all present in their files! The "no evidence" title of their chapter alone is nothing but global in its implication—and perhaps in its intent. This title, and the "negligible effect" global claims of chapter one and the REAP Executive Summary, are simply contradicted by key and strong remaining chapters.

One overriding trap REAP steps into is their search for "causality." This quest sounds reasonable, especially in the face of some exaggerated claims that the authors rightly identify and decry. This causation criterion drove the authors to find quantifiable interventions with quantifiable outcomes; this compelled a focus on standardized test scores because meta-analyses require similar quantifiable outcomes across studies to be reviewed. The REAP design was to find true experiments that are better at supporting claims of causality than other types of studies. This assertion we should grant. And this is a long-standing tradition of educational psychology research.[5]

But consider the following example, which raises questions about establishing "true causality" in scientific studies. "Causality" in the case of pharmaceutical products, according to the canons and standards of the United States Food and Drug Administration, demands double-blind experiments, and nothing short of this will do. These are studies where neither the program providers (doctors, theater teachers) nor the patients (diseased persons, characters in a United States history dramatization) know whether they are getting an experimental treatment or some placebo. Of course, double-blind experiments in the arts are practically if not absolutely impossible. The theater treatment group would know they were getting a dose of theater and so would their teachers.

A first-rate medical researcher would survey research in the arts, including the experiments held high in REAP, as flawed and not supportive of true causal claims. The upshot? Causality in research is based on the canons of individual fields of inquiry. Even the most expert of educational psychology research is not at the top of any respectability pecking order. But it does a better job of exploring causal relationships than many weaker and casual research designs.

The Primary Implication of this Discussion

Focus on experiments does not mean that other research methods do not capture causal relations correctly. Many expert researchers, along with teachers themselves, gain tremendously important insights through lenses far removed from experiments.

REAP simply cannot say we lack evidence of the effects of the arts on academic success, either in the Winner and Cooper chapter or in the whole of the report. Even REAP's own chapters on theater and verbal skills, as well as on certain music experiences and spatial reasoning ability, show large, significant effects on test scores and crucial academically related abilities. And it just so happens that all but twelve of the nearly two hundred experimental or quasi-experimental studies reported in REAP are in music and drama, where the effects are strongest.

When these well-established and sizable positive effects of REAP are juxtaposed with REAP's small positive effects reported in the visual arts and dance, what is a reasonable overall message of REAP? I would say the message is:

> *We found considerable effects of the arts on standardized test scores in verbal domains and in spatial reasoning ability (so far), and there is more available in other studies we excluded. And since we excluded by design all studies other than experiments with test scores as outcomes, we cannot, on the basis of reap, make any overriding claims about the effects of the arts on basic skill, not to mention more higher-order academic and academically related outcomes.*

Winner and Cooper seem quite in agreement on some of this spirit in their own chapter:

> *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. The arts are important human ways of understanding and knowing, no less important than the sciences.*[1]

The Real Message of REAP

Most arts educators and myself also could not agree more. Researchers have found good indicators of such outcomes, but not mainly through experiments. REAP, along with Winner and Cooper, simply cannot say that we have "no evidence (yet) for causal claims" about the effects of the arts on academic success.

This is the message proffered by Winner and Cooper, and also in the summary of the whole report. This message is wrong.

The correct message is this:

> *Our examination of experimental research across many art forms and basic academic skills growth generally reveals positive linkages; moreover, where we benefit from numerous studies, namely in drama and music, the effects are strongest. Where we found a paucity of studies, namely in the visual arts and dance, the effects are still positive but we clearly need more research.*

And finally, the disconnect. Winner and Cooper, along with the volume editors, are in a considerable rush to defend the arts on the basis of what the arts alone can do—i.e., on the basis of arts outcomes and not on academic outcomes. They seem to believe that by belittling research that shows academic benefits of the arts, their own rationale for supporting the arts will rise in its fortune. People will buy into the art-for-arts'-sake message if academic outcomes research can be shown false.

Well, the research in question is not false and amounts to much more than Winner and Cooper or the REAP co-editors care to admit. The leaders of REAP are sending a message unrelated to the results of their research reviews, and a message that thus must rise or fall on its own merits. It is a meritorious message, but one not linked to the ample research on academic outcomes as the authors would somehow have us believe.

REAP's wish for greater societal valuation of the arts is a mere idle wish until it attracts sponsors, leaders, a strategic plan, and soldiers. There is perhaps no better game plan than implementing more arts and arts-integration programs in the schools so that children will grow up with greater appreciation for and skills in the arts. Arts programs are proliferating across the nation partly through a spirit connected to the research analyzed by REAP. Educators, parents, and school boards have grown to believe, correctly, that the arts do more for human development than increasing arts skills—even if we do not yet know the full story.

The idle dream of greater societal appreciation for the arts is rhetoric without a game plan. Such urging will have little effect on society now, just as it has failed to demonstrate an impact on society for the last half century.

Endnotes

1. The Getty conferees were given page proofs of REAP chapters in advance of the Getty Conference discussing this Project Zero report. The authors also gave conferees copies of the final published work, which appeared in the *Journal of Aesthetic Education*. The citation is: Winner, E. and Cooper, M. (1999). Mute those claims: No evidence (yet) for a causal link between arts study and academic achievement. *Journal of Aesthetic Education* 34 (3/4), 11–76.
2. This CAPE evaluation is reported as a chapter in *Champions of Change: The Impact of the Arts on Learning*. Washington, D.C.: The President's Committee on the Arts and Humanities (sponsored and published also by the GE Fund, the MacArthur Foundation, and the Arts Education Partnership (Washington, D.C.).
3. Podlozny, A. (1999). Strengthening verbal skills through the use of classroom drama: A clear link. *Journal of Aesthetic Education* 34 (3/4), 239–276.
4. Winner and Cooper find small effects on "the arts" and verbal skills. The Podlozny chapter, based on many more studies, says convincingly the opposite.
5. Winner and Cooper have a very constrained view of their target. They largely focused on a very few quantifiable experiments with standardized test scores as outcomes—fewer than two hundred studies out of ten thousand originally culled. A reasonable guess is that considerably more than one thousand of the remaining studies make some valid contribution to understanding the academic effects of the arts.

COMMENTARY

Response to James Catterall

ELLEN WINNER AND LOIS HETLAND

James Catterall states his critique of REAP as follows:

> *The primary flaw in Winner and Cooper, as well as in the whole of* REAP*, is that the lead author/editors claim that they have important all-encompassing things to say about the lack of academic effects of the arts—namely that we do not yet have evidence of any effects of the arts on academic success.*

We respectfully differ with this representation of the Reviewing Education and the Arts Project (REAP) report and wish to address three questions raised by this statement: (1) Do we claim that our findings are all-encompassing? (2) Is there evidence that arts affect academic success? (3) Does REAP have important things to say?

Comprehensiveness of Findings

We claim comprehensiveness, not all-encompassing results. We reviewed all types of empirical literature exhaustively over three years. We classified it by design categories into experimental, correlational, and pre-experimental (designs with no control groups). We submitted to meta-analysis those studies that had sufficient data. We explained the strengths and limitations of the meta-analytic method. We reported studies that we identified but did not include so that the work was acknowledged and its conclusions were readily accessible to the field. We interpreted our results by standard procedures of social science research. Although it is the most systematic and comprehensive synthesis in the field of arts education research to date, we acknowledge that REAP includes only work that has sufficient data to be meta-analyzable.

We neither dismissed nor disenfranchised non-experimental research models. Winner and Cooper did not dismiss the research on motivation. Rather, they reported that they were unable to meta-analyze these studies because of insufficient reported data. Far from ignoring the research, Winner and Cooper listed each of these studies in a table so that a record of results would be available. However, Winner and Cooper could not and did not conclude that the arts unquestionably motivate children, because a number of these studies were correlational rather than experimental in design. Nor did

REAP dismiss non-experimental studies. Rather, in our introductory article,we characterized correlational research as not able to demonstrate causality, and experimental and quasi-experimental research as able to demonstrate causality. We did not demand the strict double-blind procedures used in medical research, because these procedures are rarely practicable in a school setting. Nonetheless, we stand by the distinction between what one can conclude from correlational versus experimental research, which is part of the canon accepted by social science researchers.[1] The "proven methods" mentioned in the critique (observation, surveys, interviews, examination of student work) are all excellent research tools that contribute substantially to understanding. But are they proven ways of finding out whether the arts boost academic achievement? It all depends on the research design. Only an experimental or quasi-experimental design allows an answer to this question.

Is There Evidence that Arts Affect Academic Success?

Catterall's critique centers on the report by Ellen Winner and Monica Cooper entitled "Mute Those Claims: No Evidence (Yet) for a Causal Link Between the Arts and Academic Achievement." Essentially, the criticism is that Winner and Cooper state (in their title and in their article) that the research thus far does not demonstrate a causal relationship between arts and academic achievement, and that this contradicts other articles in the REAP report in which causal links were reported, such as that by Ann Podlozny and the two by Lois Hetland. We see no contradiction. The title does not refer to all ten of the REAP meta-analytic studies, but only to the research meta-analyzed by Winner and Cooper. The titles of the Podlozny and Hetland analyses also reflect their (positive) conclusions. The title of the entire *Journal of Aesthetic Education* volume is *The Arts and Academic Achievement: What the Evidence Shows*. This title reflects the scientific tone of the enterprise, which was to find out what conclusions the evidence thus far allows.

Winner and Cooper's title speaks to a commonly heard generalization about unqualified transfer effects from combined forms of "arts" instruction, and it accurately summarizes the conclusion reached in their analysis. This analysis combined studies testing the claim that studying the arts (undifferentiated by art form, so study could be either of several separate arts disciplines, or arts integrated into the academic curriculum, or both) leads to verbal and mathematical improvement. The analysis showed no significant positive effects. Criteria for inclusion required more than one art form to be employed, so studies such as those in Podlozny's and Hetland's analyses were not included. The distinguishing feature of included studies is that in all cases students studied a combination of art forms. Thus, neither taking a variety of kinds of

arts courses nor exposure to an arts-integrated curriculum has been demonstrated empirically to boost verbal or mathematical skills.

In three cases, the evidence found by REAP was strongly positive, both in size and in its ability to generalize to other studies. Podlozny examined studies investigating the effect of a specific art form (drama) on verbal achievement when it was integrated into classroom instruction and found a strong positive effect. Hetland examined studies investigating the effect of another art form (music) on spatial reasoning and also found positive effects —temporary (ten to fifteen minutes) enhancement from music listening for adults, and longer-term enhancement for children given active music instruction for up to two years. Of these three strong, positive, and generalizable findings, only Podlozny's speaks directly to school success. Hetland's "Mozart effect" analysis is not about learning, but about how the mind works. Hetland's analysis of music instruction studies showed effects on spatial tests whose results will not necessarily increase school success. Although they are an essential dimension of human intelligence, many schools do not routinely utilize or assess spatial abilities.

Does REAP Have Important Things to Say?

We believe REAP has important things to say. Throughout the project and the report, we strove to be accurate and judicious in all our analyses and interpretations. We set out to discover what the research evidence *thus far* allows us to conclude. We were careful to separate our opinions from the results of the research. The evidence from the type of studies that allow generalization to broader populations does not support the claims we generally hear. Thus, the case is still to be made. We call for more—and more rigorous—research on transfer, starting with qualitative studies that result in testable hypotheses. We also call for refocusing on learning in the individual art forms. We stand by our conclusions.

We also wish to affirm our opinion that professional disagreement among scholars representing differing viewpoints is constructive. It is through conversation about such disagreements that understanding advances.

Endnote

1. Cook, T. D. and Campbell, D. T. (1979). *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton-Mifflin.

COMMENTARY

Comments on the Question of Transfer

ELLIOT EISNER
Stanford University

At the most fundamental level the use of anything learned in one context and applied to another is an instance of transfer. However, we can distinguish between proximal transfer and distal transfer, specific transfer and general transfer, in-domain transfer and out-of-domain transfer. It is one thing to apply what one has learned to another situation and quite another to use the information as a means for inventing a way to make the information useful in a different domain.

For example, learning how to experience the changing relationships in music could enable someone to recognize the changing relationships on a soccer field or basketball court. This ability is an example of distal transfer rather than proximal transfer, out-of-domain transfer rather than in-domain transfer, general transfer rather than specific transfer. While no crisp line can be drawn between the distinctions I am describing, the differences between them are important enough to draw the distinctions.

To teach an art form, say drama or visual art, and to find that students enrolled in drama courses improve their reading ability and that students who learn to see visual patterns in visual arts are also able to see them on the walls of a building, is to say that the skills learned in drama and in the visual arts are essentially the same ones they employed in these other domains. Transfer here is specific rather than general, proximal rather than distal, in-domain rather than out-of domain.

What I believe we are seeking is the transfer of learning in domains that are dissimilar to the one in which the initial learning took place. That is, we are trying to determine whether learning in music increases performance in mathematics, whether learning in the visual arts predicts performance in science, whether learning in physics predicts ability to think scientifically about social affairs.

If an arts course integrates history with the arts it should come as no surprise if performance in history was promoted as a result; after all, history, its content and its methods, was being taught at the time the arts were being taught. This may very well be a fine way to teach both the arts and history, but we have to be careful about the claims we make about transfer. Put another way, the kind of transfer in which we are interested is one that enables students

to use what they learn in one field in another without being taught how to do so. To the extent to which explicit teaching or practice of skills or ideas occurs and to the extent to which those skills and ideas need little or no modification by the new context, the claim for transfer is difficult to make.

Some Comments on the Report

The significance of differences in a relationships among variables are of two kinds: statistical and, in our professional context, educational. It is quite possible to have statistically significant outcomes that are educationally trivial. Thus, assessing the significance of an outcome requires the application of educational as well as statistical criteria. The former depends upon judgment, the latter on algorithm.

To have educational significance academic performance needs to be conceptualized as outcomes that matter educationally. To determine what matters educationally one could use, for example, the various performance standards for subject matters constituting the school curriculum, standards that have been promulgated by professionals working in those fields. Better still, one could generate with teachers a complex of outcomes that represented significant educational achievements and then try to assess their realization as a function of experience in the arts. No material in the report provides such evidence.

There is typically little or no interpretation of the significance of differences from an educational perspective between experimental and control groups exposed and not exposed to arts courses. For example, the magnitude of the increase reported in SAT scores between those enrolled in arts courses and those who are not is educationally trivial. The mean verbal SATs for those who take three years of arts compared to those who do not is the difference between a mean score of 432 and a mean score of 413. The difference is nineteen points. A difference in nineteen points on the SAT is a function of three multiple-choice items. Three items, I submit, is not an educationally significant difference, even when the populations are huge and the score differences are statistically significant.

There is very, very little replication of the studies reported. The general picture that I received from this well-documented and carefully executed meta-analysis of studies is that there is a desperate search for significant differences on forms of performance that are educationally marginal. The differences sought are derived from research studies that do not describe the educational programs related to the outcomes they measure and pay no attention whatsoever to learning in the arts, which, from a rational perspective, is what would account for higher levels of academic achievement.

To get serious about outcomes that matter in education we would need to look far beyond standardized achievement test scores. We would need to determine the extent to which students actually use what they learn in school in out-of-school settings. It is in these out-of-school settings where they can do what they really want to do.

There is a huge difference between determining what a student can do and knowing what a student will do. It's not the greatest of educational victories to enable a student to read who does not choose to do so.

The conception of educational assessment I have described is currently far beyond our customary practices and I do not suggest that it necessarily needs to be pursued in this context. I use it simply to illustrate the distance that eventually we will need to travel in order to understand what we need to do to make school programs educationally consequential.

Legitimating the educational significance of standardized test scores, in my opinion, exacerbates one of the largest problems we have in education. It trivializes outcomes that matter, it distracts us from attending to issues and practices that do matter, it neglects dispositional outcomes, and it promotes, at its most egregious, cheating and teaching to the test. It provides a false sense of security to those who want an unambiguous index of educational quality. It coerces teachers to stick to the curriculum guide no matter what, which, as everyone knows, is the surest road to hell in a classroom.

Ironically, academics, whose understanding of education is theoretically deeper and wider than that of the lay citizen, often exacerbate educational problems by legitimating assessment practices that often have mis-educational effects.

My overall appraisal of the research pertaining to the relationship between arts experiences and academic achievement is that the claims made about the strength of these relationships have exceeded their evidentiary justification. In addition, most of the studies are theoretically barren. In an article entitled "Does Experience in the Arts Boost Academic Achievement?" published in the journal *Art Education* in January 1998, I had this to say after having reviewed a number of studies done in this domain:

> *What I have not mentioned is that in order to understand why arts courses have such effects, if in fact they do, requires a theory that relates the skills developed through the arts to the demands made upon students in academic classes. Perhaps it is not skills at all that arts courses develop, perhaps it's the promotion of certain attitudes, attitudes that promote risk-taking and hard work. Perhaps the effects—if effects there are—of arts courses on academic achievement is due*

to the motivational effects of arts courses; perhaps students in arts courses enjoy school more and therefore attend more regularly. If higher motivation is the source of academic achievement it might be that other motivating experiences might have the same effects. What is needed is more than correlations or statistically significant differences between groups, what is needed is a theory that links experience in the arts with academic achievement. What we need is a theory that explains the connections between the cognitive skills work in the arts develop and the function these skills perform in academic work of other kinds. To create these links the arts programs in which students are enrolled need to be carefully described, as do their academic programs."

I am pleased that the Winner-Hetland meta-analysis project substantiates on virtually every count the conclusions that I drew in that 1998 publication.

Where Do We Go from Here?

The critical question now is how we design studies that will enable us to answer that question. Using conventional or traditional assumptions about research design, how might we proceed? Let me suggest the following.

First, we would need to determine the kinds of academic outcomes that we cared about; second, we would need to decide what would count as evidence; and third, we would need to determine how that evidence might be secured, measured, and evaluated. Since this is a thought experiment let's simply assume that we have a reliable and valid test that measures important educational outcomes in the social studies. Let's assume further that these outcomes are those the faculty of the school in which the study is going to be done believes to be not only important, but consistent with what they are teaching and trying to accomplish. Let us assume further that the faculty has examined the social studies tests we propose to use and regards them as fair and appropriate.

In setting up the experimental design one group of ten-year-olds would receive the art curriculum, the other would not. The same design conditions would apply to the groups of twelve-year-olds.

Once having made the selection and assignments, we would want to make sure that the competencies of the teachers were about the same for each of the four classes, for if some teachers were far better than others it could bias the results in favor of the better teachers. Assuming about equally able teachers, we might then want to design or acquire an art curriculum that from a theoretical perspective was most likely to develop the cognitive

skills that gave students in that art program an advantage in learning the social studies. To do this we need a theory of cognition that enables us to identify the kind of cognitive demands activities in the art curriculum made and the relationship of those demands to academic performance in the social studies. If we could design curriculum activities that fostered such skills, the probability of finding positive relationships between performance in the arts and the social studies would be increased.

Once having designed or acquired such a curriculum, equalized the teacher's teaching abilities, and randomly selected and assigned the students to experimental and control groups, we could proceed. Of course we would need to determine the length of the experimental treatment, and we would need to monitor that treatment to make sure that the program as intended was actually provided; many experiments fail because the treatment investigators thought was provided wasn't. To avoid this problem we would need to make sure the art curriculum was one teachers could teach. If no, we have an in-service problem to deal with as well.

In addition, to avoid the Hawthorne effect it would be good to give a placebo to the control group in order to equalize novelty as a motivating factor in the experimental groups performance.

How long should the experiment last? Many experiments expect too much in too short a period or use a short treatment time in order to maximize control over potentially confounding variables. Until rather recently too many experiments in education were like commando raids. Let's say that we will conduct this experiment for at least five months, preferably ten. We want to provide enough time for the art program to have effect. So we now have a five-to-ten month experiment in which an art program designed, in part, to develop artistically relevant cognitive skills is used to determine if the skills developed in the art program increase academic performance in the social studies.

But is that all? Don't we want to know something about what the students have learned in art? Performance levels in art are proxies for cognitive development in that field. We need to know if those skills had been developed, and the students' art work is a good source. In addition, what if the students did very well in social studies, but paid for it by poor performance in art? Is the trade-off worth it?

We are working with the tacit theoretical assumption that the kind of thinking students do in art is related positively to high-level performance in social studies. How shall we know about the performance level of students in art? One thing we can do is to look for it in student work and listen for it in student discourse. We can read it in what students write. What counts in art

education can include not only the art students can create, but also what they are able to say and write about art. We need assessments in these areas too.

I won't, given the time limit, try to describe how these art assessments might be undertaken; I only want to point out that they ought to be done. There are other research designs that can be used, but this description, I hope, provides some sense of one kind of experimental effort that is needed

To sum up in highly distilled forms the features of the research undertaken to assess transfer from the arts to other domains, I list them here as succinctly as I can: (1) The studies do not describe the programs in the arts that were used to teach the students enrolled; (2) the studies do not describe or measure the effects of the arts programs on students' artistic learning; (3) the studies use a highly restricted array of outcome measures to assess transfer; (4) the "effects" secured are generally small and, when there are effects, they are often educationally trivial; (5) the "effects" are mainly the result of correlational studies from which causality cannot be claimed; (6) the subject areas most likely to show effects are music to spatial ability and drama to reading; (7) there are only a few studies that have tried to determine the durability of effects when effects were identified.

Each one of the foregoing observations suggests the kind of research that need to be designed in order to secure more confident answers and more robust conclusions to questions concerning the effects of arts experience on academic achievement.

Another (Somewhat) Different Approach to Assessing the Effects of Transfer

There is another way to think about transfer: it is this. The situations within which students live and work are replete with a wide variety of tasks and opportunities. They are situations pervaded by norms and models, they impose constraints, and they invite forms of thinking that are often unpredictable. In a sense, these situations from a developmental and cognitive perspective are resource rich.

To understand what students learn and how their cognitive processes are stimulated, developed, and refined, we need to be in a position to analyze the varieties of interaction that characterize those situations. What does an art room make available to the student? What opportunities and forms of thought are engendered by being a member of a string quartet? In what ways does a dance class promote thinking, and what kind of thinking does a dancer need to do in order to get better at what he or she does?

Of course, what complicates the analysis of such situations is that unlike inert material, the human being is a construing organism. Perception is

itself selective and the meanings that individuals make of the situation depend not exclusively, but significantly, on what they bring to it. Thus the analytic task is to understand how students make sense of the situations they inhabit. This task might be considered a kind of cognitive ethnography, but whatever it is called, it reflects a realization that the context or situation itself is not unidimensional and that the student learns more than one thing at a time.

It is out of an analysis of the richness of the context and the meanings made by individuals in it that we might be able to identify, despite individual differences and personal histories, common elements or themes that the various arts exemplify and the modes of thought they promote.

But even given the way in which I have characterized the conditions of learning, how do we account for transfer? Again, by transfer I want to emphasize distal transfer rather than proximal transfer, general transfer rather than specific transfer, out-of-domain transfer rather than in-domain transfer. In other words, the challenging task is to determine how individuals are able to make the connection between ideas and skills learned in one domain to another quite different domain.

The best I can come up with is related to what Arthur Koestler called bissociation. Bissociation is the ability to relate two previously separated independent conceptual fields so that they have a productive payoff. He regarded bissociation as the cognitive core of creative thinking, a core rooted in our capacity to create humor. The ability to see connections where they are not obvious is a form of cognitive invention, a kind of creative thinking. This ability may be related to a *g* factor in the uses of intelligences. I have no doubt that there is a *g* factor traversing whatever collection of defined intelligences one can identify. Making connections, that is, using what one has learned in one setting in another quite unlike it is a mark of both perceptivity and an indication that transfer has or is occurring.

Yet, what is troublesome about the explanation I have just provided is that it comes very close to being merely a redescription of the very process we are trying to explain. Yet it does broaden considerably our purview of the conditions that would account for the individual's cognitive development in an educational program, and it does suggest that the dependent variables influenced by those conditions are far wider than the variables measured by standardized tests. It implies, at the very least, that while ideas are being developed by the student in the situation, so too are norms and so too is affect. More than one thing is going on, as any good primary school teacher will acknowledge.

Thus, one potentially productive arena for research is the close analysis of the conditions that characterize the situations within which students learn, including their pervasive quality and the connections students make

between the forms of learning engendered and other theoretically relevant forms of thought and action outside of that domain. Such work cannot be completed within the confines of a research project. It requires a research program. If such a program provides useful methods and concepts for studying the conditions of learning and their effects on the arts, I believe it will have significant implications for learning in other fields.

Studies Needed

We need highly detailed examinations of students engaged in the process of creating art forms. We need to be able to make inferences about the forms of thinking that various problems and materials impose upon them and how it is they resolve those problems and cope with those materials.

We need studies of teaching practices in order to understand how pedagogical intervention with respect to prompts by the teacher, modeling, and problem constraints imposed by the teacher stimulate, challenge, and develop students' modes of thinking.

We need studies that examine the classroom as a community of practitioners functioning collaboratively within a set of social norms. We need to situate students in a social context in addition to understanding individual psychodynamics.

We need studies of the classroom milieu as a whole, studies that integrate all of the foregoing dimensions I described. Such work could give us a comprehensive picture of the impact of the environment as a whole on the thinking processes of students in the context of an art form.

In all of the foregoing we need to take into account the quality of the work students do. We also should be doing cross-sectional and longitudinal studies in order to determine changing patterns in performance.

We need studies that would enable us to determine whether deep experience in the arts has significant consequences of a dispositional type for students; are they more apt to pursue aesthetic forms of experience outside of the classroom, are they more apt to think metaphorically about the expressive character of situations they encounter, do they have a greater willingness to address what William James called "the vague" and what is sometimes called the ambiguous? Do they have less of a need to be literal, to have boundaries well defined, to know precisely what the destination of an inquiry is to be?

Finally, the most telling manifestation of educational consequences in any field emerges out of the school setting rather than in it. The primary aim of schooling is not to do well in school, but to do well in life. We ought to be looking to life outside of school to find out.

COMMENTARY

What Research in the United Kingdom Shows About Transfer from the Arts

JOHN HARLAND
National Foundation for Educational Research, United Kingdom

I would like to begin by adding my voice to the chorus of admiration for the meta-analysis work undertaken by the Reviewing Education and the Arts Project (REAP) team. It's a great achievement: it's highly systematic, thorough, and rigorous; it's insightful, yet well-balanced and measured in its conclusions; it's extremely useful in stimulating creative discourse between researchers; and it will be appreciated and valued in Europe, as well as here. From discussions with colleagues who know the literature better than I do, it is clear that the central drift of the REAP report is in broad agreement with our own (less comprehensive) reading of the existing literature.

I have only two critical observations to make on the report. I'll leave one to the end of my commentary and offer the other now. The first, however, is more of a criticism of the approach to meta-analysis and literature review the REAP team have bought into, rather than a comment on the work itself.

The adopted approach is exclusively concerned with quantitative studies, more precisely, quantitative studies of a particular type: what Hamilton and Parlett, many years ago, described, if my memory serves me right, as the "agricultural-botantist" model of educational research. While completely accepting that studies in this mold have a lot to offer investigations in the field of arts education—although, as Ellen Winner and her colleagues suggest, far too many are based on spurious forms of testing and many lack ecological validity—to adopt a review methodology that excludes other models of research is a serious flaw.

Thus, for example, the authors go to great lengths to avoid the so-called "publications bias," yet overlook a much more fundamental methodological bias that is inherent in the meta-analysis model they have adopted. The assumption is made that if studies do not possess a correlation coefficient, they do not warrant a place in an overview of the research in the given field. I cannot accept that assumption because it constitutes an incomplete social science and leads to a very partial and myopic view of what it is we are trying to understand. As a number of participants have suggested in the discussions, human beings differ from plants and physical matter in that they

create meanings. Consequently, any holistic research strategy must incorporate methodologies that acknowledge and allow for that fact.

I found it very depressing that throughout yesterday's proceedings, although we heard a succession of very stimulating papers, there was hardly any mention of a single study from an interpretive, qualitative, phenomenological or ethnographic paradigm. A particularly important casualty of this omission was the absence of a perspective from the learner. Their constructions of the meanings surrounding arts education—the importance, or lack of it, it has for them, what they do or don't get out of it—are crucial elements in our efforts to build up our knowledge of the effects and processes of arts education. Interpretive, phenomenological, and ethnographic methodologies are needed to access these meanings.

I do not wish to argue that such qualitative research, by itself, is a sufficient condition for research in this field, but I would submit that it is a necessary one, that the approach to meta-analysis adopted by the REAP team fails to embrace qualitative research and that future reviews of the evidence should redress this imbalance. Indeed, the challenge for the developers of the meta-analysis model is how to improve their techniques in order that qualitative studies can be fully incorporated. The jigsaw will not be complete without them.

By way of offering further comment on the REAP report, I would like to outline some results from a study recently completed by my colleagues and me in England. Lacking sufficient time to summarize the whole report, I will focus on the findings most pertinent to the concerns of the REAP meta-analysis, particularly those transfers discussed in the papers this morning.[1]

The results of the study will be published in October under the title *Arts Education in Secondary Schools: Effects and Effectiveness*. Launched by the Royal Society for the Encouragement of Arts, Manufacturers and Commerce (RSA) in 1997, the research was conducted by the National Foundation of Education Research (NFER). The main sponsors of the research were the Arts Council of England and the Local Government Association. An interim report, cited in the REAP review, was published in 1998.

The project was designed to address four aims: to document and evidence the range of effects and outcomes attributable to school-based arts education; to examine the relationship between these effects and the key factors and processes associated with arts provision in schools; to illuminate good practice in schools' provision of high-quality educational experiences in the arts; and to study the extent to which high levels of institutional involvement in the arts correlate with the qualities known to be associated with successful school improvement and school effectiveness.

To pursue these aims, the research drew on evidence collected through: case studies of five secondary schools with good reputations in the arts—these included annual interviews with two cohorts of pupils (approximately seventy-nine in total each year) who were performing well in the arts, interviews with school managers and arts teachers, and video observations of arts lessons; analyses of wider-ranging information compiled through NFER's QUASE project—data on a total of 27,607 pupils from 152 schools in three cohorts of year eleven pupils taking GCSEs between 1994 and 1996 were analyzed; questionnaires completed by 2,269 year eleven pupils in twenty-two schools, with related information on their GCSE results, prior attainment scores, and key stage three national test results, along with responses to a school questionnaire; and interviews with a cross-section of twenty employers and some of their employees.

Accordingly, the study was mainly based on qualitative self-reports from pupils. These established ten main categories of effect: seven as direct outcomes for pupils and three other types of effect. The seven direct outcome types and their frequencies are: (1) an increase in the knowledge and skills associated with particular art forms (1,177); (2) advances in personal and social development (736); (3) a heightened sense of enjoyment, excitement, and fulfillment, and therapeutic release of tensions (371); (4) effects that transfer to other contexts, such as learning in other subjects, the world of work and cultural activities outside of and beyond school (363); (5) the enrichment of communication and expressive skills (356); (6) the development of creativity and thinking skills (199); and (7) enhanced knowledge of social and cultural issues (98).

The study also contains two correlation analyses, one with a large sample and data on a limited range of variables, another with a smaller sample but with data on a wider range of variables. It is these analyses that are especially germane to the type of studies reviewed by the REAP report. Taken together, the two analyses revealed that there was no sound evidence to support the claim that the arts boost general academic performance in school-leaving examinations (GCSE) at the age of sixteen. For example, when social class and prior attainment were controlled for, taking two-year art, drama, or music courses from fourteen to sixteen was not positively associated with examination performances in English and mathematics. Where such variables are not controlled for, it is maintained that any correlations disclose more about the characteristics of the pupils taking different arts-related courses than any purported impact on general academic performance. Hence, essentially, the findings are in close accord with the general drift of Winner and Hetland's conclusions.

I would now like to consider our evidence for each of the transfers raised this morning.

Verbal Skills Through Drama

Once social class and prior attainment in English were controlled for, there was no significant positive association between taking courses in drama or participation in extracurricular drama and GCSE performance in English. However, in the qualitative testimonies, pupils volunteered accounts of gains in interactive communication skills, language development and expressive skills:

> *Mainly how to silence speech. You know, if I like shut up now, you'd be thinking "What should I say next?" (year nine, drama).*

Drama helped a lot with work-experience interviews—speaking to people in a friendly manner, not being too frightened of questions. . . . (year ten, drama).

Music to Maths

Once social class and prior attainment in maths were controlled for, there was no significant positive association between taking courses in music or participation in extracurricular music and GCSE performance in maths. Art (39) and drama (40) attracted more references to transfers to learning in other subjects than music (13). Only one of the latter cited a transfer to maths:

> *. . . music and maths. It might not seem like it, but you have to time bars and things and add them up and that process helps you in other subjects (year seven).*

Music to Reading

Once social class and prior attainment in English were controlled for, there was no significant positive association between taking courses in music or participation in extracurricular music and GCSE performance in English. None of the pupils alluded to possible transfers from music to reading.

Visual Arts to Reading

Having controlled for social class and prior attainment in English, there was a significant negative association between taking courses in art and GCSE performance in English. The pupils did not mention transfers from art to reading, but some described transfers from art to writing (e.g. observation, imagination, and presentational skills).

Dance and Reading/Non-Verbal Reasoning

Due to the small numbers taking dance courses at the fourteen-six stage, it was not possible to explore associations between this subject and general academic performance. Similarly, none of the pupils mentioned such transfers.

Arts to Creative Thinking

No statistical associations could be investigated, because of the lack of outcome measures for creativity thinking. However, the qualitative testimonies from pupils offered insights into outcomes relating to different aspects of creative thinking and displayed some interesting variations by art form.

Few pupils, however, alluded to the thinking skills and problem-solving dimension of creative thinking. One exception was:

> *Good for your mind as well—playing the piano, have to train the brain to be able to think . . . [gives example of different rhythms in each hand] . . . split your mind so you're not quite thinking of either but both of them at the same time (year ten).*

For two pupils, the ability to think more clearly was facilitated by listening to music. For one, in year eleven, this was about being more able to focus the mind, in particular through listening to background music while doing homework. (He mentioned, however, that it would be "a bit of a bummer in the exam," though, because there would be "total silence, no music.")

Although such references to thinking skills and problem solving were fairly infrequent, they were more likely to be associated with drama, music, and the arts in general. Art and dance each attracted only one such comment.

Compared to thinking skills and problem solving, there were more mentions of creativity per se, creativity as experimentation and creativity as imagination. One example of each will have to suffice.

> Creativity per se:
> *I think with the composition it encourages my creativity, I can, knowing how to write a piece of music, being able to, what I hear in my head, being able to put it down on paper, that's something really good and interesting to do, and that's been built up over this past year (year eleven, music).*

> Creativity as Experimentation:
> *Any form of making things up and experimenting with it, usually through sound and sight. But I think it is anything to tantalize your*

senses really, through what you can see and what someone else has given to you (year seven).

Creativity as Imagination:
We just sort of . . . on the spot you just think of anything and just make it up as you go along. . . . And so it sort of makes your brain try to imagine things. It's not like you have to act out a scene in a park or something. You just think of anything and just do it from there (year seven, drama).

Overall, different art forms seemed to be generating different levels and types of contribution to the development of creativity. Art, for example, attracted low scores for thinking and problem-solving skills, but high scores for creativity, experimentation and imagination. Drama also registered a high number of nominations for imagination, but had proportionately fewer for creativity, although more for thinking and problem-solving skills. Dance was similarly high on imagination, but low on thinking and problem-solving skills. Although attracting some references to thinking and problem-solving skills, music was comparatively low for creativity, experimentation, and imagination. In addition to such variations between art forms, the data suggested that the type and level of creativity outcomes varied according to differences in teacher pedagogies and course content.

To conclude, this focus on creativity brings me to the second of my critical observations. When reading the REAP report, I was struck and somewhat surprised to find that the authors treated the development of creativity as a transfer effect, even if a "near" transfer effect. Most teachers of the arts in England would consider the nurturing of creativity as a legitimate and direct objective of teaching within their own subject area. The arts would probably be seen as one of the most, if not the most, important carriers of creativity. Accordingly, from this perspective, to construct creativity as a transfer effect is to give too much away to other areas of the curriculum. At the very least, it raises questions about what we mean by "transfer" effects.

This, however, is a relatively minor point, when seen against the backdrop of a great deal of common ground between the conclusions reached by the REAP team and the findings from the study outlined above.

Endnote

1. For the full report of this study, see Harland, J. et al. (2000). *Arts Education in Secondary Schools: Effects and Effectiveness*. United Kingdom. The Library, National Foundation for Educational Research.

III.

The Relationship Between Music and Spatial Reasoning: A Summary of Two Meta-Analytic Studies

LOIS HETLAND
Project Zero, Harvard University

This article summarizes my two meta-analytic syntheses of empirical research that explore the relationship between music and spatial reasoning. The first analysis reviewed a group of laboratory experiments with adult subjects who listened briefly to music before taking tests of spatial reasoning.[1] The second analysis reviewed the enhancing effects on spatial reasoning for children ages three through twelve from programs of active instruction in music.[2]

There are both theoretical and empirical reasons to suspect a relationship between music and spatial reasoning. Karma[3] and, more recently, Leng and Shaw[4] have put forth theoretical reasons for such a connection. Shaw's "trion" theory motivated much of the empirical research reviewed here. The theory suggests that columns of neurons in the cortex that process "complexly structured music, regardless of style or period"[5] have three levels of activation (hence the term "trion"), and "prime" neurons used for spatial tasks.[6] The "trion" theory could explain any effect on spatial reasoning resulting from music instruction. Neural networks that process musical and spatial information may be either proximal or overlapping such that development of one influences the other.

However, a near transfer theory could explain an educational effect equally well.[7] That is, learning music may require synchronized use of both musical and spatial processing areas of the brain, even if those areas process information entirely separately. Thus, learning music might result in development of both processing areas.

In 1993, an empirical study testing the "trion" theory was published in the prestigious journal *Nature*.[8] In this laboratory experiment, adults (college

students) listened briefly (for about ten minutes) to three stimuli: (1) the first movement, "Allegro con spirito," of Mozart's *Sonata for Two Pianos in D Major*, K. 448; (2) silence; and (3) a tape of verbal instructions designed to induce relaxation and lower blood pressure. After listening, participants took spatial subtests of the Stanford-Binet: Fourth Edition. The music condition temporarily enhanced performance on one subtest (Paper Folding and Cutting) for about fifteen minutes, but not on two other subtests, Matrices and Pattern Analysis. Shaw and Rauscher used the term "spatial-temporal" to define the specific category of spatial tasks that music enhanced. Their definition specifies that such tasks require mental rotation (flipping and turning objects mentally) and spatial visualization (manipulating figural information through multiple steps) in the absence of a physical model.

The effect of music listening on the spatial-temporal task of Paper Folding and Cutting was not trivial, and to give readers an intuitive grasp of its size, the authors compared it to an IQ equivalent of 8–9 points. These results suggest that areas of the mind/brain that process musical and spatial information are related more closely than many cognitive psychologists would predict. For example, Gardner's theory of multiple intelligences posits that musical and spatial information are processed independently.[9]

This experiment attracted a great deal of attention from cognitive scientists. Researchers attempted to replicate the experiment with mixed results, some affirming and some negating the effect, and experiments addressing the question increased to a high of fifteen replications appearing in 1999. I reviewed such replications in my first analysis, "Listening to Music Enhances Spatial-Temporal Reasoning: Evidence for the 'Mozart Effect.'" As a result of these replications, other theories have been proposed as explanations for the effect. These include arousal (music may optimally arouse attention to the tasks so that performance improves[10] and its related hypotheses about preference (if subjects like a condition better, it may arouse them more[11]) and mood (some conditions may arouse more because of the feelings they evoke[12]). Parsons and colleagues have offered another "priming" hypothesis, suggesting that rhythm alone may be the element of music that influences spatial tasks.[13]

Scientific interest in this initial experiment, along with the effects of intense publicity, marketing opportunities, and arts advocates' interests in potential justifications for beleaguered and undervalued music programs, resulted in a great deal of misunderstanding about the implications of this original finding. For example, some states began to hand out free music tapes to all newborns in the hope that this would lead eventually to higher SAT scores. Mozart CDs for babies were widely marketed under the slogan that

listening to Mozart would make babies grow up smarter. But none of these claims flowed from the research. The effect was with adults, not children or infants; it was limited to a specific form of spatial reasoning; and it was temporary. The "Mozart effect" studies may ultimately tell us important things about the mind, but the road from a laboratory experiment to practical educational effects is long and rocky.

Although neither the original laboratory study[8] nor my analysis of the body of laboratory studies has direct implications for education, educational implications are addressed by the body of work analyzed in my second analysis, "Learning to Make Music Enhances Spatial Reasoning." These studies, conducted in schools and other instructional contexts, are of interest to the field of education because they address the question of transfer—that is, whether learning in one area can increase learning in another. Because this body of work has direct implications for education, and because it is often not clearly distinguished from the "Mozart effect" experiments, I summarized it in a second analysis with three subanalyses, defined by the type of spatial task employed as a measure of music's effect.

A study by Rauscher, Shaw, Levine, Wright, Dennis, and Newcomb[14] is typical of this body of work. Three- and four-year-old children were given music instruction (in this case, piano and singing), and children in control groups, similar in age and background, were taught basic computer games and skills or were given no special program for one school year. The children were tested on spatial measures shortly before and after the instructional programs, and those who received music instruction were found to have greater skill in completing spatial-temporal tasks. Such a result suggests that music instruction may have the bonus effect of enhancing skill in a specific type of spatial reasoning.

This effect is of particular interest to arts educators and advocates who wish to justify music programs that are in danger of being cut from school curricula. Spatial reasoning is valued in many disciplines, ranging from physics and engineering to surgery, sculpture, navigation, archaeology, and air traffic control. Spatial abilities could be used to understand many school subjects, including mathematics, if teachers taught those subjects in a way that drew on their students' spatial strengths. Thus, any influence that learning music might have on increasing skill in spatial reasoning could potentially be valuable in education, and an enhancing effect of music instruction on such skills could serve as a lobbying tool for keeping music programs in schools.

Unfortunately, promising individual studies are not enough to inform policy responsibly. For research to inform policy, one must synthesize a body of research, because no single study, however well designed, can demonstrate

a causal relationship. In addition, instrumental justifications of music programs are not without cost, since programs tend to emphasize the learning goals they are intended to produce. If music can be shown to have an enhancing effect on spatial reasoning, music educators might be distracted from musical purposes and shape music programs to develop spatial rather than musical abilities. Thus, despite well-intentioned efforts on the part of advocates to keep music in schools through demonstrating the enhancing effects of music instruction on spatial reasoning, such an effect could ultimately weaken music instruction.[15]

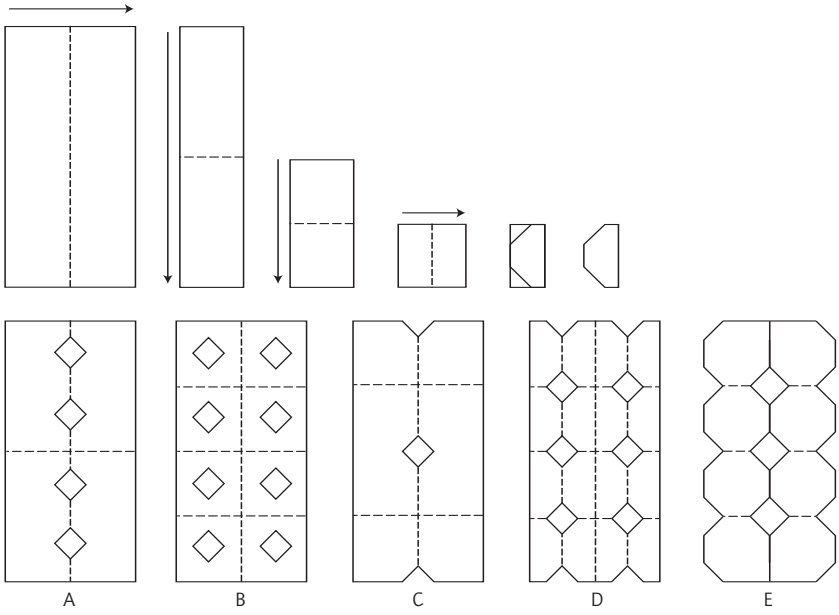
Does Listening to Music Temporarily Improve Spatial Reasoning?

An exhaustive search identified thirty-six experiments (2,469 subjects) that could be synthesized through meta-analysis. Those experiments included were conducted in laboratory settings with adults who listened briefly to a musical stimulus that was predicted to enhance spatial reasoning and/or listened to at least one non-musical control condition. Researchers also employed some measure of spatial reasoning and made enough data available to compute an effect size (i.e., the degree of relationship between the musical condition and the score on the outcome measure).

None of the replications exactly reproduced the original experiment. Many used the same music as the 1993 experiment, but some researchers predicted that enhanced spatial reasoning would result from other movements or pieces by Mozart, from other classical music (e.g., Schubert, Mendelssohn), from a piece by the contemporary composer, Yanni, and from musical stimuli comprised only of pure rhythm or pure melody.

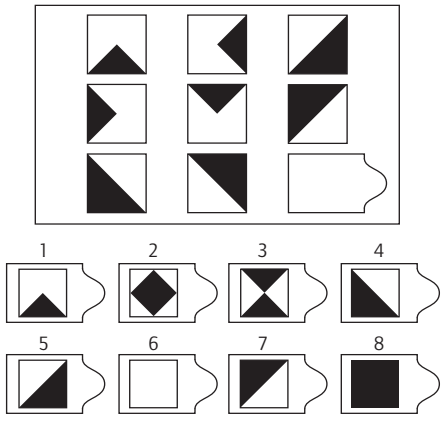
Replications also varied by the measures used to index spatial reasoning. "Spatial reasoning" is a term that encompasses a range of intellectual processes, much as the term "heart attack" refers to variety of medical traumas. The Paper Folding and Cutting subtest of the Stanford-Binet: Fourth Edition, a task used in the original experiment and many replications, is a good example of the type of task that Rauscher and Shaw call "spatial-temporal" and that they predicted certain types of music would enhance.[16] A sample item is shown in Figure 1. This task requires subjects to imagine folding and cutting paper in ways similar to actually folding and cutting paper snowflakes. Researchers attempting to replicate the original experiment used a variety of other tests as well, some of which did not meet the criteria Shaw and Rauscher set for spatial-temporal tasks. For example, matrices tasks do not qualify as spatial-temporal. Figure 2 shows a sample item from Raven's Progressive Matrices.[17, 18] In these tasks, one figure is missing from a gridded pattern, usually three by three, of figures that are arranged vertically and

Figure 1: An Item from the Paper Folding and Cutting Subtest of the Stanford-Binet: Fourth Edition.



Subjects imagine what a piece of paper would look like when it is unfolded after folding and cutting it as shown at the top. They then select one of the possible solutions. Thorndike, R. L., Hagen, E. P., Sattler, J. M. (1986). The *Stanford-Binet scale of intelligence*. Riverside, IL: Chicago.

Figure 2: An Item From Raven's Progressive Matrices



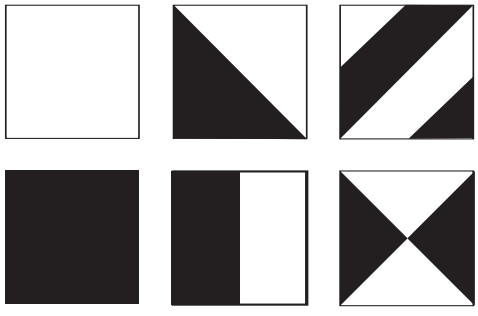
Subjects select one of the possible solutions to complete the pattern shown in the three by three matrix. Example from Gregory, R. J. (1996). *Psychological Testing: History, Principles, and Applications* (2nd ed.). Boston: Allyn and Bacon.

horizontally according to rules of logic (add the figures, subtract the figures, enlarge the figures in specific ways). Such tasks do not require flipping and turning objects mentally, nor doing so in sequential steps. Other tasks, such as the Pattern Analysis subtest of the Stanford-Binet, also do not qualify as spatial-temporal, because, while they do require flipping and turning objects mentally, they provide a model for subjects to match and compare while solving the task.[16] Figure 3 presents a sample item of the Pattern Analysis subtest.

Another important variation in replications is that studies employed different types of control conditions, including silence (used in about three-quarters of the experiments), audio tapes of instructions for relaxing that were designed to lower blood pressure (used in about half of the experiments), natural and man-made sounds (five of thirty-six experiments), texts read aloud (three of thirty-six experiments), and music that researchers thought was not complex enough or sufficiently like Mozart to enhance spatial-temporal skills (used in about one-fourth of the experiments). (Note that experiments often used more than one control condition.) For example, five of thirty-six experiments used a piece by Philip Glass called *Music with Changing Parts* that is almost hypnotically repetitious, and others used various "relaxing" music

Figure 3: An Item from the Pattern Analysis Subtest of the Stanford-Binet: Fourth Edition.

Subjects have several cubes, each of which has these designs on the six sides.



Subjects use their cubes to create images such as this one:

Thorndike, R. L., Hagen, E. P., Sattler, J. M. (1986). *The Stanford-Binet Scale of Intelligence*. Riverside, IL: Chicago.

(one was described as "angelic female voices"). Still others used disco and rock music, presumably thought to be distracting.

I conducted six preliminary analyses to determine whether experiments with such diverse controls could be combined responsibly into a single analysis. The first two preliminary analyses replicated and compared results from my sample to a previous meta-analysis,[10] and the other four directly compared scores on spatial-temporal tasks given after listening to different control conditions for those studies that used more than one control (Silence versus Relaxation tapes, Silence versus Noise, Silence versus Nonenhancing music, Relaxation versus Nonenhancing music).

The Music versus Silence analysis yielded a moderately sized average effect of $r = .24$ (equivalent $d = .48$), compared to Chabris's small average effect equivalent to $r = .07$ (equivalent $d = .14$). The Music versus Relaxation analysis yielded a moderate-to-large average effect of $r = .33$ (equivalent $d = .70$), compared to Chabris's similarly sized average effect equivalent to $r = .29$ (equivalent to $d = .57$). Because my sample is more representative of all the studies conducted on this research question (due to the exhaustive nature of the search and the fact that I included both published and unpublished studies), my results are more likely to represent the true effect size for the theoretical "universe" of studies on this research question.

Note that the relative size of the effects for the first two preliminary analyses is similar to Chabris's analysis (with Music versus Silence having a smaller effect than Music versus Relaxation). At face value, this finding lends support to the arousal theory—that music enhances spatial performance because it arouses, and, unless over-stimulated, an aroused person performs better on tests—since we might expect relaxation to produce lower arousal, on average, than merely sitting in silence. However, the third preliminary analysis suggests that arousal does not account for the difference in effect sizes, because when scores following silence and scores following relaxation were compared directly, they were essentially the same ($r = -.02$, with the negative sign indicating that scores following relaxation were actually slightly higher on average). The remaining preliminary analyses suggest that the differences in scores following various controls that could be directly compared were not consequential or systematic (Silence versus Noise, $r = .02$; Silence versus Nonenhancing music $r = -.05$; Relaxation versus Nonenhancing music, $r = -.02$). As a result of these analyses, I determined that the various control conditions used in the experiments produced essentially similar results and, thus, that these experiments could be combined legitimately into one single analysis. Including all of the identified experiments lends my analysis considerable statistical power and summarizes all the laboratory data with adults that

I could identify in relation to the question of music's enhancing effect on spatial task performance.

The first main analysis (thirty-six experiments, 2,469 subjects) compared tasks that qualified as spatial-temporal (31/36) to other types of spatial measures (5/36). The moderately sized and highly generalizable mean effect ($r = .22$, $d = .46$, ninety-five-percent confidence interval $r = .14$ to $r = .31$; t test of the mean $Zr = 5.34$, $p < .0001$) results from higher effect sizes in experiments that used spatial-temporal measures. The average weighted effect of the experiments employing spatial-temporal measures alone is $r = .20$. Experiments employing only nonspatial-temporal measures yielded a weighted average effect of $r = .04$, and experiments that used a combination of spatial-temporal and nonspatial-temporal measures showed an intermediate effect size (weighted $r = .15$). Thus, this analysis allows the conclusion that music's influence is specific to spatial-temporal, rather than to all types of spatial measures.

The second main analysis included only those thirty-one experiments (2,089 subjects) that employed measures that qualified as spatial-temporal. Again, the analysis showed a moderately sized relationship between listening briefly to music and enhanced performance on spatial-temporal measures ($r = .25$, $d = .50$), which is highly generalizable (95% confidence interval: $r = .14$ to $r = .35$; t-test of the mean $Zr = 4.84$, $p < .0001$). However, two problems limit conclusions that can be drawn.

First, the effect sizes of the individual studies varied too much to be considered as a sample from a single population of studies (Range: $r = -.20$ to $r = .67$, SD $= .25$, heterogeneity test, $\chi^2_{(30)} = 101.90$, $p < .0001$), and only some of the variation could be accounted for by moderator variables. Of the seven potential moderator variables identified, four did not influence the size of effect significantly (type of enhancing music used, subject gender, carry-over from previous spatial activation, and publication status). The remaining three did explain some of the variation. Experiments that employed a Relaxation tape control did have larger than average effect sizes ($r = .34$). However, since my preliminary analysis that directly compared scores following Silence versus Relaxation showed no difference in scores, it is likely that unidentified procedures of the laboratories that used relaxation as a control account for the systematic differences in effect sizes, rather than the control condition itself. This conclusion is affirmed by the results of a sensitivity analysis in which I temporarily removed the labs that contributed five or more experiments using a relaxation control. Both the Rauscher studies (average $r = .40$) and the Rideout studies ($r = .42$) had higher than average effects.

Such an observation leads to speculation about the procedures used by various labs. An analysis of study quality showed that experiments with stronger designs (that is, designs that were less vulnerable to threats of internal validity) had higher average effects, and both the Rauscher and Rideout experiments ranked average or above on these criteria. Thus, the difference remains unexplained by quality and cannot be attributed to errors by the researchers. Rather, the most likely explanation for the effect is that these two laboratories emphasized to subjects the importance of attending to the music closely. It is possible that doing so allowed the music to have an effect, while other experimental procedures allowed subjects' attention to wander. Such an explanation should be addressed in the design of future studies.

The second limitation is that a mechanism could not be unequivocally identified as causing the effect. Experiments did not provide enough data to explore plausible alternate hypotheses to the "trion" priming model, alternatives such as arousal and its related preference and mood theories, the effect of the element of rhythm as the link between musical and spatial processes, or the variation resulting from subjects who are musically sophisticated and trained to listen analytically compared to non-trained musicians as subjects.

In summary, these findings are of scientific interest, because the highly significant, moderately sized effect indicates that a relationship does exist between musical and spatial reasoning, as far as can be assessed from the studies conducted to date. It appears that spatial and musical processing areas of the human mind/brain are not entirely independent, but it is uncertain whether they influence each other because they are nearby, such that activation of one "primes" activation of the other, or because they overlap, such that development of certain musical processing areas would simultaneously develop the particular type of spatial reasoning defined as spatial-temporal. Further research needs to disentangle the cognitive mechanism that causes the effect. For example, neither priming model—either Shaw's "trion" or Parson's "rhythm" models—is conclusively affirmed or refuted, although both remain promising. In addition, future research needs to distinguish the effect conclusively from potential artifacts of procedures (e.g., subjects' attention to musical stimuli, or subject or experimenter effects that align results with unconscious expectations of subjects or researchers) or research design variations (e.g., control stimuli that are equally preferred by subjects or that can be measured as equally arousing or mood-altering). The analysis does not have direct implications for education, since the experiments were not about learning, but rather about how the human mind processes two types of information, musical and spatial. However, the result does suggest that studies in which subjects are taught

music could plausibly result in spatial learning. A group of such studies was synthesized in my second analysis.

Does Learning to Make Music Lead to Improved Spatial Reasoning?

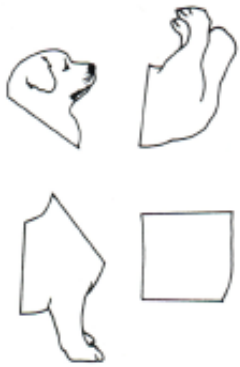
The search for my second analysis identified nineteen studies in which children ages three through twelve engaged in programs of active music instruction for up to two years. (One study lasted for three years, but I was only able to analyze results from the first two years[19]). Studies were conducted in schools or other instructional settings and used a variety of musical pedagogies and measures of spatial reasoning. To be included in the analysis, studies had to have one or more control conditions, with or without an alternate treatment. About one-third had an alternate treatment for controls consisting of instruction in language, instruction in reading or math on the computer, or passive instruction in music. Almost all had a non-treatment control (17/19), either in addition to a treated control group or as the only comparison group.

I defined active music instruction as including combinations of the following: singing, playing musical games, learning notations, improvising or composing music, playing instruments, and moving responsively to music, including clapping. The instruments used in the programs were combinations of voice, piano, xylophones, snare drums, and classroom rhythm instruments (triangles, tambourines, rhythm sticks, finger cymbals, hand-chimes, and bells).

Measures used in the studies varied widely, and because the results of the "Mozart effect" analysis indicated that only spatial tasks defined as spatial-temporal were enhanced by music, I conducted three analyses defined by the types of tasks employed as measures. The first analysis included studies that employed tasks that qualified as spatial-temporal, the second included studies employing tasks that qualified as nonspatial-temporal, and the third included studies that employed a variety of spatial tasks that could not be clearly distinguished by the criteria for spatial-temporal tasks.

The first instructional analysis included fifteen studies (701 subjects) employing such spatial-temporal tasks as the Object Assembly subtest from the WPPSI-R or WISC-III, in which children assemble a puzzle of a familiar object without seeing a model of the completed image. (See Figure 4 for a sample item.) Studies using other tasks were also analyzed, including a program designed by Matthew Peterson in Gordon Shaw's lab (Spatial-Temporal Animation Reasoning or STAR) and spatial subtests of other standardized tests for children (i.e., Developing Cognitive Abilities Test, the Wide Range Assessment of Visual Motor Abilities, and the Kaufman, Woodcock-Johnson, and McCarthy batteries). The average effect size was large by meta-analytic standards ($r = .37$, $d = .79$), and the results were highly generalizable (t-test

Figure 4: An Item from the Object Assembly Subtest of the Wechsler Preschool and Primary Scales of Intelligence, Revised.



Subjects assemble a simple puzzle without a model of the completed image.
Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary Scale of Intelligence*. New York: The Psychological Corporation.

of the mean *Zr* was 7.50, $p < .0001$). Most interestingly, despite great variation in the music programs and spatial-temporal measures employed, there was relatively little variation in effect size among the studies included. All had effects greater than zero, the 95% confidence interval was $r = .26$ to $r = .48$, the SD was less than half the size of the effect at .16, and a heterogeneity test was decidedly not significant with $\chi^2_{(14)} = 20.37$, $p = .12$. We can conclude from these results that the analysis is highly robust.

Meta-analysis is important not simply for determining average effects, but also for exploring the reasons for those effects. I examined seventeen potential moderator variables through contrast analysis. The most interesting finding is that thirteen of these moderators proved not to influence the size of the effect systematically, even though many of them are factors that often have been found to influence learning. These potential moderators include socioeconomic status, duration of instruction, parental involvement, test reliability, teacher and experimenter expectancy effects (unconscious expectations of subjects or experimenters that bias results), the Hawthorne effect (a tendency of any new program to have a positive impact), methods of group assignment, and study quality. In addition, and of particular interest to music educators, keyboard instruction proved no more influential than the other forms of active music instruction tested, despite a reasonable assumption that the spatial layout of the keyboard might be an important contributor in enhancing spatial

outcomes. In addition, effect sizes did not vary for those studies that used different keyboard instruments (pianos and xylophones), nor for studies that either did or did not use responsive movement in the music program, nor for those that either did or did not ask students to create or improvise musically. In other words, the large effect found for the analysis is very stable in relation to a host of variables that might have affected it one way or the other. The effect is not an artifact.

There were, however, two moderator variables that definitely had an impact on the size of effect. Effect sizes were somewhat larger in studies with individual rather than group lessons, and in studies in which children learned standard notation (rather than either no notation, or preparatory types of notation such as Kodaly hand signs). However, the more important finding, in my view, is that large effects were obtained in both group and individual formats (group lessons $r = .32$, individual lessons $r = .48$) and with and without standard notation (no notation $r = .36$, standard notation $r = .39$).

There were also two moderator variables that were nearly significant. The first is the publication status of the article (published articles $r = .29$, unpublished articles $r = .47$). Because most (11/15) studies in the analysis appeared in 1998 or 1999, however, it seems unlikely that many of them will remain unpublished. Thus, I doubt that the results of this contrast index a tendency of unpublished studies to inflate the effect. This opinion is confirmed by the quality analysis that showed no difference in average effect sizes between studies with higher and lower ratings on threats to internal validity. The other variable of interest was subject age (comparing three- to five-year-olds to children six years of age or older). Since the comparative effect sizes of the two groups were fairly large (three to five years, $r = .44$, ≥ 6 years $r = .27$), it is worthy of note. Future research should test whether enhancing effects from music programs are greater on younger children, as is the case here.

The second instructional analysis (five studies, 694 subjects) included studies with Raven's Matrices as the outcome measure. I expected a lower effect size, based on the results of the contrast on measures in the "Mozart effect" analysis, which found a lower weighted average effect ($r = .04$) for nonspatial-temporal measures compared to spatial-temporal measures (weighted $r = .20$). The average effect for the nonspatial-temporal measures analysis in the instruction studies ($r = .08$, $d = .16$) was much lower than the average effect of the spatial-temporal measures analysis ($r = .37$, $d = .79$). The average weighted r was even lower ($r = .03$, $d = .07$), which may be the more informative statistic, since four of the studies were similar in size (ranging from 147–179 subjects) and only one differed (forty subjects). The effect was not generalizable (the 95% confidence interval spans zero at $r = -.10$ to

$r = .27$, t-test of the mean $Zr = 1.23$, $p = .29$), and the studies were from a single population ($\chi^2_{(4)} = 5.72$, $p = .22$). This result provides support for the claim that spatial-temporal tasks are influenced by music training, but not nonverbal tasks, such as Raven's, that rely more on general logic.

The third instructional analysis included nine studies (655 subjects) that employed a range of spatial measures not readily classifiable as either spatial-temporal or nonspatial-temporal. Thus, this analysis tested whether the enhancing effects of music instruction extend beyond spatial-temporal measures to other, less clearly defined, types of spatial reasoning. Some studies mixed spatial-temporal and nonspatial-temporal measures (e.g., those that used several spatial subtests and only reported a global score), some used tests that may be spatial-temporal but that are difficult to classify (e.g., Children's Embedded Figures Test, "drawings and words presented in lacunary and ambiguous form"[20]), and one used a task that relies mainly on spatial memory (Bead Memory task from the Stanford-Binet: Fourth Edition). The average effect ($r = .26$, $d = .55$) is lower than the effect in the spatial-temporal analysis, but it is still of moderate size. In addition, it is generalizable (95% confidence interval $r = .16$ to $r = .36$; t-test of mean $Zr = 6.11$, $p = .0003$) and represents a single population of studies ($\chi^2_{(4)} = 8.87$, $p = .35$). From this we can conclude that music instruction may not be limited to spatial-temporal tasks but may enhance some other spatial reasoning more broadly. Further research is needed to affirm this finding, however, since the measures are quite diverse.

Cautions in Interpreting and Using these Results

In light of the current educational climate in which positive results from research have been overstated, I urge considerable caution in interpreting these results.

Researchers will realize that the lack of mechanism for the "Mozart effect" finding means that the effect is still questionable. Future research may yet demonstrate that the effect is an artifact of research design. However, the most important message from the "Mozart effect" analysis is that, to date, the effect appears to hold up, but that future research needs to test specific hypotheses about the mechanism underlying this effect. Further, we must realize that this laboratory finding with college students implies nothing for the education of children, much less infants in utero. If parents or teachers wish to play classical music for themselves or their children, they should by all means do so. But they should not expect that listening to music alone will aid children's future scores on standardized tests of academic achievement.

For the instructional analysis, there is a solid, generalizable finding that active instruction in music—not listening alone, although listening is a

component of such instruction—enhances performance on a specific type of spatial task classified as "spatial-temporal." Further, the third instruction analysis for mixed spatial measures suggests the possibility that this enhancement may extend more broadly to some nonspatial-temporal forms of reasoning, although not to matrices tasks (as shown in the second analysis).

However, there are still important questions about the value to education of such an effect. Remember that not all types of music programs have been tested, and that, in fact, the musical treatments combined may be different from each other in important ways. More research specifying the components of music instruction is needed to clarify just what teachers and students do in music instruction that aids skill in spatial reasoning. Further, the music studies that I analyzed were only for students between ages three and twelve, so we cannot generalize to infants, toddlers, or adolescents. Further, because the spatial tests were conducted within a few weeks of the end of the music instruction, we do not know how long any enhancing effect lasts. And because only one longitudinal study extending beyond two years currently exists, and it showed that students without music instruction caught up to those with piano instruction during the third year of instruction,[19] we do not know if music instruction is effective in fostering spatial reasoning after the first two years of instruction.

Perhaps even more important is the question of whether the effects of music instruction on spatial tests translate to better success in school. They might, or they might not. First, "real world" spatial problems, whether found in mathematics or the block corner or the ball field, may or may not be predicted by success on paper and pencil or table-task tests such as those used in these studies. Second, a corollary to this problem is that many classrooms do not give students a chance to use spatial skills, because instruction may not offer an opportunity for students to apply spatial reasoning to school subjects. In such cases, unfortunately, enhanced ability would not necessarily lead to improved success in school. To reap the benefits of any enhancement of spatial reasoning resulting from music instruction, therefore, schools would need to insure that instruction emphasizes spatial approaches to learning. Third, because spatial reasoning is multidimensional (consider the differences in designing a bridge, packing a car trunk, or finding your way around a new city, for example), it is not clear where the effects of the specifically "spatial-temporal" tasks would show up.

Finally, I urge caution in justifying music programs on the basis of their "bonus" effects. The benefits of music—those dispositions and concepts that are at the core of the discipline of music, such as understanding melody and harmony, thinking about the implications of combining timbres or of rhythmic variations, interpreting the aesthetic and narrative meanings of

musical pieces, or expressing personal convictions of thought and feeling through musical elements—are worthy in their own right. All societies throughout time have had music in their cultures, and all humans are born with the capacity to perceive and appreciate music. Thus, it is the responsibility of education to develop this potential in our citizenry, for the good of individual children and for our collective good. An education without music and the arts is an impoverished education, and educators who understand these arguments must take the lead in educating those who make decisions. Music is as "basic" as science and mathematics, and it is not a "frill." While there may be "bonus" benefits of music education that teachers of other subjects can exploit in helping students engage with and learn other subjects, "core" reasons must remain the justification that music educators use and the standards to which music educators are held responsible. Unless music educators retain vigilant integrity to their own discipline, they unwittingly support a mistaken belief that music is a handmaiden to other, more important learning.

Endnotes

1. Hetland, L. (2000a). Listening to music enhances spatial-temporal reasoning: Evidence for the "Mozart effect." *Journal of Aesthetic Education* 34 (3/4), 105–148.
2. Hetland, L. (2000b). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education* 34 (3/4), 179–238.
3. Karma, K. (1979). Musical, spatial, and verbal Abilities. *Bulletin of the Council* for *Research in Music Education* 59, 50–53.
4. Leng, X., and Shaw, G. L. (1991). Toward a neural theory of higher brain function using music as a window. *Concepts in Neuroscience* 2 (2), 229–258.
5. Rauscher, F. H., and Shaw, G. L. (1998). Key components of the Mozart effect. *Perceptual and Motor Skills* 86, 835–841.
6. Shaw, G. L. (2000). *Keeping Mozart in Mind*. San Diego: Academic Press.
7. Salomon, G. and Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist* 24 (2), 113–142.
8. Rauscher, F. H., Shaw, G. L., and Ky, K. N. (1993). Music and spatial task performance. *Nature* 365 (6447), 611.
9. Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: BasicBooks.
10. Chabris, C. (1999). Prelude or requiem for the "Mozart Effect"? *Nature* 402, 826–827.
11. Nantais, K. M., and Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. *Psychological Science* 10 (4), 370–373.
12. Steele, K. M., Bass, K. E., and Crook, M. D. (1999). The mystery of the Mozart effect: Failure to replicate. *Psychological Science* 10 (4), 366–369.
13. Parsons, L. M., Martinez, M. J., Delosh, E. L, Halpern, A., and Thaut, M. H.

(1999). *Musical and visual priming of visualization and mental rotation tasks: Experiment 1*. Manuscript in preparation, San Antonio: University of Texas.

14. Rauscher, F. H, Shaw, G. L., Levine, L. J., Wright, E. L., Dennis, W. R., and Newcomb, R. L. (1997). Music training causes long-term enhancement of preschool children's spatial-temporal reasoning. *Neurological Research* 19 (1), 2–7.
15. Reimer, B. (1999). Facing the risks of the "Mozart effect." *Music Educators' Journal*, 37–43.
16. Thorndike, R. L., Hagen, E. P., and Sattler, J. M. (1986). *The Stanford-Binet Scale of Intelligence*. Riverside, IL: Chicago.
17. Raven, J. C. (1986). *Raven's Progressive Matrices*. San Antonio, TX: The Psychological Corporation.
18. Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn and Bacon.
19. Costa-Giomi, E. (1999). The effects of three years of piano instruction on children's cognitive development. *Journal of Research in Music Education* 47(5), 198–212.
20. Zulauf, M. (1993/1994). Three-year experiment in extended music teaching in Switzerland: The different effects observed in a group of French-speaking pupils. *Bulletin of the Council of Research in Music Education* 119, 111–121.

COMMENTARY

Research and Justification in Arts Education: An Ill-Fated Romance

BENNETT REIMER
Northwestern University

When I'm asked to offer remarks about some topic or issue, I usually try to give my reflections a title. That helps me concentrate my mind: it gives me direction and focus. The title I've chosen for this response to the two papers I was assigned allows me to deal with what interests me most about the studies we're considering here: their use as a basis for justifying the arts in education. My interest in research in the field of the arts has mostly to do with whether such research can help us do a more effective job of enhancing people's satisfactions from the arts, and, if it can, how we may best take advantage of what research suggests so as to improve our educational practices. Given that orientation, that is, a devotion to what has often been called aesthetic education, and my hopes that research can be a powerful guide for providing it effectively, I find myself dealing, in this assignment, and in a good deal of research on the arts over the past several years, with a focus opposite to that which I cherish and to which I have devoted my career.

In this I am quite unexceptional. Many, perhaps most, arts educators (I'm tempted to say practically all arts educators) believe that the arts (and, of course, their particular art above all) are so valuable to humans for their distinctive characteristics as to warrant their inclusion as a basic component of education. In this cause an ocean of blood has been shed during the course of well over a hundred and fifty years.

So it's a bit ironic—amusing in a kind of twisted way—that we are going through a period in which research, which has only fairly recently been embraced by the arts education fields as possibly being relevant to their desires, has seemingly turned those desires on their head. Massively, pervasively, and at an astonishingly popular level, research in the arts has given the impression that it has flip-flopped in its purpose, focusing on how the arts can enhance just about everything one can imagine other than aesthetic satisfactions. Never mind that most research in arts education has nothing to do with the effects of arts study on academic achievement. All that research is completely unknown to the public. In the popular perception, and trickling down alarmingly into the arts education fields themselves and to some efforts in those fields to justify their existence, recent research has seemed to

establish—conclusively—that the arts are valuable because they contribute—significantly and causally—to learnings having nothing whatsoever to do with the quality of aesthetic/artistic interactions. Those contributions to academic learnings, it is hoped by some and perhaps many, will, finally, provide the arts with the justification they need in order to be accepted as necessary subjects in the school curriculum. Our romance with research seems to have left us embracing a lover whose mind is clearly not concentrated on the aesthetic/artistic benefits of the arts to which our hearts have for so long been devoted. O jilted love!

Disappointed lovers, not always in a clear state of mind, can rationalize their situation to make it seem to be what they had hoped for. Of course! It's all for the best! Our lover's eyes may not be on us, but in gazing into another's eyes they may find there our own true value, thereby returning to us what we crave. If the arts can enhance SAT scores, creative thinking abilities outside the arts, spatial-temporal reasoning, mathematics skills, reading ability, verbal skills, and cognitive functions in a host of non-arts subjects, all these will finally provide us with the esteem, the respect, the affection, if not love, that we have yearned for all these years, and we will be redeemed in our self-estimation. Pathos, indeed. I envision a "personals" ad in the newspaper. Arts education, sensitive, fun-loving, creative, heart in its right place, seeks love (will settle for respect) from anyone willing to accept it for what it is. Committed relationship desired. One-nighters considered.

While my heart dwells primarily on the arts as unique domains for creating and sharing human meanings, my mind, and some of my heart, is and always has been enamored with quantitative research—with the realities, the factualities, it attempts to deal with. So when I examine the research on music's effects on spatial-temporal reasoning part of me is able to put aside my concern for the aesthetic/artistic functions of the arts—the basis for their existence, after all—and to immerse myself with great interest in the methodologies and findings having to do with brain function; in this case interactions between music listening and spatial-temporal reasoning and music performance and spatial-temporal reasoning. These are topics compelling in and of themselves, quite apart from matters of aesthetics, of education, and of justification. Whatever we can learn about how our brains work is worth learning; in fact is terribly important for us to learn. If we take the admonition "know thyself" seriously, we must learn as much as we can about the mechanisms allowing us to know ourselves. So let me, for a moment, focus on the research I was asked to review—the two chapters by Lois Hetland on music and spatial reasoning—as strictly research intended to explore an interesting, even intriguing, aspect of brain function.

I admire a great deal about the studies Hetland reviews so thoroughly and painstakingly. They exhibit a high degree of imagination in their clever attempts to pin down several very elusive sets of human capacities and to get a handle on how they might or might not interrelate. But we must remember that listening to music and performing music are each, in and of themselves, among the most complex, most multifaceted, most culturally saturated, and most individually determined endeavors in which humans engage. They each involve mindfulness at its highest levels, which includes and subsumes, by necessity, the feelings and the body as integral components. "Music-think," in any of its manifestations—composing, performing composed music, improvising, listening, conducting, and so forth—depends on a unity of mind, body, and feeling in the cause of creating and sharing meaning as only sounds organized for that purpose are capable of doing. That endeavor takes place at three interactive levels: the universal level reflecting the inherent capacity of all humans to be musical, the cultural level in which that generic capacity is given particularity in the socially constructed regulatory system each culture has devised, and the individual level, in which all humans both adopt and adapt their musical culture to serve their singular, personal proclivities and capacities.

We are presented, in music, with an activity as rich and dense and complicated as anything humans attempt to do. This activity is a function of mind, not simply of brain. That distinction is crucial. Mind is dependent on brain function, of course, but far surpasses it in both breadth and depth. Here is how Antonio Damasio puts it, in his 1994 book *Descartes' Error: Emotion, Reason, and the Human Brain*:

> *There may be some Cartesian disembodiment behind the thinking of neuroscientists who insist that the mind can be fully explained solely in terms of brain events, leaving by the wayside the rest of the organism and the surrounding physical and social environment—and also leaving out the fact that part of the environment is itself a product of the organism's preceding actions. I resist the restriction not because the mind is not directly related to brain activity, since it obviously is, but rather because the restrictive formulation is unnecessarily incomplete, and humanly unsatisfactory. To say that mind comes from brain is indisputable, but I prefer to qualify the statement and consider the reasons why the brain's neurons behave in such a thoughtful manner. For the latter is, as far as I can see, the critical issue.*[1]

When dealing with the arts, human functioning at the level of the thoughtful is indeed the critical issue. Research at the level of brain function

must, because of its focus, deal with particles of mind, perhaps eventually being able to be aggregated to approach the level of the mindful, but realistically limited to what neurological research is interested in—human functioning at the brain activity level. Even at the level of brain activation, no research modality, no research technique, no research tool, is capable of embracing and accounting for the totality of a phenomenon as complicated as how the brain is activated when it is engaged in mindful human pursuits. So selections must be made. From the whole, a part, or a few parts, must be chosen for examination. And from all the many possible ways to examine those chosen aspects, only one or a few examinations can be made; those that are feasible given our limited research knowledge, our restricted choices of subjects, limited amounts of time and money, and the severe restrictions in the ways available to us to measure—to specify with exactitude—what is actually going on in even a modest set of mindful behaviors and what effects are actually being produced by them. What I am suggesting here is that brain research on a high-order human cognitive capacity such as musical experience, itself presenting complexities of such magnitude as to force us to atomize our focus, compounded by seeking relationships between that capacity and others perhaps equally complex, leaves us in a position of great vulnerability as to any expectations of attaining much more than modest insights, if that, into selected, highly restricted, probably isolated aspects of the phenomena to which our research efforts are directed.

These, as we all know, are the realities of brain research—the tough conditions every brain-function researcher must face. That requires us to appreciate such research, to appreciate the courage it takes, the skills and the cunning it calls upon, and the modesty it requires, given that it attempts to open windows onto the most complex phenomenon in nature of which we are aware—the workings of the human brain. We must be grateful to those who devote themselves to such difficult work, and who present us with findings they have been intelligent enough and lucky enough to have hit upon, however insecure, partial, puzzling, and, yes, even inconsequential they may be. It is easy to criticize researchers for not presenting us with significant, flawless, and guaranteed results. I am suggesting that we must temper our expectations to what is achievable, and be generous in our respect for what is, after all, a daunting task.

In that spirit of appreciation I would like to offer a few suggestions for further conduct of research such as I was asked to review that might lead it in the direction of mind—in the direction of what Damasio calls "the critical issue."

In regard to the so-called "Mozart effect" caused by listening to his (and perhaps others') music, it is important to recognize that listening to music, in addition to being complex as I have already mentioned, requires an

act of meaning-construction on the part of the listener. To portray music listening as essentially passive, as is often done, seriously misconstrues what it entails. Music listening indeed can be passive, meaning inattentive and perfunctory, as can looking at visual art, watching a dance or a play, or reading a novel or poem. The same is true of "making" music, which, in many people's minds mistakenly means only performing, forgetting that every musical involvement requires music to be "made." Performing can also be inattentive and perfunctory, as every performance teacher knows full well. The same applies, of course, to "making" in all the arts.

The simple admonition "pay attention to the music," given by Rauscher and Rideout to their subjects, seemed to produce in them higher effect sizes than found in other studies where music was played without that reminder. That's very interesting. Just imagine what might happen if genuine instruction in listening, such as music educators are supposed to be able to provide, which activates perceptual, contextual, affective, and kinetic dimensions of creativity in the listening act, were supplied in abundance to subjects, or at least to the levels supplied in the performing instruction studies. Would serious music education, devoted to enhancing musical intelligence as it is particularly manifested in the music listening role, affect spatial-temporal measures more positively than when no instruction is given, or equally positively, or less positively, or, perhaps, negatively? And what dimensions of active listening are most salient to spatial-temporal responses?

Apply the same ideas beyond the performance of composed music, to improvising, and to composing. We would then begin to better understand how genuine musical experience, such as music education attempts to cultivate, might be implicated in other brain functions compatible with and affected by musical experience when such experience is at its higher rather than lower levels. We would get closer to understanding how learning affects brain function both within and outside musical experience itself. We would get closer to understanding the implications of brain function for mind.

But a serious effort must be made to reduce the level of triviality of the measures used to ascertain effects in this research. The most common measure of "spatial-temporal reasoning" has been the Paper Folding and Cutting subtest from the Stanford-Binet IQ test. The concept of intelligence embodied in that test and others like it represents a now discredited or at least severely questioned construct of what intelligence is, how it works, and how it might be measured. To call by the term "reasoning" a task such as, for example, identifying what pattern is made in a folded paper napkin with holes cut in it and then unfolded, is to stretch to unacceptable limits any concept of reasoning that might be meaningful outside a behaviorist laboratory. Unreal

in its construal of what musical experience entails (in these cases listening and performing), unreal in its approach to measurement, in which an atomistic instrumentation substitutes for the role-based making of meaning that reasoning requires, we are left with a few shards, a few glimmers, of possible relationships, at the sub-meaning level, between two artificially represented phenomena of mind.

I believe we can do better than this if we elevate our concept of measurement to include aspects of thoughtfulness—aspects of mind. I believe that is doable, if we can relax, to some reasonable extent, the statistical/experimental mind-set governing present brain research. We can indeed admire the pioneering research that has taken us this far, and we can hope that more holistic approaches, added to the more traditional experimental paradigm so far employed, will provide a much-needed dimensionality to research on this topic.

To return to the issue of justification, we cannot ignore the reality that, in the world of education policy-making, the research we are reviewing has caused a mountain to be made of a molehill. I do not fault the researchers for this, although a bit more evident modesty on their part might have forestalled the painful and embarrassing exaggerations and misconceptions that have occurred. Hetland ends her chapter on Music and Spatial Reasoning by saying that my concern, expressed in my article "Facing the Risks of the 'Mozart effect,'"[2] that "music educators might be held accountable for their students' having learned spatial skills, which could corrupt the quality of programs in developing musical understanding, can be laid to rest for now." She must excuse me for continuing to not rest easily. Her very next sentence explains why. "However," she says, "music educators and policy makers will need to be vigilant to ensure that music programs are designed to teach music, and music educators are held accountable for musical, rather than spatial, understanding."[3] Indeed. Vigilance *is* called for, rather than rest.

Throughout the entire Winner-Hetland report we are wisely reminded, time and again, that the arts cannot be and should not be justified on the basis of any possible and as yet only poorly and partially substantiated contributions they might make to academic improvement. Fine. But let's get real. All of us here can produce, at a moment's notice, dozens if not hundreds of recent examples of claims for the value of the arts in education because of "research-based proof" that they contribute to academic learnings, to "general intelligence" (as if there were such a thing), and to a host of other values for which they are a magic bullet, and, on that basis, that they can be justified as part of schooling. The Winner-Hetland study demonstrates that, in educational policy-making, facts play a minor role as compared with credulity. This

is the case, unfortunately, not only in regard to the arts, but also in areas as "high stakes" as teaching verbal literacy. I urge you to read the paper "The Politics of Literacy Teaching: How 'Research' Shaped Educational Policy" in the November 1999 *Educational Researcher*, which reports on how completely misguided representations of supposedly unequivocal findings about the superiority of a code-emphasis approach to reading led to adoption of curricula for which no evidence could actually be found in the research.[4]

So apparently we are not alone in being in a compromised position in regard to research and justification. Nevertheless, we must acknowledge the moral dilemma with which this romance we have stumbled into confronts us (as romantic involvements often do.) Do we try to take advantage of this unexpected interest in the arts, despite its shaky foundation, on the argument that anything at all that keeps our foot in the door is worth having, warranted or not, and we can then surreptitiously go about our aesthetic business? Or do we stand firm on aesthetic justifications, refusing to have our virtue (such as it is) sullied by lesser (at least to us) concerns?

If we do take that high road, we are then faced with the obligation to explain, in a clear, understandable way, just what it is about the arts that, on their own terms, is valuable enough to deserve precious educational time and effort. While philosophers of arts education continue to probe the complexities of that issue, the profession as a whole is, in fact, able to articulate, quite simply and precisely, explanations that most people are likely to understand. For example, the Winner-Hetland report says that "The arts are a fundamentally important part of culture, and an education without them is an impoverished education leading to an impoverished society. Studying the arts should not have to be justified in terms of anything else. The arts are as important as the sciences: they are time-honored ways of learning, knowing, and expressing."[5] For purposes of public advocacy a statement like this, and many others similar to it, seems to me intellectually respectable and quite easily comprehensible. Yet offering such explanations, in a great variety of ways, as arts educators have done for a very long time and continue to do, seems not to be persuasive enough to provide a secure and important place for the arts in education. This is what drives arts educators crazy—that we exist in a culture more enamored with secondary benefits of the arts than with primary ones. It tempts them to abandon or at least weaken their resolve, and to reach out to whatever might improve their sense of self-worth, even if their virtue becomes slightly tarnished by doing so.

I have taken the position[6] that it is possible to accommodate both intra- and extra-aesthetic benefits of arts education in advocacy efforts, and even in curriculum, so long as our fealty to aesthetic learnings remains

resolute. On the continuum from complete capitulation to unwarranted but popularly effective claims, to complete rejection of any concerns other than aesthetic ones, each of us will have to find our level of moral acceptability. It comes down, I suppose, to what we'll do for love.

Endnotes

1. Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G. P. Putnam's Sons, 250–251.
2. Reimer, B. Facing the risks of the "Mozart effect" (July 1999). Grandmaster Series, *Music Educators Journal* 86, no. 1. Reprinted in *Arts Education Policy Review 101*, no. 2, November/December, 1999. Reprinted in *Phi Delta Kappan*, 81, no. 4, December 1999.
3. Hetland, L. (2000). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education* 34 (3/4), p. 224.
4. Allington, Richard L. and Woodside-Jiron, Haley. The politics of literacy teaching: How "research" shaped educational policy. *Educational Researcher* 28, no. 8, November, 1999, 4–13.
5. Winner, E. and Hetland, L. (in press). The arts and academic achievement: What the evidence shows. Arts Education Policy Review (executive summary).
6. In, for example, "Facing the Risks . . . ," n. 2, above.

COMMENTARY

Revolution in Math Education Fueled by Music Training

GORDON L. SHAW
University of California, Irvine

The role of arts education has taken on enormous political aspects, as this conference has made perfectly clear. Many arts academics at this conference rightly argue that arts should be taught for their own sake and not because they can enhance performance in non-arts domains. This is like saying you are for motherhood or apple pie. However, the reality is that arts are being dramatically cut in public schools in recent years, in particular, music. To ignore the potential benefit to all our children of having music training enhance their ability to grasp difficult math through their innate spatial-temporal reasoning abilities is untenable in our high-tech society.

Lois Hetland, in her very thorough meta-analyses, "Listening to music enhances spatial-temporal reasoning: evidence for the 'Mozart effect'"[1] and "Learning to play music enhances spatial reasoning,"[2] has provided very important evidence for the causal link of music enhancing spatial-temporal reasoning. However, meta-analyses as used here,[1, 2] by their very nature, average over a number of widely different types of experiments and tend to wash out an effect. Further, these meta-analyses do not allow the examination of all the extremely relevant theoretical, behavioral, and neurophysiological studies that present a coherent picture. This is particularly clear when looking at the meta-analysis by Kathryn Vaughn, "Music and Mathematics: Modest Support for the Oft-Claimed Relationship."[3] The full benefit of music training for enhancing math performance must involve further specific, well-documented educational steps as outlined below.

Rather than replying in detail to these related three meta-analyses,[1, 3] I will review our entire collaborative twenty-year project, "Music as a window into higher brain function," so that you can see the complete picture. Otherwise, it looks like a bunch of dramatic, yet puzzling and separate pieces of research. Then I will detail the striking results of the MIND Institute's Music Spatial-Temporal Math Program and describe its present "educational reality" phase in twelve schools with over twelve hundred second- and third-graders. If this educational reality proves successful, then all children should be given this opportunity to use their innate spatial-temporal abilities to develop their math reasoning and thus be competitive in our high-tech world.

The connection between math and music goes back to the ancient Greeks, who considered music as one of the four branches of math. It has long been observed that there is a correlation between music training for children and their math performance. The predictive theory of music training causally enhancing math abilities began with the seminal work of Xiaodan Leng.[4]

A series of landmark papers started with the trion model[5, 8] of higher brain function based on the Mountcastle[9, 10] columnar organization principle of the cortex. In the structured trion model, the innate internal language of higher brain function is represented by spatial-temporal memory firing patterns. The brain's innate ability to relate (through symmetry operations) these memory patterns is the unifying physiological mechanism of higher brain function.[7, 8] The finding that the structure of these memory patterns and their symmetry relationships were that of recognizable styles of music led to the realization[4] that "music could be used as a window into higher brain function" and the prediction that specific music could enhance spatial-temporal reasoning. Spatial-temporal reasoning involves maintaining, transforming, and comparing mental images in space and time using symmetry operations, and it is fundamental in learning and using math (and science) concepts. Children can understand the math concepts through the spatial-temporal approach so that they then can solve the standard quantitative equations and word problems.

The trion model[4, 8] motivated and guided these innovative, collaborative behavioral and neurophysiological experiments, all with major results.

"Mozart Effect" Listening Experiments:

1. "Mozart effect" causal experiments with Frances Rauscher:2, 11, 12 College students, after listening to the first ten minutes of the Mozart Sonata for Two Pianos in D major (K. 448), showed subsequent short-term (ten- to fifteen-minute) enhancement of spatial-temporal reasoning. These results received an enormous amount of attention in 1993 and were called the "Mozart effect" by the media.
2. Alzheimer patients,[13, 14] after listening to the Mozart *Sonata*, had enhanced short-term spatial-temporal reasoning.
3. Exposure in epileptic patients, even in a coma, to the Mozart *Sonata* reduced neuropathological spiking activity.[15, 16]
4. Long-term exposure to the Mozart *Sonata* enhanced learning of a maze by rats; the enhanced performance lasted more than four hours after the last exposure to music.[17] You definitely do not want to give a child such long-term exposure to any music. However, this does open the possibility of some long-term enhancement just from a moderate amount of listening to specific music over a long time period. (We are

just starting the first experiments ever done with young children to test this possibility of enormous interest.)

5. EEG (surface brain wave) coherence study[18] gave evidence for a carry-over from the Mozart *Sonata* listening condition to the subsequent spatial-temporal task in specific cortical regions. This gave the first neurophysiological evidence for the Mozart effect. The idea is that the Mozart *Sonata* "resonates"[8] with the innate columnar structure of the cortex.[6] Functional Magnetic Resonance Imaging (fMRI) studies comparing cortical blood flow activation by the Mozart *Sonata* versus other music gave striking results.[19, 20] In addition to expected fMRI activation in cortical regions associated with music, substantial activation was found in cortical regions important for spatial-temporal reasoning. Further fMRI studies (along with EEG studies) should not only be extremely valuable in determining the neurophysiological basis for the Mozart effect, but in determining which other music might give similar enhancements in spatial-temporal reasoning.

Direct tests of the trion model of higher brain function:

6. A consistency test of the trion model[4, 8] and the Mountcastle[9, 10] columnar spatial-temporal code for higher brain function came with the demonstration of highly accurate mental rehearsals.[21] The temporal durations of mental rehearsals of pieces of music lasting up to a few minutes were extraordinarily reproducible.
7. Neurophysiological studies by Mark Bodner[22] confirmed, as predicted, the presence of families of spatial-temporal firing patterns related by symmetries during higher brain function.

Music training enhances spatial-temporal reasoning and learning math for children:

8. Preschool children[23] who received piano keyboard lessons for six months improved dramatically, with the effect lasting for several days, on an age standardized spatial-temporal reasoning task.
9. Spatial-Temporal Animation Reasoning, STAR, is an ingenious math software developed by Matthew Peterson. We contend that these nonverbal math games utilize spatial-temporal abilities built into our structured cortex. Inner-city second-graders given piano keyboard training along with STAR training scored significantly higher[24] on proportional math and fractions than children given control training along with STAR. A special interactive version of STAR by Peterson is available on CD-ROM enclosed in Shaw.[8] We contend that STAR not only exploits spatial-temporal reasoning but specifically those operations that are especially innate to our structured columnar brain.

10. Second-graders given piano keyboard training, STAR lessons, and lessons bringing this spatial-temporal approach into their regular language based math curriculum did as well on advanced math concepts as fourth-graders from a higher socio-economic school not having our training.[25] They also greatly enhanced their performance on nationwide standardized math tests (Stanford nine). Our three-component Music Spatial-Temporal Math Program has an enormous potential for enabling all children to learn difficult math concepts.

These eleven distinct highly successful experiments all are extremely relevant to our general theme, "Music as a Window into Higher brain Function." They are all supportive of the basic underlying trion model predictions that motivated them and must be considered as an entire coherent body of knowledge.[8] The meta-analyses[1, 3] cannot take these distinct experiments into account and thus do not offer pathways for further research. This fundamental scientific research continues: we are just at the very beginning. It is just as if we have found a vein of gold leading to the gold mine of understanding higher brain function. The trion model provides a road map to our understanding of higher brain function. We all know that the connections in the young child's brain are rapidly being modified as a result of experience and learning. The crucial distinguishing concept of the trion model (as derived from the Mountcastle columnar principle of cortex) is that the infant's brain starts with an innate, well-defined, and common internal neural language and grammar.[26] As we learn even more about this innate neural language from experiment and theory, we will better be able to help children learn to think and reason.

This summarizes the present key milestones showing how music not only helps us understand how we think, reason, and create but how it can enhance these higher brain functions through our innate spatial-temporal abilities. Now let's turn to our present Music Spatial-Temporal Math Program.

Revolution in Math Education

As we enter the third millennium, if one thing is clear, it is the growth of technology. It is impossible to predict all the ways that (just to give a few examples) biological engineering, computers, lasers, microelectronic circuits, and robotics will effect our lives and job structure in the next decades. However, it is safe to say that the number of semi-skilled jobs of the past will continue to diminish. Furthermore, businesses are no longer competing with those next door or even in the next state, but throughout the world. There will mainly either be high-tech jobs or low-tech ones. I suggest that a child who does not learn math will not have a choice.

The Third International Math and Science Study, which involved 500,000 students from forty-five countries, showed[27] that American eighth-graders perform poorly (twenty-eighth) in math and, in particular, in proportional reasoning (comparing ratios of quantities). For decades, researchers have understood that proportional reasoning is notoriously difficult for children and for many adults.[28] If your child cannot do proportional math, the child will have difficulty with even the most basic science, for example, rate problems in biology, chemistry, and physics. Once a child is lost at these beginning levels, the higher-level math, science, and engineering courses may prove to be insurmountable. This will essentially rule out any high-tech job.

The American educational system relies almost entirely on language-analytic reasoning and grossly neglects the complementary spatial-temporal reasoning. This is a major flaw in our educational system.

Spatial-temporal reasoning involves maintaining, transforming, and comparing mental images in space and time using symmetry operations, and it is crucial in math and science. Both types of reasoning, language-analytic and spatial-temporal, are necessary and complementary. Language-analytic reasoning, for example, is better suited to solving equations or word problems that lead to a quantitative result. Spatial-temporal reasoning is more useful, for example, in reasoning about chess and understanding math and science concepts.

The most brilliant scientist of the twentieth century, Albert Einstein, made many references to his own dependence on spatial-temporal reasoning:[2]

> *The words or the language, as they are written and spoken, do not seem to play any role in my mechanism of thought. The entities that seem to serve as elements in thought are certain signs and more or less clear images that can be voluntarily reproduced and combined. . . . Conventional words or other signs have to be sought for laboriously in a secondary stage.*

Richard Feynman, perhaps the most important and influential physicist of the second half of the twentieth century, described the relationship between spatial-temporal and language-analytic reasoning:[3]

> *It's a crazy mixture of partially solved equations and some kind of visual picture of what the equation is saying is happening. Strange! I don't understand how it is that we can write mathematical expressions and calculate what the thing is going to do without being able to picture it.*

Despite its crucial role in the thinking processes of most scientists, mathematicians and engineers,[31] spatial-temporal reasoning is grossly neglected in United States schools. This does not make any sense.

The three-component Music Spatial-Temporal Math Program[25] redresses this neglect because it capitalizes on spatial-temporal reasoning in helping children master difficult math concepts. The Music Spatial-Temporal Math Program has the following components: (1) Piano keyboard training, which enhances the child's innate cortical ability to solve spatial-temporal tasks; (2) STAR software, which allows the child to learn difficult math concepts using spatial-temporal reasoning; (3) Math Integration, which bridges the spatial-temporal approach to the standard language-analytic methods of symbols, simple equations, and word problems.

After eight months of participation in the Music Spatial-Temporal Math Program, our class of eighteen second-graders in the 95th Street School dramatically increased their average on the national Stanford nine math scores from the thirtieth percentile to the sixty-fifth percentile; half of these children performed in the top twentieth percentile nationally.

In this first year of the Music Spatial-Temporal Math Program, we concentrated on the advanced math concepts of proportional reasoning, fractions and symmetry that cannot be learned by rote. These concepts are not covered in any depth at the second grade, but are typically introduced in some detail in the fourth and fifth grades and continued through eighth grade. Our Advanced Math Concepts test assessed students' ability to apply their understanding of proportional math, fractions, symmetry, graphs and pre-algebra problems in a manner that did not make demands on language skills. When we wanted to gauge results on the Advanced Math Concepts test, our comparison group consisted of fourth-graders from the higher performing school SA. The second grade Music Spatial-Temporal Math class from the 95th Street School performed at the same level as fourth-graders from the higher socio-economic school SA on our Advanced Math Concepts test. This result will be a benchmark for our future work.

One huge remaining hurdle is to demonstrate that schools with a variety of socio-economic characteristics can use the Music Spatial-Temporal Math Program and obtain results similar to those we had obtained. Also, in order for such a program to "scale up" it must be economically viable: the dollar cost must be such that all schools can afford it, and the benefits in math performance must be large enough to clearly warrant the investments of time and resources. This is an educational "reality test," not a scientific test.

We developed second grade curricula for all three components of the Music Spatial-Temporal Math Program. We have trained each of the class-

room teachers in our program. Each one now teaches the math integration, the STAR lessons (or acts as the aid to a STAR computer instructor), and acts as the aid to the piano keyboard instructor. We now have a third-grade curriculum and thus can follow our present second-graders into the third grade. We have thirteen schools with over fifteen hundred second- and third-grade children in our program this school year 2000–2001.

Our goal in the next few years is to have a complete Music Spatial-Temporal Math Program for grades kindergarten through six available to schools around the country. The Music Spatial-Temporal Math program is intended to complement, not replace, the standard language-analytic math programs. Clearly, excellent teachers and the determination of the students are crucial for learning difficult math concepts. However, the teachers and children must be given the appropriate tools.

Golf Analogy to Music Spatial-Temporal Math Program

It is always interesting to include a simple yet good analogy to allow one to immediately say "Oh, yeah, I get it," when a new complex concept is being explained. So, here is our golf analogy to our three component Music Spatial-Temporal Math Program:

You are going to learn to play golf and you decide to play with just your right arm. You soon realize that this will not work and you must start using both hands. Step one is to get your left arm in shape: this is analogous to the music training getting your brain in shape for spatial-temporal reasoning. Step two is to get used to using your left arm for the golf swings: this is analogous to STAR training using spatial-temporal reasoning to master the difficult math concepts. Step three is to now use both hands together: this is analogous to the math integration, which brings the spatial-temporal approach to learning the math concepts together with the standard language-based math programs in the school system. It makes sense to start playing golf using both arms: Clearly it makes sense to start learning math using both spatial-temporal and language based learning.

The Music Spatial-Temporal Math Program presents the beginning of a revolution in education allowing all children, especially those from disadvantaged backgrounds, to learn difficult math and science concepts using their innate spatial-temporal reasoning abilities. All children must be given the opportunity to learn to think mathematically and thus be able to compete in our high-tech world.

Finally, since setting research priorities was a large component of this conference, I propose that there is strong indication that a program for neurophysiologically handicapped children of music training along with training on

an appropriate version of STAR would prove to be of large clinical benefit in helping them to better cope with their world.[8] Certainly, this is a great use of research funds in testing such a program of huge potential benefit. I would hope that arts academics would support and welcome such a program.

Acknowledgments

The work that I summarized here, "Music as a window into higher brain function," represents the long-term, collaborative research efforts of many highly talented individuals from beginning undergraduates to world-renown scientists. In particular, I wish to thank my colleagues at the MIND Institute (see our web site www.MINDinst.org and my book[8] for more details). The funding for this research during the past ten years all comes from the private sector. I gratefully thank these visionaries: the principal ones include John Chambers, Paul Folino, Ralph and Leona Gerard Family Trust, Herbert Lucas, Lyons Share Foundation, National Association of Music Merchants, Marjorie Rawlins, Samueli Foundation, Seaver Institute, Ted and Janice Smith and the Texaco Foundation.

Endnotes

1. Hetland, L. (2000). Listening to music enhances spatial-temporal reasoning: Evidence for the "Mozart effect." *Journal of Aesthetic Education* 34 (3/4), 105–148.
2. Hetland, L. (2000). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education* 34 (3/4), 179–238.
3. Vaughn, K. (2000). Music and mathematics: modest support for the oft-claimed relationship. *Journal of Aesthetic Education* 34 (3/4), 149–166.
4. Leng, X. and Shaw, G. L. (1991). Toward a neural theory of higher brain function using music as a window. *Concepts in Neuroscience* 2, 229–258.
5. Shaw, G. L., Silverman D. J., and Pearson, J. C. (1985). Model of cortical organization embodying a basis for a theory of information processing and memory recall. *Proceedings of the National Academy of Sciences* 82, 2364–2368.
6. Shenoy KV, Kaufman J, McGrann JV and Shaw, G. L. (1993). Learning by selection in the trion model of cortical organization. *Cerebral Cortex* 3: 239–248.
7. McGrann, J. V., Shaw, G. L., Shenoy, K. V., Leng, X., and Mathews, R. B. (1994). Computation by symmetry operations in a structured model of the brain. *Physical Review* E49, 5830–5839.
8. Shaw, G. L. (2000). *Keeping Mozart in Mind*. Academic Press, San Diego.
9. Mountcastle, V. B. (1978). An organizing principle for cerebral function: the unit module and the distributed system. In: Edelman, G.M., Mountcastle, V.B. (eds.), *The Mindful Brain*, Cambridge: MIT: 1–50
10. Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain* 120, 701–722.
11. Rauscher, F. H., Shaw, G. L., and Ky, K. N. (1993). Music and spatial task performance. *Nature* 365, 611.

12. Rauscher, F. H, Shaw, G. L., and Ky, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Toward a neurophysiological basis. *Neuroscience Letter* 185, 44–47.
13. Johnson, J. K., Cotman, C. W., Tasaki, C. S., and Shaw, G. L. (1998). Enhancement in spatial-temporal reasoning after a Mozart listening condition in Alzheimer's disease: A case study. *Neurological Research* 20, 666–672.
14. Johnson, J. K., Shaw, G. L., Vuong, M., Vuong, S., and Cotman, C.W. (2000). Short-term improvement on a spatial- temporal task after music listening in Alzheimer disease: A group study. Submitted for publication.
15. Hughes, J. R., Daaboul, Y., Fino, J. J., and Shaw, G. L. (1998). The "Mozart effect" on epileptiform activity. *Clinical Electroencephalography* 29, 109–119.
16. Hughes, J. R., Fino, J. J., and Melyn, M. A. (1999). Is there a chronic change of the "Mozart effect" on epileptiform activity? A case study. *Clinical Electroencephalography* 30, 44–45.
17. Rauscher, F.H., Robinson, K. D., and Jens, J. J. (1998). Improved maze learning through early music exposure in rats. *Neurological Research* 20, 427–432.
18. Sarnthein, J., von Stein, A., Rappelsberger, P., Petsche, H., Rauscher, F.H., and Shaw, G. L. (1997). Persistent patterns of brain activity: An EEG coherence study of the positive effect of music on spatial-temporal reasoning. *Neurological Research* 19, 107–116.
19. Muftuler, L. T., Bodner, M., Shaw, G. L., and Nalcioglu, O. (1999). FMRI of Mozart effect using auditory stimuli. Abstract presented at the Seventh Meeting of the International Society for Magnetic Resonance in Medicine. Philadelphia.
20. Bodner, M., Muftuler, T., Nalcioglu, O., and Shaw, G. L. (2000). FMRI study of the Mozart effect: Brain regions important for spatial-temporal reasoning. Submitted for publication.
21. Brothers, L. and Shaw, G. L. (1989). The role of accurate timing in human performance and the code for higher cortical function. In Cotterill, R. (ed.). *Models of Brain Function*. Cambridge, United Kingdom: Cambridge University Press.
22. Bodner, M., Shaw, G. L., Gabriel, R., Johnson, J. J, Murias, M., and Swanson, J. (1999). Detecting symmetry patterns in EEG data: A new method of analysis. *Clinical Electroencephalography* 30, 143–150.
23. Rauscher, F. H., Shaw, G. L., Levine, L. J., Wright, E. L., Dennis, W. R., and Newcomb, R. L. (1997). Music training causes long-term enhancement of preschool children's reasoning. *Neurological Research* 19, 2–8.
24. Graziano, A. B., Peterson, M., and Shaw, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal reasoning. *Neurological Research* 21, 139–152.
25. Peterson, M., Bodner, M., Cook, S., Earl, T., Hansen, J. S., Martinez, M. E., Rodgers, L., Vuong, S., and Shaw, G. L. (2000). Music spatial-temporal math program for second-graders enhances advanced math concepts and Stanford 9 math scores. Submitted for publication.
26. Sardesai, M., Figge, C., Bodner, M., Crosby, M., Hansen, J., Quillfeldt, J. A., Landau, S., Ostling, A., Vuong, S., and Shaw, G. L. (2000). Reliable short-term memory in the trion model: Toward a cortical language and grammar. *Biological Cybernetics*, 174–182.
27. National Research Council (1999). Global perspectives for local action: Using

TIMSS to improve American mathematics and science education. Washington, D.C.

28. Karplus, R., Pulos, S., and Stage, E. K. (1983). Early adolescents' proportional reasoning on "rate" problems. *Educational Studies in Math* 14, 219–233.

29. Hadamard, J. (1945). *The Psychology of Invention in the Mathematical Field*. Princeton, NJ: Princeton University Press.

30. (1998) Kemp M. Feynman's figurations. *Nature* 394, 231.

31. Ferguson, E. S. (1977). The mind's eye: Nonverbal thought in technology. *Science* 197, 827–835.

COMMENTARY

The Effects of Early Music Experiences

RICHARD COLWELL
Music Educators National Conference

Lois Hetland has conducted two meta-analyses, one on the effects of music listening and one on the effects of music performance on measures of spatial awareness. Both have been carefully conducted; no music educator or educational researcher is likely to find fault with her procedures or her interpretation of her findings. The music education research community has produced no other meta-analysis of comparable quality. My comments will, therefore, be focused on the meaning of her research for instruction and research in the profession. As with any meta-analysis, Dr. Hetland was limited by the quality and quantity of studies that have addressed the connection between music experiences and measures of spatial awareness. Thus, my first comment is that Hetland's results indicate that "messing around" with music performance leads to improved competence on measures of spatial awareness. I'm using the term "music experiences" and "messing around" rather than "music instruction" because the authors of the research she analyzed provided only the most minimal description of the instruction conducted and its goals, and no valid assessment of any musical objectives. I suspect that the instruction was typical of most music experiences at this preschool and kindergarten level and that any direct instruction on improvement in spatial awareness was minimal or absent. My second comment is that a meta-analysis such as the two studies of Hetland is important as it focuses on an outcome for music education that is influencing the public's perception of the value of music education.

Music Educators as Researchers

As one who reads almost all of the research studies conducted in music education, I was pleasantly surprised that a careful meta-analysis produced consistent results from a variety of studies, some of which were not well designed. That is, of course, what meta-analysis is supposed to accomplish. Music educators have produced some excellent research results but, on the whole, research has never been a priority in our field and few individuals have developed competence in conducting or interpreting research. Our research consists primarily of doctoral dissertations, and in recent years their quality has been declining. Meta-analyses on the primary objectives of music education are needed, and these would provide Hetland with in-field comparisons.

Comments on the Project

The authors of the larger study indicate that a journal entitled *Music in Education* was searched. This journal was unknown to me and to those of my colleagues whom I queried. I determined that it was a British journal that ceased publication nearly thirty years ago and can be found only in seven United States libraries, three of them in Washington, D.C. That's thorough research!

Of interest would be the extent to which the keyboard experience of students in the studies required reading treble and bass clefs, as that task would provide assurance of a possible connection between keyboard and spatial tasks. Music educators have conducted a few studies on music reading using eye trackers: single-line music reading follows roughly the same sequence of saccades as found in language reading, but when the reader must read bass and treble clefs, both vertical and horizontal eye movements are required. If "keyboard" is to be the intervention, the experience should be authentic to piano instruction. The use of xylophones and similar classroom instruments should not be combined under the classification of keyboard in a meta-analysis as any results (musical or nonmusical) would be confounded. When using Orff-type instruments music educators teach more by rote and with simplified notation than they do in teaching individual or group piano.

I applaud the use of Robert Rosenthal's contrast analysis for Hetland's primary statistical procedure. Perhaps surprisingly, few music educators have used this technique; we seem to be "spooked" by the need to search for any possible interaction and thus lose sight of the primary comparison.

The use of length of instructional time as a consideration raises an interesting question. My personal belief is that length of instruction is most often only nominal data in music; the frequency and quality of music instruction varies considerably. One who has taken piano lessons for ten years may not be as competent–by any measure–as another who has studied for five years. If length of instructional time is a consideration, it must be accompanied by some measure of competency gained.

The most important point for music education is Hetland's discussion of the laboratory effect. This effect appears to have resulted from the clarity of instructions provided the students. We have a small body of research in music education that indicates that when attention is focused on a sound, a different response is produced than is obtained with casual listening. When one listens for specifics, selective attention changes the neuronal response to acoustic stimuli because there is a decreased responsiveness to other frequencies or attributes of the sound. This change is due to the focus.[1] Within the past five years, there has been considerable interest in the research findings

of McPherson,[2] Sloboda,[3] Davidson,[4] Lehmann, [5] and others who are demonstrating the importance of and difference in results when practicing is focused.

What could be misleading for the casual reader would be any impression that the studies analyzed by Hetland contrasted music instruction with no music instruction. I have long wanted to conduct, at any public school level, such a contrast study if for no other reason than to obtain an idea about what music understandings students simply absorb from their culture and from maturation. Armed with that knowledge we could focus our instructional efforts on what the student does not know. Students have major encounters with music outside the school, but school music is not Walkman® music. Few public (or private) schools would sanction a research project where random students receive instruction and others are deprived of music even with a guarantee of a double dose of instruction the following semester or year. Interestingly, there is no moral dilemma for administrators in reducing or eliminating music completely from the curriculum, but if instruction is available in any form, it must be fairly apportioned. The comparisons that Hetland reported were with pre-primary school students, students electing keyboard, and one non-United States research report, by Zulauf, where the context is unknown.[6]

Hetland's speculations on possible musical outcomes or at least experiences that might have been investigated seem to be focused on composition, creative imagery, and having students verbalize by reflecting on their musical experiences. These are of interest but of no greater importance than the teaching of the skills used in numerous musical experiences. In an investigation of this magnitude, the failure to provide any information about the attainment of any musical competencies must surely be seen as deserving of comment.

In the twenty-first century, there is considerable interest in and advocacy for nonmusical outcomes from music instruction. The relative importance of temporal or nontemporal visual-spatial competence as compared to numerous other likely outcomes that contribute to the many intelligences (a discussion of which would be a fascinating paper to write) is an unanswered question. Hetland focused on these two because of their present influence on arts policy, but there is also a need to look at relationships to scores on personality, sensory-motor, neuropsychological, other subtests to the WISC, and other abilities that educators believe are important or abilities that appear to have a narrow window of opportunity for the most effective instruction.

The reader must also be aware that, through no fault of Hetland's, students tested were primarily of one age group, and these effects might not result if the tests were replicated with older students.

A Sampling of Acoustic and Neurological Research in Music

The connection between music and the brain has not come as a surprise to music educators. Music educators have known for one hundred years that musical stimuli affect humans physically and psychologically. It requires only a crude study to demonstrate changes in bodily responses and a bit more data for evidence of changes in one's mental attitude as a result of musical stimuli. We have heart and pulse-rate studies, skin-conductivity studies, respiration studies, blood pressure, muscular tension and motor activity studies, motor-postural response studies, finger or peripheral skin temperature studies, blood volume studies, pupil dilation, pilomotor response, biochemical response, and even stomach contraction research.[7] Determining a relationship between music experiences and spatial-temporal ability is new but only one of many effects of music upon the human being. The human auditory system seems to include the functions of perception, attention, learning, and memory, thus any stimulus could be of major importance depending upon its type and strength. Unfortunately, few of the studies have been replicated, thus we know little about the specific circuits that underlie specific phenomena, and even when the circuits are known it is unclear why the circuit behaves the way it does. Today's cognitive approach to music education must consider the complexities related to musical perception. Pitch and meter are not the only factors in perception. Timbral aspects have been used to provide serial organization to music. The cognitive psychologist interested in music's effects will encounter a century's worth of interesting research, although little of it has been related to education. We can safely say that there is a spectrum of sensory, perceptual, emotional, cognitive, mnemonic, aesthetic, and performance processes involved in music besides the basic auditory stimulus.[8]

We can guess that the reason that one musical selection produced the effect under consideration and another did not was due to the very complexity of music and the many examples in the literature where what was perceived does not match with what was produced.[9]

This basic research that constitutes much of the field of psychology of music has been of little practical use to music educators and has not affected practices, the curriculum, or policy. The research from psychology of music that has been of most interest has been studies related to stage fright and that related to the acquisition of perfect pitch or high levels of relative pitch.

Developmental Studies

As I have suggested, there is a close relationship between cognition and perception, but they are not the same. Perception has been the focus of psychologists who recently have attempted to trace its development and relationship to

musical and nonmusical tasks. The first book on the psychology of musical development was published in 1986,[10] with the first large-scale studies of music perception by young children being that of Helmut Moog in the mid 1970s. Moog found that students who were handicapped from birth did not develop perceptual skills in the same sequence as the nonhandicapped.[11] He reported the phenomenon but did not explain it, a characteristic of most developmental research in music education. As linguists believe that infants must babble before they begin to talk, Edwin Gordon suggests that there is a music babbling stage prior to the production of discrete musical pitches.[12]

Infants at five months (according to some) can identify a new or different song as indicated by a heart-rate deceleration. These babies don't notice the insertion of a few notes in the melody but do notice a change in the contour of a song.[13] Studies show a considerable difference in the age at which music perception begins. For example, the ability to discriminate between major and minor mode in music occurs between the ages of three and eight, depending upon the research study. There seems to be agreement that eighteen months is the earliest age for producing discrete pitches but the major advancements in responses to music occur after the age of seven or eight. Jones suggests that the concept of meter develops after the age of nine and a half years;[14] Mary Louise Serafine finds that it is generally in place by the age of nine.[15] Thus, it appears necessary to relate the research in developmental psychology of music to an interpretation of the results of music experiences with the three- and four-year-olds used in the meta-analyses analyzed by Hetland. To what extent were the children in the studies analyzed by Hetland responding to acoustic stimuli? To what extent were they perceiving and understanding? The distinction between these two responses may be important for those who attempt to explain the results according to rhythm and similar theories.

The best-known developmental studies are those by Zimmerman and Secrest, Swanwick and Tillman, Serafine, Bamberger, Moog, Davidson and Scripp, and Trehub and associates.[16] None of these investigators attempted a sequenced instructional program; at most they taught students the task to be performed. These researchers do agree, however, that the music must be apprehended by the listener or the investigation is not about cognitive processes in music. To what extent does a three- or four-year-old child apprehend sounds in a musical-cognitive sense? We do not know. To think in music, one must organize the sounds perceived, usually by patterns. Is it possible that the five- and six-year-old students in the studies analyzed by Hetland were engaged in these musical/cognitive processes? Past research in the discipline would indicate the possibility is doubtful. This possibility does not preclude

the importance of music in early childhood education, but suggests rather that there are important premusical experiences that have an effect on later musical and nonmusical outcomes. A type of "music bathing" occurs with Muzak and many uses of music for therapeutic purposes. Without more research we cannot say whether the experiences with music engaged in by the students in the studies analyzed by Hetland were musical or a nonmusical "bathing." One can hear students much older than the students in these studies playing piano compositions without perceiving that there are melodic and rhythmic patterns that are intended to organize the sounds, and it seems entirely possible that one can have a listening experience by reveling in the interesting timbres without engaging any perceptual/cognitive competencies.

A Research Agenda

Education experiences should have multiple outcomes. We in music education have given scant thought about the priority of the possible outcomes that result from traditional or nontraditional school music experiences. Traditionally the most valued outcomes by the public have had some relationship to performance. Music education has long been comprised of two programs, but these have seldom been as distinct as they are at present. The general music program required of all students was initiated by Lowell Mason in 1837. It was based on attaining objectives of physical development (lungs and other vital organs), moral outcomes (happiness and contentment), and intellectual growth that Mason believed to be improved memory and attention span. We have had little agreement on the outcomes, or even the content, of the general music program since the beginning of the twentieth century when doubt as to the importance of music reading skills first surfaced. Throughout the twentieth century the general music program was accepted but unfocused and today consists of a variety of experiences. Hetland's meta-analysis, more than any other event, suggests the importance of knowing *why* all students should partake in general music experiences and demands that a rigorous research agenda support any adopted purposes.

The second music education program is the performance program, focused in the secondary school but found in the primary grades as well. Bands and orchestras (and later choruses) began as after-school activities around 1870, becoming part of the school day and curriculum shortly after the beginning of the twentieth century. The public seems satisfied that students in performing ensembles learn how to budget their time, to be responsible, to know the meaning of excellence, to work cooperatively, to finish tasks, and build self-esteem and self-confidence. Agreement on musical outcomes is markedly less clear. Recent indications that students who participate in these

ensembles for four years have higher scores on many standardized tests, including those that measure language arts and math abilities, raise research questions similar to those addressed by Hetland's meta-analysis. The importance of clarity of objectives and outcomes is evident in that the indescribable general music programs are presently struggling for time and support while the high school ensemble programs are at an all-time high in quantity and quality.[17] The attention given to music in the educational reform movement is due largely to claims about, on the one hand, improving temporal-spatial ability in the general music program and, on the other hand, higher SAT scores in the secondary performance program, giving rise to considerable confusion about the role of music education.

The research agenda in music education cannot be based upon changes in the curriculum that have occurred as a result of the crisis in education. Few of the educational reform issues of the past seventeen years, including detracking, theme-based curricula, changes in evaluation, and smaller schools, apply to music education.

In the history of music education, the initial and often seminal studies were done by psychologists with an interest in music. Hetland, Catterall, Sloboda, and others appear to be contributing to this tradition. Music educators must begin to accept responsibility for issues about which they should be well informed. There must be comparable research on the experiences that contribute to *musical* outcomes that are unique in the educational experience. Educational psychologists are interested primarily in the cognitive outcomes, but, as important as these are, they are not the sum total of outcomes of musical experiences. There must be research on the unique aspects of perception and on affect.

Research comparable to that investigated by Hetland needs to be conducted with the full range of heterogeneous students found in the schools. Cognitive research in music education has been focused on the youngest students primarily because this age is available. Once quality public performance is a focus, teachers have been unwilling to provide minimal research time.

Authentic music and authentic musical experiences are themselves complex, and when perceived by complex students raise research issues that have never been addressed.

A review of the research indicates that little is known about the psychology of musical development; indeed, Hetland's results show little difference has been found by student age or by the music experienced. To conduct dependable research, music researchers and doctoral advisers need to be as well-versed in research design and interpretation of results as are researchers in educational psychology. In the 1960s the federal government funded two

week-long research training institutes for forty doctoral advisers, one conducted by Jason Millman and Gene Glass in experimental research and one conducted by Robert Lathrop in descriptive research. This training, brief as it was, made a major difference in the quality of research produced by the doctoral students of the institute participants. At the same time, the *Bulletin of the Council for Research in Music Education* was established, a periodical that rigorously reviewed research, becoming the conscience of the profession. At present there is no means for a scrupulous attention to research findings similar to that the Getty Education Institute provided to critique the findings of Lois Hetland, Ellen Winner, and others involved with the project. Public and printed critiques of substance are needed if the profession is to advance in instruction or research.

A compelling argument can be made for research and assessment in the area of musical outcomes from education. Not only is there an absence of evaluated practice, there are in the doctoral dissertations of the past decade or so few dependent variables that have been systematically developed. A needs assessment for program evaluation would be very difficult to conduct.

One of the highest priorities for music education is research on teaching and learning how to listen to music. In the mid-twentieth century Leonard Bernstein attempted to structure the listening experience through his excellent Norton lectures; today every civic orchestra is attempting concerts for youth; but there has been almost no research in this area for more than a half century. There is a difference between learning to enjoy music and learning to perceive the qualities of music that make it great. Those differences need to be identified for purposes of research and instruction.

As Hetland's meta-analysis has demonstrated, the research agenda should not be based on folklore about music or the arts. There is a bit of magic to the musical experience, a magic that distinguishes music from visual arts. It is easier to be mechanical about production in visual art than it is with performance in music; with a skillful conductor and competent peers, participation in musical performance can approach the big thrill, the aesthetic experience.

Endnotes

1. Lipscomb, S. (1996). The cognitive organization of musical sound. In Hodges, D. (ed.). *Handbook of Music Psychology* (2nd ed.). San Antonio: IMR Press, 150; Weinberger, N. (1999). Music and the auditory system. In Deutsch, D. (ed.). *The Psychology of Music* (2nd ed.). San Diego: Academic Press, 55, 59.
2. McPherson, G. (1995). The Assessment of musical performance: Development and validation of five new measures. *The Psychology of Music* 23, 142–161.

3. Sloboda, J. A. (1991). Musical expertise. In Ericsson, A. and Smith, J. (eds.). *Toward a General Theory of Expertise: Prospects and Limits*. Cambridge, United Kingdom: Cambridge University Press.
4. Davidson, J., Howe, M., and Sloboda, J. (1995). What motivates instrumental learning. Paper presented at the Seventh European Conference on Developmental Psychology. Krakow, Poland, August 23–27, 1995.
5. Lehmann, A. (1997). Acquired mental representations in music performance: Anecdotal and preliminary empirical evidence. In Jorgensen, H. and Lehmann, A. (ed.). *Does Practice Make Perfect?* Oslo, Norwegian State Academy of Music.
6. Zulauf, M. (1993–1994). Three-year experiment in extended music teaching in Switzerland. *Council for Research in Music Education Bulletin* 119, 111–121. Her most recent two-year study found no difference between the experimental and control groups and she reports that "interest has shifted from the possible extra-musical effects to the benefits in the musical realm" (*Psychology of Music* 28 [2], 2000, 125).
7. Radocy, R. and Boyle, J. D. (1997). *Psychological Foundations of Musical Behavior*, Springfield: Charles C. Thomas, 276–280.
8. Weinberger, N. (1999). In Deutsch, D. (ed.), *The Psychology of Music* (2nd ed.), San Diego: Academic Press.
9. Duke, R. (1987). Musicians' perception of beat in monotonic stimuli. Unpublished research paper, MENC Southern Division Conference, Orlando, FL; Geringer, J. M., Madsen, D. K., and Duke, R. A. (1993/1994). Perception of beat note change in modulating tempos, *Council for Research in Music Education Bulletin* 119, 49–57; Schouten, J. F. (1938). The Perception of subjective tones, *De Koninklijke Nederlandse Akademic voor Wetenshappen Proceedings* 41, 1086–1093, cited by Pierce, J. (1999). The nature of musical sound. In Deutsch, D. (ed.). Op cit, 15–16; Weinberger, N. (1999). In Deutsch, D. (ed.). Op cit, 63, 81; Deutsch, D. (1999). Grouping mechanisms in music. In Deutsch, D. (ed.). Op cit, 309; Radocy, R. and Boyle, J. D. (1997). Op cit, 132; Deutsch, D. (1999). Grouping mechanisms in music. In Deutsch, D. (ed.). Op cit, 325; Stevens, S. S. (1934a). Tonal density. *Journal of Experimental Psychology* 17, 585–592; (1934b). The volume and intensity of tones. *American Journal of Psychology* 46, 150–154; (1935). The relations of pitch to intensity, *Journal of Acoustical Society of America* 6, 150–154.
10. Hargreaves, D. (1986). *The Developmental Psychology of Music*, Cambridge, United Kingdom: Cambridge University Press.
11. Moog, H. (1976). *The Musical Experience of the Preschool Child*, London: Schott, B., cited in Kassner, C. Research on music in early childhood. In Colwell, R. (ed.). *Handbook of Research on Music Teaching and Learning*, New York, Schirmer Books (1992), 639–640; Radocy, R. and Boyle, J. D. (1997). Op cit, 126.
12. Gordon, E. (1997). *A Music Learning Theory for Newborn and Young Children*, Chicago, GIA Publications, Inc, 4–6.
13. Chang, H. and Trehub, W. E. (1997a). Auditory processing of relational

information by young infants. *Journal of Experimental Child Psychology* 24, 324–331; (1977b). Infant's perception of temporal grouping in auditory patterns. *Child Development* 48, 1666–70.

14. Jones, R. L. (1985). A dialectic analysis of selected contradictions among definitions of meter in music. *Council for Research in Music Education Bulletin* 83, 43–56.
15. Serafine, M. L. (1988). *Music as Cognition: The Development of Thought in Sound*. New York, Columbia University Press, 223–228.
16. Pflederer, M. and Lee, S. (1968). Conservation-type responses of children to musical stimuli. *Council for Research in Music Education Bulletin* 13, 19–36; Swanwick, K. and Tillman, J. (1986). The sequence of musical development: A study of children's composition. *British Journal of Music Education* 3, 305–339; Serafine, M. L. (1988). Op cit; Davidson, L. and Scripp, L. (1989). Education and development in music from a cognitive perspective. In Hargreaves, D. J. (ed.). *Children and the Arts*, London, Milton Keynes, Open University Press; Moog (1976). Op cit; Trehub, W. E., Bull, D., and Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development* 55, 821–830; Bamberger, J. (1991). *The Mind Behind the Musical Ear: How Children Develop Musical Intelligence*, Cambridge, MA: Harvard University Press.
17. 2000 Survey of school music budgets. *Instrumentalist*, August 2000, 18–22.

IV.

Strengthening Verbal Skills Through the Use of Classroom Drama: A Clear Link. A Summary of a Meta-Analytic Study

ANN PODLOZNY
Researching Education and the Arts Project Team

Over the past thirty-five years, nearly two hundred experimental studies have assessed the relationship between drama instruction and various kinds of academic ability. These studies have tested the hypothesis that involvement in drama leads to improved abilities in other areas of the academic curriculum. More than forty percent of those studies examined the relationship between drama instruction and academic achievement in verbal areas, such as reading comprehension, oral language ability, and writing. There have been four substantial integrative reviews of the literature examining the effect of drama on cognitive outcomes (including two meta-analyses), and each reported primarily positive results. In the meta-analyses described here, I focused only on the relationship between classroom drama and verbal achievement.

The domain of classroom drama differs from the other outcomes explored in the REAP research (and described in this volume) in that it is explicitly designed to be part of the curriculum. Classroom drama differs from theater. By "theater," we refer to electives students take that involve play production. By classroom drama, we refer to acting out stories that are used in the regular academic curriculum. Classroom drama is used as a way of supporting the curriculum and is an integral part of the curriculum.

A meta-analysis published by Kardash and Wright in 1986 synthesized sixteen studies on transfer effects of classroom drama and reported a strong positive relationship between drama and a variety of cognitive outcomes, including reading, oral language development, self-esteem, moral reasoning and various drama skills. These researchers reported an average effect size of $r = .32$, which is fairly large and equivalent to a d of .67.

Kardash and Wright found that the strongest relationships between drama and other outcomes were those that came from published (rather than unpublished) studies, studies published earlier rather than later (that is, between 1970 and 1983), studies with younger rather than older children, and studies assessing typical rather than "special" populations (such as learning-disabled or low-SES children). They also found that the strength of the relationship between drama and other outcomes was positively related to the number of minutes of instruction. However, strangely, the strength of the relationship between drama and other outcomes was inversely related to the number of weeks of drama instruction.

A second meta-analysis was a dissertation authored by Conard in 1992. Conard examined the effect of classroom drama on a variety of language outcomes, as well as on students' self-concept and creativity. Like Kardash and Wright, Conard found a positive relationship between classroom drama and academic outcomes, with an average effect size of $r = .23$ (equivalent to $d = .48$), which is slightly lower than that reported by Kardash and Wright.

Consistent with the studies by Kardash and Wright, Conard found that higher effect sizes were associated with published studies, younger participants, and typical populations. Contrary to Kardash and Wright, however, Conard also found larger effect sizes associated with more weeks of drama instruction and less rigorously designed studies.

Neither Kardash and Wright nor Conard were able fully to account for the heterogeneity of their groups of studies, and neither were they able to tease apart any particular aspect of drama instruction that might predict achievement in the various outcomes. Finally, neither study compared across outcomes to see if certain areas of academic achievement were more strongly related to drama instruction than others.

I sought to replicate the findings of both of the previous meta-analyses, and also to extend and deepen their analysis by sifting the studies through a finer sieve. To begin with, I was able to locate eighty usable studies—four times more than in any previous synthesis. Additionally, I sorted the studies into seven different outcomes in the hope of teasing apart any effects that might be outcome-specific. I also performed a construct analysis on the generic label of "drama instruction," identifying three salient dimensions of drama. I identified these dimensions in the hope of examining whether one or more of them played a role in determining the strength of the relationship between drama and verbal outcomes.

The first dimension to consider is "enactment." All dramatic instruction must involve enactment of some pretend situation. This enactment, however, can take on a myriad of forms: children can re-create a story or

explore a theme through verbal enactment (by creating dialogue while sitting in a circle on the floor), or they can pantomime the actions of a story or theme without using words, engaging in physical enactment only. The enactment can either be performed by the child (what I call "self"-enactment) or puppets or other toys can be used (what I call "distanced from self"). Finally, enactment can include a combination of any of these four features (verbal action/self/distanced from self). Unfortunately, I was unable to test the role of the form of enactment because I found scant variation among the studies along the enactment dimension. Seventy-five of the eighty studies engaged children in verbal, physical, self-oriented drama. I mention it here, however, because this may be an important dimension to investigate in future research.

The second dimension to consider is the degree of structure in the enactment. When children are given a story or script to enact, the enactment is highly structured. In contrast, when they are simply given themes to act out, the enactment is unstructured. I coded each study as involving structured or unstructured enactment, or some combination. Finally, the third dimension consists of the teacher's level of involvement in the enactment. The teacher can engage as a participant, taking on a character and modeling the dramatic behavior for the children "in-role."[1] Or the teacher can work outside the dramatic frame, answering questions, assigning roles and encouraging dramatic behavior as a facilitator, while not actually playing a character or modeling dramatic behavior. Finally, the leader can be removed from the action, answering questions, but not serving as a driving force for the activity.

In addition to analyzing groups of studies in terms of different types of outcomes and dimensions of drama instruction, I also examined types of populations. While both previous meta-analyses combined "special" populations into one group, I distinguished children with learning disabilities from those with low-SES. Finally, and in some ways most importantly, I chose to code studies according to whether children were tested on material they enacted in their drama sessions (what I call "direct") or whether they were tested on new material (what I call "transfer"). The purpose of this distinction was to allow me to determine whether acting out a story simply helped children read and understand and recall that particular story better, or whether the experience of acting out a story helped children's verbal skills when applied to new material never enacted. Of course, the second possibility would be the more powerful finding.

I examined seven outcomes and performed a meta-analysis on each one. The seven outcomes were:

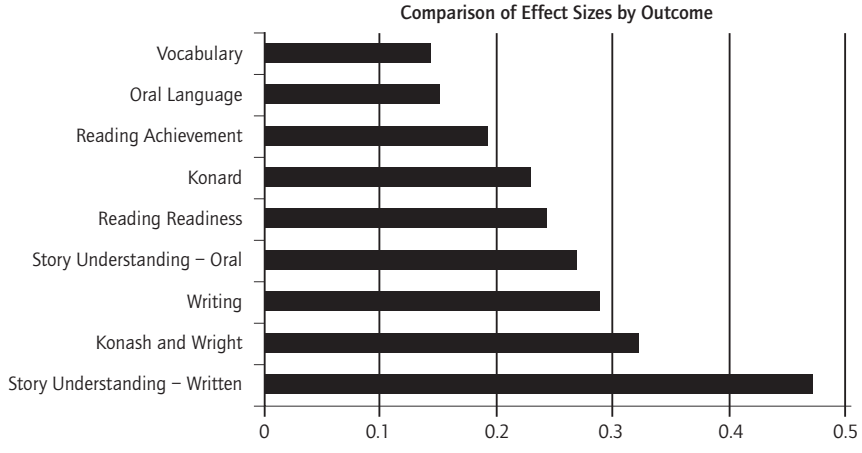
1. Oral measures of story understanding. Seventeen studies tested the effects of classroom enactment on story recall as measured orally. These studies tested recall both of the stories enacted and of new stories never enacted. Students in the drama group acted out the stories. Students in the control group had equal exposure to the stories but did not act them out.
2. Written measures of story understanding. Fourteen studies tested the effects of classroom enactment on story recall as measured in writing. The drama group read and then enacted the stories; the control group read the stories and then discussed them and were drilled on vocabulary words from the stories. In this body of studies, children were tested only on the stories enacted; thus transfer to new stories could not be tested.
3. Reading achievement. Twenty studies tested the effects of classroom enactment on reading comprehension. The drama group in these studies typically read a story or play and enacted it. The control group did not read the story or play but continued with their regular reading classes. Both groups were then given standardized reading comprehension tests. Thus in this body of studies children were always tested on new material. Hence, any effect demonstrates transfer of reading comprehension skills to new material.
4. Reading readiness. Eighteen studies assessed the effects of enactment on reading readiness in young children. Typically, children in the drama group heard a story and acted it out. Those in the control group either heard the same story and discussed but did not enact it; reenacted themes from field trips or other experiences (and hence did not hear the story); or engaged in cut and paste and categorizing activities (here they neither heard the story nor engaged in any enactment). This body of studies again only tested children on new material.
5. Oral language development. Twenty studies assessed the effects of classroom drama on oral language. Typically students in the drama group engaged in creative dramatics (storytelling, role-playing, puppetry) as well as discussion. The control group engaged in other arts besides drama (visual art, music, watching film strips). Later the oral language of all children was assessed, sometimes when talking about new material, other times when talking about the stories that they had enacted.
6. Vocabulary. Ten studies assessed the effects of classroom drama on vocabulary. In a typical study, children in the drama group engaged in creative drama activities, including role-playing, pantomime, move-

ment, and improvised dialogue. The control group had no special treatment. Later all children were given a vocabulary test. In some studies, the vocabulary came from the stories enacted (and here one would expect the drama group to do better since they were exposed to these words); in other studies, the vocabulary consisted of words that were not a part of the stories enacted.

7. Writing. Eight studies assessed the effects of classroom drama on writing skills (including audience awareness, organization, and elaboration). Typically children in the drama group first participated in a discussion about writing, and then engaged in improvisation, pantomime, and movement, developed story ideas, improvised the stories, and then drafted them. The control group also participated in a discussion about writing, but then they simply continued with their regular language-arts program before drafting their stories. Stories were then analyzed according to a narrative writing scale. In some of the studies, children wrote stories based on what they had enacted. In others, they wrote stories based on new material. These latter studies thus assessed transfer of skills acquired in drama to new material.

Results showed that classroom drama had a positive, robust effect on six of the seven verbal outcomes examined here. The size of this effect varied considerably by outcome, as shown in Figure 1. The largest effect size was found for studies assessing the effect of drama on written measures of story understanding, which yielded an average weighted effect size of $r = .47$. Studies assessing the effect of drama on writing, story understanding as

Figure 1



measured orally, and reading readiness followed (mean weighted $r = .29$, .27, and .25 respectively). Classroom drama was shown to improve reading achievement (mean weighted $r = .19$) and oral language (mean weighted $r = .15$). Vocabulary appeared to be enhanced by drama as well (mean weighted $r = .14$). However, this latter effect size, unlike the other six, was not statistically significant (the 95% confidence interval for this effect spanned zero). Hence research has not as yet demonstrated a reliable relationship between drama instruction and vocabulary development.

The type of plot used in the drama instruction influenced the effectiveness of the instruction. I had hypothesized that structured enactment, which requires participants to work within the story frame, increasing the children's exposure to story structure and familiarity with story concepts, would facilitate greater growth of verbal skills. This occurred when the outcomes measured were writing, reading achievement, and reading readiness.

For those studies measuring oral language development, however, I found that either unstructured enactment or a combination of structured/unstructured enactment was associated with larger effect sizes than structured enactment alone. This finding is understandable, however. The outcome of oral language development does not have anything directly to do with story structure and comprehension. Thus, it is likely that unstructured enactment's emphasis on extemporaneous and improvised speech is more facilitative than structured enactment, with its emphasis on staying within the confines of a particular story or script. It is also interesting to note that all types of plot were equally effective for oral measures of story understanding and vocabulary.

The role of the leader proved to be significantly related to the effectiveness of drama instruction only for those studies measuring story understanding. Following Jeffrey Dansky's "multi-stage" model of effects,[2] I had hypothesized that leader "in-role" might increase the occurrence and/or quality of dramatic play, which, in turn, might increase academic achievement. Indeed, it seems that this might be the case in the story understanding studies. I have two caveats, however: first, it is possible that increased attention to the students may be influencing the effect (although many studies tried to control for this, not all succeeded), and also, the story understanding studies were the only ones to report the role of the leader in detail. It was therefore difficult to classify the leader in many of the studies.

Recall that some bodies of studies compared children's verbal skills as applied to the stories enacted versus as applied to new materials. If students' verbal skills improve only when applied to the stories enacted, we would have to conclude that not much transfer had occurred. But if verbal skills improve even for new, non-enacted materials, we can conclude that

classroom drama does indeed lead to transfer of verbal skills. Studies whose outcomes were oral comprehension, vocabulary, oral language, and writing all allowed a comparison of verbal skills applied to enacted versus new materials. In three cases, students did as well on new material as on enacted material and in one case (oral language) students actually performed better on new material. In addition, the bodies of studies assessing the effect of enactment on reading achievement and reading readiness tested children only on new material, and both of these bodies of studies resulted in positive and significant effect sizes. Thus, we can conclude that the experience of acting out stories increases children's understanding of new stories; it increases their vocabulary, oral language, and writing skills, whether applied to the stories enacted or to entirely new material, and it increases their reading ability for new material.

Note, however, that studies whose outcome was story understanding as assessed in writing yielded the largest effect size overall, and in these studies, children were only tested on stories they had enacted. Likewise, the Reading studies (both Achievement and Readiness) required all participants to transfer skills from drama instruction to new material, and they also had lower effect sizes overall. Thus, looking across the seven meta-analyses, it appears that drama does have more power to inculcate skills applied to enacted texts rather than to inculcate skills that transfer to new texts. Because of the elusiveness of transfer, this is what one would expect. What is surprising and encouraging, however, is that the analyses also show that drama does promote skills that transfer to new material. What is remarkable is not that drama's strongest effects are the direct ones, but rather that drama does have the power to foster skills that transfer.

How much drama instruction is actually needed to promote achievement? The answer to this question is mixed. While it appears that more time spent with drama is important for reading readiness and vocabulary outcomes, those studies in which children spent less time with drama had higher effect sizes when the outcomes were story understanding, oral language development, and writing. However, the studies that found more drama time associated with higher effect sizes had very high amounts of drama instruction overall, averaging between 1200 to 1640 minutes of classroom time (about twenty to twenty-seven hours). On the other hand, the studies that found less drama associated with higher effect sizes had much shorter periods of instruction, averaging only 315 to 720 minutes of instruction (about five to twelve hours), or less than half the time spent in the other studies. This leads me to hypothesize, therefore, that there may be a plateau effect that occurs when instruction is of medium duration: a little bit of drama may lead to detectable

effects, but then one needs to increase the time spent with drama quite a bit to see an increase in this effect.

With respect to the age at which drama is most likely to result in enhanced verbal skills, the evidence was inconsistent. While the studies by Kardash and Wright and by Conard both found that drama was more effective for younger children, five of the seven meta-analyses showed no relationship between age and effect size. Of the remaining two, one showed that the effect was stronger for younger children (the outcome here was writing achievement), while the other found that the effect was stronger for older children (the outcome here was oral language).

Also contrary to both of the previous meta-analyses, five of the seven meta-analyses found that drama was equally effective for average, low-SES and learning-disabled samples. The remaining two analyses (those assessing written story understanding and reading achievement) found that drama was actually more effective in promoting verbal skills when the children involved were from low-SES populations. This finding is consistent with Smilansky's report that exposure to drama increases the achievement levels of poor students.[3] One explanation is that children from disadvantaged backgrounds may not have had the opportunity to engage in creative, dramatic play, nor to experience success through engaging instruction. Classroom drama instruction may provide a "boost" to these students, helping them acquire a deeper level of story understanding. There may, however, be a ceiling to these effects after a certain point in achievement. If average populations have already achieved this level, the effects of drama instruction may not be pronounced.

The most important finding of these meta-analyses is the demonstration that drama not only helps children master the texts they enact, but also often helps them master new material not enacted. The transfer of skills from one domain to another is generally not thought to be automatic: it needs to be taught. In the field of classroom drama, however, transfer appears to be naturally designed into the curriculum, even if teachers are not labeling it as such. If drama teachers did more to teach explicitly for transfer, these effects might be even stronger.

There are still many unresolved questions about the effect of drama on other kinds of outcomes. For example, how much drama, in what doses, is most effective? And does theater (rather than classroom enactment) lead to the same kinds of enhancement of verbal abilities?

It is my hope that this synthesis has provided fertile ground for new studies based on some of the more provocative findings. If we work to build on what we already know, conducting well-designed, carefully planned studies,

we can only increase our understanding of how drama can serve as a creative and effective tool for learning that extends beyond drama itself.

Endnotes

1. I borrow this term from Sara Smilanksy.
2. Dansky, J. L. (1980). Cognitive consequences of sociodramatic play and exploration training for economically disadvantaged preschoolers. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 21, 47–58.
3. Smilansky, S. (1968). *The Effects of Sociodramatic Play on Disadvantaged Preschool Children*, New York: Wiley.

COMMENTARY

Learning in Drama

JOHN SOMERS
University of Exeter, United Kingdom

I come to this conference as a cultural stranger, to the United States and its arts teaching in schools, and to the research paradigm used to review the research conducted. How I envy you some of the certainties you express in your research report: "A file drawer analysis revealed that 451,924 more studies averaging null effects would be needed to yield a non-significant Stouffer's *Z*." Many researchers would kill for that kind of certainty.

I am not going to comment directly on Ann's and Ellen's presentations, although my reflections on what they have said are implicit in what I say. I wish to communicate something about the nature of drama, the learning that takes place within it, and the need for more research of the drama-learning process. Most of my points are made in relation to drama and I leave you, where appropriate, to extrapolate to the other arts.

Ann mentioned that researchers in the arts do not talk to each other. This was so until recently in drama, and the situation continues to be fraught with difficulty as we do not possess a common lexicon. The term "Theater in Education," for example, has a different definition in many countries, including the United States and the United Kingdom. We will need to solve that problem soon. Nor do we share common practice, for the kinds of drama found in the American school system tends to spring from theater, whereas that in the United Kingdom, while now more easily embracing theater form, has as its source structured children's play. In this presentation I intend to identify some of the central concepts of drama and drama learning.

The Nature of Drama

Drama activity takes place in a range of contexts, moving across age and cultural function. In education it can range on a spectrum from children's structured play to straight theater. A teacher[1] may work at any point on this spectrum; the important issue is for them to know where they are working on that continuum at any one time. At its heart drama involves the development of models of the human struggle to make meaning of the world and our place in it. All arts would claim some commonality in this endeavor, but what is special about drama is that we explore and express our thoughts through the

dramatic medium. We do not write poems, paint pictures or write songs about it—we model it in drama.

The engineer who is designing a bridge makes a model, probably computer-generated, of the intended structure and subjects it to a variety of changeable variables—wind speed, thickness of steel, weight to be carried, for example. So in drama we create models of life and test them for their efficacy. We too are able to change the variables—"let's try that again where the mother is more understanding"; "I wonder what will happen if the family can't afford the medical treatment?" Drama could be seen as a laboratory in which we examine the human condition, and particularly the dilemmas through which humans pass. The fact that it is not "real" allows us to release ourselves into it without fear.

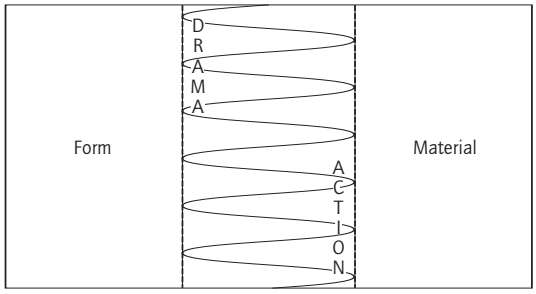
When working in this manner, students use performance as a way of communicating the distilled wisdom that has accrued from the drama investigation. As such they are embryonic playwrights, and they should be introduced to other playwrights in that way: "Let's see what Miller has to say about how men's dreams can be shattered"; "I wonder how Shakespeare saw Lear being treated by two ungrateful daughters"; "Beckett also wrote a play about the meaninglessness of life."

Form and Material

All drama activity is constituted of two components in dynamic relationship: form and the dramatic language and material, what the drama is about. These two elements meet in drama action, as shown in Figure 1, and it is this characteristic that is unique to learning in drama.

Although the material may be examined independently of the form and, to a certain extent, form can be experienced with little attention to material, the fusion of dramatic action renders them inseparable. In process, there

Figure 1



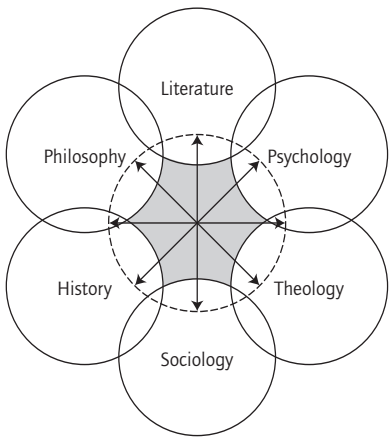
is an incompleteness that drives forward the exploration of the human condition within the medium. When participants have explored the dynamic relationships between specific types of form and material, they often want to capture what they have discovered to both clarify their own thinking and feeling, and to transmit the outcomes to others (classmates or theater audiences). When they have harnessed the meanings that have emerged to the best use of the medium, the dynamic energy is spent and it is time to move on.

It is relatively straightforward to plan a course dealing with drama form—the skills of the medium. What is more difficult (and for me, interesting) is the progression we adopt in relation to material. In choosing material to explore, I think the teacher or facilitator is applying an often intuitive child development model. She knows what kinds of material will challenge children of differing ages. Successful drama teachers constantly scan the work of children to feed their decision-making with the impulses and responses they see around them. My research also supports the notion that the teacher is also on a journey within their material, using the students as a means of exploring something that has significance for the teacher.[2]

The Project Zero research team has stressed at a number of points in its report that the arts must not be justified on the basis of their usefulness in transfer of learning to academic areas of the curriculum. I agree, of course, but we should not ignore the potential for drama to work in cooperation with other disciplines. If the form is unique to drama, the material we are exploring within it is often shared with other disciplines.

The characteristic that is unique to drama is the way in which we treat these topics—through the dramatic language. This is represented by the shaded area of Figure 2.

Figure 2



The fragmentation of learning that is at the heart of the school curriculum subverts "joined-up" learning. Acceptance of the mutual interest implied in my diagram encourages teachers to discuss with fellow teachers how topics within other areas of the curriculum could relate to their own work. In particular, drama teachers should discover how these topics might form the focus of work in drama. Where examination strictures allow, there is scope for interdisciplinary inquiry across the curriculum. Consideration of the Depression in the United States, for example, might include work in History (human and other forces at work), music (the songs of the period), Economics (analysis of the economic factors), English (stories and poems of the time), Geography (migration and settlement effects), Philosophy (the place of religion) and Art (examination of contemporary photographs)[3].

On the matter of transfer, so assiduously sought in the Project Zero research, I doubt in this case whether we would need to "teach for transfer." It is probably more effective to provide the conditions in which transfer can happen, a condition that Ann described as transfer "naturally designed into the curriculum."

Art's Educative Power

Art has been the forum in which humankind has expressed its questions, wonder and worship since time began. The humans who crawled through the earth in 18,000 BC to reach the Lascaux caves, there to paint pictures of huntsmen and their quarry were trying to make sense of the vagaries of their world. We are still engaged in this fundamental process, although it is sometimes difficult to recognize the more noble intentions of art in modern artistic output. Art does not provide a quick fix. It deals with the moral, the spiritual, the stuff of dreams, the metaphysical, and the ineffable. It deals in values and beliefs, transcending the mundane, characterized by the flood of one-star films that clog our media output. It explores what it is to be human. This morning on *NPR News* I learned how a Native American had been denied access to a sweat lodge before execution. This is the sort of event that sends the stone looping into the pool of drama as moral modeler.

Art questions our actions and motives and their moral contexts. An example is found in the film *Schindler's List*, when, taking time off from rounding up Polish Jews, a young SS man sits at a piano in an otherwise empty room and plays, exquisitely, a Bach prelude. It is the function of art to disturb, in the productive sense, to provide a counter story to the dominant story, to gnaw away at the foundations of the status quo.

The Ownership of Knowledge

The ownership of much curriculum material resides somewhere other than with the student. They know that last year's class did this topic, from that book, and that next year's class will do it too. The teacher is the conduit through which knowledge passes to the child (Figure 3). In this model, the teacher is gatekeeper. In drama there is a much more democratic, dynamic relationship between the teacher, the student, and knowledge (Figure 4).

The aim of the teacher of drama is to lead the student to an autonomous relationship with the art form, so that she may use and appreciate the medium in her wider life and the world beyond school. This changed function of the teacher may relate to the phenomenon identified in Ann's study that a teacher in role can, in some circumstances, enhance the learning transfer.

Narrative

We story the world to make meaning of it and our place in it. We do not experience directly most things we "know," we are told stories about them. The curriculum is chiefly an amalgam of stories that society has decided its

Figure 3

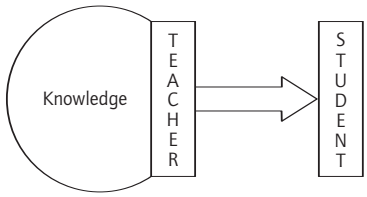
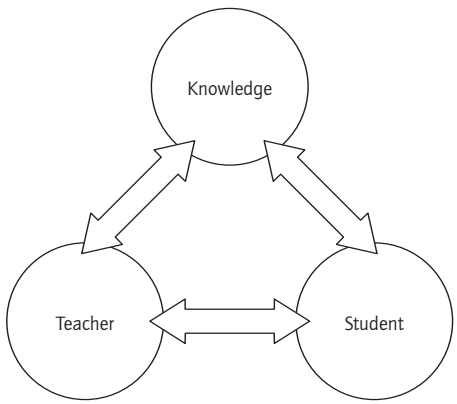


Figure 4



children should hear—and be tested on. Many of these stories concern giving children answers to questions they have not asked. As individuals, we know who we are by the story we tell of our past, and the story in which we predict a future. We are very selective about what we integrate into our story. Novitz maintains that we create our personal story much like a work of art.[4] Possessing a positive personal story is crucial to our self-esteem. We need only perceive the effect of Alzheimer's disease to understand the destructive power of story absence. Stories also give identity and meaning to societal groups—family, village, town, nation, for example.

An individual's personal story has potential for modification through intertextuality—its interaction with other stories. Much of education is of transitory interest, learned for examination purposes, failing to connect with our personal story-generating mechanisms. It is "out there," it passes us by. Significant stories do interact with our own, and good teachers strive to create conditions to make significant educational experience in which intertextuality can occur. It matters from where these modifying stories come. Stories, once the preserve of the oral tradition and carrying the wisdom of the society in which they were held, have now become commodities. They surround us; seventy-five percent of all television output is narrative of some form, mostly film. We have come to believe, been persuaded, that what we experience virtually is more important than the real. British people could be more concerned about the accidental death of Princess Diana, who most had never met, than the death through neglect of a neighbor, who they might see every day.

It is essential that schools support children in being producers of stories, counteracting the potentially overwhelming function of being story receivers. And we must ensure that they are making stories that matter, avoiding the trivial, the trite. The true function of art is the improvement of the human condition, not amusement.

The Role of the Teacher

Arts stories are meant to provoke, to disturb. The arts, and in particular, drama, constitute a debating chamber in which society can hammer out its views and values. Often we should feel uncomfortable in the presence of effective art, for it is not meant to stroke and console. To achieve this a teacher must prepare the context for effective arts teaching in the school. The students, parents, and other teachers should understand the impulses behind such work. It will require sustained and tireless effort by the teacher, but if she operates with integrity and purpose, the arts can become central in the life of the school and its students.

Teachers of the arts must engage in informed advocacy. They have most chance of succeeding when working with colleagues sympathetic to the arts' aims. It is particularly important, however, that all colleagues understand arts' rationale. Open evenings at which the students explain and demonstrate their use and knowledge of the arts can be an effective way of broadcasting intent and achievement, for the students are often our best advocates. In spite of institutional retrenchment in preparing drama teachers in the United Kingdom and widespread misunderstanding about what drama is for, drama has been among the fastest growing subjects at sixteen-year-old and eighteen-year-old examination levels. In my own university drama department, approximately twelve hundred applications are received for fifty undergraduate places. However, we still need better evidence of drama's effectiveness, and for that to be achieved I believe that more teachers must be engaged in research.[5]

University-based researchers must support teachers in their research efforts and teacher education should include research as a significant strand of theory and practice. To achieve this level of involvement and, importantly, to release the necessary resources, we will have to: (1) prove that the researching teacher is a better teacher; (2) establish research not as a selfish act, but as a gift to the education community; (3) show teachers that research is not "other," but closely related to their daily practice; (4) build a research community that embraces teachers; (5) properly reward researching teachers both academically and financially; and (6) find ways of allowing research outcomes to be spread more widely and deeply in the education community; among other arts teachers in the school; among all teachers in the school; in other schools; in the district; and nationally and internationally as appropriate.

We also need to widen the arts research community's understanding of appropriate research approaches, drawing from different, appropriate paradigms. During discussion about the nature of future research in the arts, a discussant stated, "we must have a beginning hypothesis." This scientific notion is limiting. Research can be conducted in many appropriate ways—and be hypothesis generating, for example.

When the research is completed, we need to ensure that the outcomes are accessible, both in terms of distribution and through the employment of elegant language. There is a tendency for some research authors to confuse worth with impenetrability of language. This will not do. We will also need to reconcile ourselves to the fact that the outcomes may not serve short-term aims, that we may not know the significance of inquiry for years to come.

Artists and arts educators are in the forefront of decisions about the nature and future of civilization. We have the choice in our cultures to embrace that which is base and selfish, or that which is noble and caring. We

make these choices in the face of the forces of globalization, where big companies increasingly set the agenda. I read today that one-third of the world's known species of flowering plants is threatened. Kew Gardens in London has been given $120 million to establish a seed bank to conserve these potential casualties.[6] Schools and communities are major cultural seed banks, containing the narratives and other representations that reflect the complexity and conundrums of what it is to be human. It is there that the students will have handed to them the stories of their heritage, and the skills with which to create new ones. Teaching and learning in the arts become crucial as cultural difference is smothered by cultural globalization.

Drama has always dealt with the eternal verities. It has echoed down the ages through Oedipus Rex, King Lear, Willy Loman, and, for me, still does through the devisings of my drama students. Arts teachers are engaged in a noble cause, and I am proud to join another community that takes these matters seriously, contributing its skills and knowledge in fresh attempts to understand the value of arts experience. The priorities will be decided in the months to come, but we could do worse than explore Elliot Eisner's conference suggestion that our three research priorities might be: (1) What do pupils do in the arts? (2) What do teachers of arts do? (3) What are the characteristics of the milieu in which this take place?

I would add that my priority in drama research is to look at three areas: (1) What is the relationship between the real and the fictional in drama? (2) How do we take children to a fictional world? (3) What benefits do we expect for the child in entering the fictional?

An extension of the Getty arts research initiative should involve original field work, working close to the essential arts impulses from which the meta-analysts engaged in this most recent enterprise necessarily distanced themselves. Whatever is researched, we need to make sure that we acknowledge the transformative power of the arts. For too long we have concentrated on simply measuring the shadow of the bird that is art. The test of our research will rest on our ability to raise our gaze from the two dimensional, gray shadow and perceive the bird in flight, with all its miracle of song and color.

Endnotes

1. I use the term "teacher" interchangeably with that of "facilitator" or "leader" to indicate that, although I locate my argument in this article in schools, the principles on which my theory rests are applicable in other contexts.
2. See Somers, J. (1999). How teachers choose what to do in drama lessons. In C. Miller and J. Saxton (eds.), *Drama and theatre in education: International conversations* 289-297. Victoria, BC: AERA Arts and Learning Special Interest Group and the International Drama in Education Research Institute.
3. For more on this, see Somers, J. (1994). Drama in the Curriculum, London: Cassell, 158–187.
4. Novitz, D. (1980). Art, narrative, and human nature. In Mitchell, W., *On Narrative*, Chicago: University of Chicago Press.
5. See Somers, J. (Summer 2000).Teachers and Research: The value of the reflective practitioner and the role of teacher educators. In *Drama Matters* 4, Columbus: Ohio State University, 51–61.
6. Radford, T. (2000). Save our seeds. In *Guardian* 2, 4.

V.

Embracing Babel: The Prospects of Instrumental Uses of the Arts for Education

DAVID PERKINS
Harvard University

The Babel Problem

The research presented at the conference *Beyond the Soundbite* at the Getty Center, Los Angeles, August 25–26, 2000, and in the special Fall/Winter 2000 issue of the *Journal of Aesthetic Education* examines with admirable depth and rigor, and largely discouraging results, the proposition that education of various sorts in the arts might enhance students' performance in other areas. Although the findings are not entirely negative, and although the limits of the analysis are carefully articulated by the authors, it is important to stand back from their findings and ask whether the game is essentially over. If it were, this would not necessarily be a bad thing. I applaud the authors' emphasis that education in the arts serves individuals and society richly and amply in itself and needs no instrumental justification.

As to the game being over, some would say that it had never really begun. The history of psychology over the past century does little to encourage looking to the arts for help with other educational problems. The best prospects for learning in the arts enhancing performance in other areas would come from a seamless concept of mind, everything intertwined with everything else, so that deep experience in one area would provoke resonances in others, the arts in their richness perhaps evoking unusually strong harmonics across the scale of human endeavor. Unfortunately, as with the biblical story of Babel, the mind's languages seem many, its tribes diverse and separated from one another. Although some psychologists argue that intelligence is a single factor, indexed by Spearman's *g* for general intelligence, numerous others have challenged this conception on many grounds (not that a single-

factor view would give much comfort to those hoping that artistic experiences might elevate intelligence, since those who hold the single-factor view generally also hold that intelligence does not respond much to efforts to expand it). Studies of transfer from one area of human endeavor to another, beginning with Thorndike's well-known investigations early in the twentieth century, have on the whole shown little transfer of learning from one to another performance unless the two have blatant common elements. Research into the nature of expertise and the cultures that form around particular disciplines and social activities has encouraged the viewpoint that knowledge is local, and sophistication about X leaves people generally naive about the Y next door, or at least the Y on the next block.

Paradoxically, it's easy to construct plausible stories about how most any X could enhance most any Y. In their meta-analyses of the literature, the authors always take pains to point out how the various kinds of learning in the arts addressed by various investigations might plausibly enhance learning beyond the arts. They thus credit the investigators with undertaking sensible enterprises with reasonable hopes. In summary, stories of plausible transfer are easy to devise, but considerable empirical research and contemporary theory, as well as the findings of the meta-analyses, suggest that the workable connections are sparse. In particular, the meta-analysis finds evidence that certain kinds of musical experience impact on spatial reasoning, and that the use of drama enhances aspects of reading and writing, but that many other potential payoffs investigated by various researchers come up short.

So, again, is the game essentially over? Well, not quite. Several questions deserve attention: (1) Why seek to enhance students' performance with indirect instruction at all, when direct instruction might do? (2) Why single out artistic experiences as promising platforms for indirect instruction? (3) When transfer fails, how can it succeed? That is, how might the apparent failures mask important prospects of transfer? (4) When transfer succeeds, how can it fail? That is, how can demonstrated transfer mask serious difficulties in capitalizing on artistic development to promote development in another area?

Why Indirect Instruction?

At first thought, the notion of indirect instruction is an odd one. If students do not show the understanding or skills desired in such areas as reading or mathematics, why not commit more time and depth to the teaching of reading or mathematics, especially in light of the Babel problem and the scant promise it attaches to indirect instruction? Indeed, more time and depth for direct instruction is certainly a good first-pass recommendation. However, that acknowledged, there remain some reasons to consider an indirect approach.

First of all, students commonly display shortfalls in several high-priority areas of learning, among them reading, particularly reading for understanding; effective writing; mathematical understanding and problem solving; and critical and creative thinking. Although one can easily say, "more and better direct instruction," all but the last of these receive considerable attention already—not always with the desired depth, but certainly without the desired results. It might be that instruction from another direction would prove more motivating, dodge some cognitive bottleneck, or simply provide added time and attention.

Second, students' difficulties include some conspicuous bottlenecks, for example mastery of proportional reasoning. Efforts to address such bottlenecks directly have often proved ineffective. Perhaps a smarter direct approach would do better, but certainly indirect approaches are worth considering.

Third, sometimes important areas of learning are also rather barren areas of learning as typically treated, dominated by rote knowledge and routine skills. Perhaps an indirect approach would enrich and energize students' experience in ways that would carry over to these curricular Saharas. In summary then, although indirect instruction is not in general the best of bets, it is a recourse worth considering in light of the magnitude and persistence of a range of educational problems.

Why the Arts as a Source?

If one plays the game of indirect instruction at all, the arts may offer some of the best ways to play it, a point that can be summed up as follows.

The Wellspring Hypothesis: In a number of ways, artistic experiences could serve as a wellspring of motivation, insight, and cognitive development superior to experiences in most other domains.

In *The Intelligent Eye* (The Getty Center for Education in the Arts, 1994), I introduced several trenchant reasons related to the visual arts, and they apply to some extent to other arts as well. I repeat them here.

- **Sensory anchoring.** Thinking, talking, and learning often gain support from a concrete physical focus—for instance, a work of art either in the original or a reproduction.
- **Instant access.** Physical presence of the work allows checking ideas by looking or listening more closely or from a different perspective.
- **Personal engagement.** Artists make works of art to draw and hold attention, helping to sustain reflection around them.
- **Dispositional atmosphere.** Works of art often create an atmosphere of heightened affect, an occasion of more involvement and commitment that might spill over to foster thinking and learning dispositions.

• **Wide-spectrum cognition.** Although different forms of art foreground different sensory channels and symbol systems (contrast music with poetry with painting for example), many works exercise multiple channels (the music of poetry, the words of song), many evoke channels other than those native to them (as when poetry or music evokes visual imagery) and all allow discussion, hypothesis formation, examination of evidence, and interpretive expression in other media and modalities.

• **Multiconnectedness.** Artists commonly fashion works of art to connect with a range of themes central to the human condition—social issues, philosophical conundrums, sour or soaring personal experiences, historical triumphs and tragedies—and various cultures, social contexts, historical periods, and so on.

Test this list against other areas of experience one might tap for indirect instruction—say the study of science or mathematics or history—and you will find that as a generalization they do not offer this list of advantages. One cannot, for example, hold an historical period before one's eyes. While some scientific phenomena can be quickly and conveniently reproduced for examination and reexamination, most cannot. Such points support the Wellspring Hypothesis: The arts do indeed appear to be special in their potential for indirect instruction.

When Transfer Fails, How Can It Succeed?

The meta-analysis reported in the special issue of the *Journal of Aesthetic Education* dashes cold water, if not dumps boiling oil, on the aspirations of arts educators to serve the rest of education. Much of the cold water/boiling oil is probably justified given the general Babel problem. Even so, some of the negative findings might mask important potentials. How could this happen? When transfer appears to fail, how can it succeed in improved circumstances?

Suppose we are seeking transfer from arts experiences A to some non-arts area X. A could be learning to play the xylophone, X the math of fractions, A immersion in the art of the impressionists, X reading for meaning, A learning a bit of modern dance, X expressive writing, or whatever. Suppose the track record so far is poor. Nonetheless, there are at least four matters that deserve close examination.

Depth of the arts learning. There is no particular reason to expect A to strengthen X if the experiences of A are brief and shallow. Given the scant time committed to learning in the arts, even in some special programs, it may often be the case that there is not much depth of A-learning, making the expectation of an X-bonus quite unreasonable.

Richer and arts-appropriate sense of X. Given the factors behind the Wellspring Hypothesis, experiences with A are less likely to contribute to rote memory and routine skills related to X than they are to engagement, understanding, and thoughtful exploration of X. If the measures of improvement in X are rather reductive, they may miss significant transfer.

Chosen to serve transfer. The six factors offered earlier in defense of the Wellspring Hypothesis also write a rough recipe for the kinds of A experiences most likely to transfer to diverse Xs. Some works of art evoke deep questions of the human condition more than others. Works of art that can be scanned quickly as they are discussed—paintings, poems, songs—sustain discourse better than extended works—novels, plays, operas (of course, sections of longer works can be used). Activities that occur around works of art can be relatively broader or narrower in their spectrum. In summary, although learners' artistic development in itself might be served in many ways, the instrumental use of the arts calls for selecting the works of art and shaping the ways they are approached to favor the prospects of transfer as indexed by the Wellspring factors.

Reflective abstraction and connection-making. If the sought-for transfer from A to X involves abstract concepts, principles, patterns of practice, strategies, attitudes, or the like, effective transfer calls for episodes of reflective abstraction and connection-making as part of the learning experience. For instance, if learners are encouraged to think about what's going on in A and how it might connect to X, they are far more likely to learn from A in the service of X. Often, the learning of A includes no such element.

When Transfer Succeeds, How Can It Fail?

The findings from the meta-analysis were not uniformly negative. One story of success concerned the positive impact of drama activities on story understanding, reading achievement, reading readiness, and writing. Another story of success involved domains superficially more remote: Brief listening to certain music yielded transient improvements in spatial-temporal reasoning, an aspect of intelligence (the "Mozart effect") and certain kinds of extended musical training yielded sustained improvements in spatial-temporal reasoning. During the conference, Gordon Shaw discussed evidence that musical training contributes to improvements in understanding of proportional reasoning. With some success apparently in hand, what could go wrong? What might be misleading about the apparently positive results? Here again, there are several answers.

Transfer that counts. Performance gains that disappear within minutes (as in the "Mozart effect") or that appear on psychometric instruments (as with extended musical training) tantalize and certainly invite further

inquiry. However, they do not in themselves guarantee improvements in anything that matters, for instance better learning of mathematics or architecture. Often, interventions that apparently increase an aspect of psychometric intelligence turn out only to have increased performance on certain test-like tasks in a way that does not translate into practical performance gains.

Better direct instruction instead. As noted earlier, direct instruction is the natural and usually superior rival to indirect instruction. If direct instruction is not doing the job, why not simply improve the direct instruction? For example, students' investment of time in music learning in experiments to improve mathematical understanding is substantial. It's hard to believe that the same time invested in a thoughtful program directly targeting mathematics would not yield superior results.

A better transfer treatment. If A enhances X, perhaps there is a different kind of indirect instruction, B, that would enhance X even more. Such a B might involve the arts or not. For example, if certain kinds of music learning enhance proportional reasoning in mathematics by boosting brain elements responsible for spatial-temporal reasoning, as some have proposed, perhaps direct attention to spatial intelligence through orienteering activities or Tangrams or other means would yield the same or greater effects much more efficiently.

What's really going on? If A enhances X, this should not excuse psychologists and educators from investigating the basis of the transfer. As noted earlier, it is relatively easy to construct alternative explanations about how A might transfer to X. Therefore, when A does indeed transfer to X, the fact does little to select among those alternative explanations or suggest others not yet considered. However, discovering the right explanation is important, because the real basis for the transfer may be quite different than it appears to be and thereby suggest other artistic or nonartistic approaches to improving X. The true basis of the transfer may in fact have nothing to do with the arts but with some seemingly incidental aspect of the A experience—for instance, involving the students in sustained analytical conversations. Were this the case, the students might as well talk analytically about whatever interests them the most—works of art, pennant races, political campaigns, personal relationships.

Have the arts been sacrificed? If A enhances X, it is important to ask whether A does so at the expense of superficial learning in the arts. For example, the authors refer with some alarm to dancing the shapes of letters as an artistic way of fixing them in mind. If this indeed worked, it would hardly represent the quintessential experience of modern dance. The concern here is not that superficial uses of the arts should necessarily be avoided, if they contribute to learning on other fronts. Rather, the point is that such uses should not be

taken as contributing genuinely to students' development in the arts, and therefore should not encroach on budgets or schedules dedicated to the arts.

How to Embrace Babel

The Babel problem introduced at the outset and the ramifications explored so far generate a vexing dilemma for investigators exploring how learning in the arts might enhance performance in other areas. It seems that nothing can be discovered with any certitude! When such investigations yield negative results, they might have missed positive ones, and when they yield positive results, they might be overlooking factors that would rob the positive results of their import.

Exasperation is the natural response. In a Calvin and Hobbes cartoon I shared at the conference, Calvin remarks that, because nothing is as clear and simple as it first appears, because ultimately knowledge is paralyzing, he, as a man of action, must simply carry on without too much thinking. Hobbes comments philosophically, "You're ignorant. But at least you act on it." In the Calvinist spirit of intolerance for the luxury of thought, perhaps investigators in this area would do better simply to act—to find what they find and move on, passing from negative results to other topics altogether or from positive results to expanded implementation of programs.

However, Calvinism is not the answer. The double dilemma of when transfer fails, it might succeed and when transfer succeeds, it might fail actually recommends a fairly straightforward set of practices designed to avoid the pitfalls. Here they are:

Design to maximize the potential for transfer, to avoid misleading negative findings. In particular:

1. Design interventions for strong learning in the arts.
2. In the non-arts area where improvement is sought, look for rich arts-appropriate outcomes, for instance understanding, not just improvements in rote knowledge and routine skills.
3. Shape the learning in the arts to serve transfer, for instance by employing works that can be present and readily accessed during discussion.
4. Include reflective abstraction and connection making as part of the learning experience, especially for far transfer.

Explore alternatives seriously, to avoid deceptively positive findings. In particular:

5. Be sure that you're measuring transfer to performances of authentic significance, not just to results on narrow tests that may not mean anything practical.

6. Investigate whether more or deeper direct instruction might serve the desired outcome better.
7. Consider whether other sources of transfer, not from the arts, might serve the desired outcome better.
8. Inquire into the basis of the transfer, recognizing that it's easy to tell alternative stories, but they need to be examined critically and tested, and that the true basis for the detected transfer might not have anything to do with the arts per se.
9. Evaluate whether learning in the arts has been sacrificed for the sake of an indirect effect, and consider whether the intervention risks being taken for adequate learning in the arts.

Let me close by expressing my own belief that the Wellspring Hypothesis is true and the potential for transfer from learning in the arts quite real. The largely negative findings from the current meta-analysis tell us not that the game is essentially over but something else very important—that the game is not very well played yet. It's played with considerable naiveté about what's likely to transfer to what and what conditions of learning are likely to support transfer. In contrast, my colleagues Shari Tishman and Tina Grotzer have designed and studied thoughtful interventions that offer reasonable prospects of transfer from the visual arts to other academic areas and that show promise of transfer.

In summary, learning in the arts needs to be designed with some finesse for effective indirect instruction, which generally it is not. There are high walls between different areas of learning. Expecting an arts experience, however rich, engaging, and energizing, to help out in areas beyond the arts is a little like bouncing stones off a trampoline and hoping that they will break down the walls around the castle next door. If we want to breach the walls between dominions, what we need is not trampolines but well-aimed catapults.

What Are the Most Promising Directions for Future Research in Arts Education?

JAMES S. CATTERALL
University of California, Los Angeles

This response essay derives from the final session of a truly unreplicable conference of well-known arts educators and arts-focused researchers from across the nation and Europe. Held at the Getty Center in August 2000, the meeting focused on the formal release of an intensive and long-term study of the academic effects of experiences involving the visual and performing arts.

The centerpiece of the discussion, obviously from the foregoing material in this proceedings publication, is now widely known as the REAP report. This refers to the Reviewing Education and the Arts Project, produced by seven scholars at Harvard University's Project Zero. Ellen Winner, Professor of Psychology at Boston College and Lois Hetland, then an advanced doctoral student at Harvard, served as principal investigators. The lead authors and five additional scholars each produced chapters for REAP. REAP has created much productive discussion and some controversy among its readers in the ensuing months. See my response to the first chapter of REAP above to get some flavor of this.

The final session of the August conference was a panel discussion chaired by Harvard Professor Howard Gardner, Co-Director of Project Zero. Expertly and wisely thinking that a final panel should somehow resonate with the entire proceedings, Dr. Gardner created the first of two questions for the panelists a mere hour before its inception.

Professor Gardner's first question emerged in two parts: (1) What do you believe to have been the most important contemporary research investment in the arts in recent years? (2) What do you believe to have been the greatest waste of money ever on research in the arts?

As the panel wrapped things up, Dr. Gardner posed a spontaneous second question that could serve as an apt capstone to the whole of the conference: What arts-related research would you propose with a grant of $500,000? With a grant of $10 million?

Of course the panelists now have had time to think about these questions; so my responses here will be more considered and clarified for the benefit of Proceedings readers.

On the most important investment in arts and arts education research to date: It is quite difficult to identify a single research effort as being paramount to any field. This is because knowledge in any domain, as scientists know, is most certainly the result of combinations of accumulated studies and professional knowledge amassed over considerable periods of time. No one study, not even a large-scale and influential effort, decides much of anything, although occasional visible studies can have the effect of gently knocking a field a bit off of its foundations. Typically, even blockbuster studies interact with existing work and add incrementally to a body of knowledge, in the final analysis. We of course should remind ourselves of a few exceptions, such as Albert Einstein's observations and conceptual creations that revolutionized the field of physics forever.

In the arts, if cornered to produce a singular answer, I would escape this trap by saying that the accumulated investments at Harvard University's Project Zero have been the single most important collective investment in arts and arts education ever done. Founded some thirty-five years ago by Nelson Goodman and shepherded though its adult life to date by Howard Gardner and David Perkins, Project Zero has been home to the exemplary thinkers and doers of arts education research. Project Zero is where the nation and world look first for studies and maps into many if not most areas touched by research in the arts.

As a clear example, with a fetus of an institute in tow a decade ago, we made an obligatory pilgrimage to Cambridge to seek wisdom, suggestions, and even collaboration on a funded but not clearly designed $10 million project. Who did we spend a helpful morning with? Dr. Steve Seidel, recently named Director of Project Zero. Ten years later what became the Galef Institute has been funded by more than thirty foundations, has worked in about twenty states, and is still going strong.

An unquestioned key to such success is the planting of strong roots. Project Zero helped us sow our ideas, and undoubtedly touched our designs. Project Zero is my answer to the first question.

As to monumental wastes of money, I choose to nominate no person or project. Projects, even huge ones, and even ones that do not find results

always teach us something. The hypotheses may have been unfounded—but the project can serve to help others see this and to shape their own thinking. Perhaps the design and methods of a grand research project amounted to unquestionable blunders. Yet—even the most egregious of errors help us all to avoid the same mistakes in the future. My heart and mind go out to others' misguided and wasted research. Such work points me toward more well-informed designs and better ways of going about my own research.

What Would I Do with $10 Million?

Since the panel was nearing the end of its allotted time and many conferees were weary or had planes to catch, I went for the $10 million question first. My response: Shouldering such a burden and opportunity, I would seek out the advice of the best minds in the world to help plan any course of action. My instincts would be to conceptualize and design ample-scale programs of research on the neuro-function analogs of skill development and experiences in the arts. Current approaches are psychometric in design; we have indicators of changes of thinking and ability effected by arts experiences. Are there observable changes in brain function that could help us understand the development of thinking skills and dispositions, and at the same time undergird the field at hand with some very hard science?

My rationale is that without question the first decade of our new millennium is going to be the decade of cognitive neuroscience where questions of human thinking and problem-solving abilities are concerned. The experts gathered would come largely from the subfields of thinking skill conceptualization and measurement on the one side, and the best researchers in the relatively new but exploding field of cognitive neuroscience on the other. Collaborating with University of California, Los Angeles, colleague Edythe London, one of the top neuroscientists in the world, we have sketched how such a program of research might be launched in a proposal for a planning grant.

My initial projection for how the money would be spent is significant support for a consortium of leading top-notch research projects in university settings across the nation or world—projects led generally by proven scholars. But I would not overlook newcomers with great ideas. The funds would support four or five research efforts attending to brain function and higher-order thinking. The funds could bring the consortium projects together periodically over a few years, to be sure they can benefit from each others' thinking.

This multi-project, geographically scattered, but integration-worthy research model was employed with great success by the Champions of Change program between 1995 and 1999. The funders supported seven highly

varied studies focused on links between the arts and human development. The projects ended up as reasonably exemplary models of approaching good questions related to the arts and human development, and the periodic gatherings supported by the funders made each project all the better. Credits for this design go heavily to Jane Polin, then of the GE Fund, and to Nick Rabkin of the MacArthur foundation and as well as to Dick Deasy and the Arts Education Partnership in Washington D.C. And of course to my Champions colleagues who taught me a great deal in the process, and who will remain indefinitely scholars and friends I can call on for guidance and ideas.

What Would I Do with a Half-Million Dollars?

To be quite honest, with $500,000 I would launch the Catterall-London neuro-imaging project sketched above. Sorry to be uncharitable and self-serving here. We would start with expert advisors and trial studies planned to help conceive a long-term program of research. The initial targets would be increased understanding of the effects of artistic and other experiences on cognitive processes and abilities. Our first experiment would be to explore the effects of significant learning in *music* on *spatial reasoning* ability. Our measures would include both traditional psychometric instruments and maps of durable changes in brain function.

A key rationale behind this type of first step is that it would be wise to look for neuro-function change in the place where we would suspect the strongest effects of the arts based on research to date. If we fail to find systematically differing patterns of brain function in such a proven ability-changing experiment, we should not expect such research to reveal important results in areas where fewer and less sizeable effects would be hypothesized. We would end up either with exciting news about actual brain function effects of the arts, or know that we need to search in other ways or in other places.

We hope also to take some risks in a parallel experiment in an under-tested domain where our intuited and professionally informed theoretical thinking could be tested. A candidate for this is research on learning in the visual arts and the transfer of any learning to other domains. Our instincts lead us toward roads such as the components of learning to "see" in the visual arts and suspected effects in other non-arts pursuits. When we stand with visual artists, our usual first observation is that visual artists see the world in ways that non-artists do not. This would be interesting and exciting to explore as a matter of artistic ability development and its relevance to other abilities, with observations supported by neuro-function research.

A Heroic Effort to Make a Practical Difference in Art Education

ELLIOT EISNER
Stanford University

Howard Gardner has challenged us to identify a research project that has made a difference, and another that has not. I want to identify a project that has strengths and weaknesses, one with which I have had a long-term affiliation, going back to 1984. The project that I speak of is the Getty Education Institute's work on Discipline-Based Art Education (DBAE). Is such a project a research project? The line between research and development is no longer as clear as it used to be. To the extent to which work of this kind not only generates benefits for students but illuminates the conditions that need to be addressed for that to happen, it makes a contribution to knowledge. The knowledge that it contributes is not of the laboratory type; it is not what one might call "hard knowledge," but is in the form of "soft generalizations."

To appreciate the significance of the Getty Education Institute's efforts in DBAE, one needs to understand something about the ideas that dominated the field of visual art education in the 1950s. The ideas that dominated the field at that time were built upon a conception of human development that saw development as occurring largely from the inside out, rather than from the outside in. Development in art was seen as a genetic unfolding of human capacity that is best served in a supportive environment with little or no instruction; the teacher was not to intervene in the delicate creative processes in which children and adolescents engaged. Instead, the teacher was to be a support, since art was not taught as much as it was caught.

With the 1960s and the developments in outer space, a race that the United States lost to the Soviet Union, there was considerable interest in this country in developing more substantive curricula, especially in the sciences and in mathematics. The idea was that the schools needed not only to focus on the sciences and mathematics, but also to do so in a way that was intellectually authentic. Jerome Bruner's very important book *The Process of Education* described the conceptual ideal that animated curriculum reform in those fields at that time.

Although the roots of DBAE can be found in the Pennsylvania State Conference of 1965 in a paper delivered by Manual Barkan of Ohio State University and in my own work at Stanford in 1967 and 1968 in the Kettering Project that I directed at Stanford, these efforts, one conceptual and the

other practical, had little impact on the field until the Getty Center for Education in the Arts (later the Getty Education Institute) came into the picture in 1984. DBAE provided a comprehensive view of art education that rested on the assumption that artistic development is not an automatic consequence of maturation and that teachers and curricula mattered. Furthermore, it promoted the idea, also finding its roots in the 1965 Penn State Conference, that making images was insufficient. What was needed was a curriculum that dealt not only with the creation of art, but which also enabled students to learn how to see and experience visual form, that addressed art in its cultural and historical context, and that eventually introduced them to some of the leading and perplexing ideas in the field of aesthetics and to consider long-standing philosophical questions such as: What is art? Is goodness in art objective? Can art provide knowledge?

What the Getty Education Institute did was provide the resources and other forms of practical support to help teachers and school administrators understand what, at that time, was a "new view" of what art education could be for the young. Advocacy conferences on a national level were held every other year or so, monographs were prepared by leading scholars in the field concerning artistic development, research projects were supported, and a strong network was created nationally to help schools embrace and pursue each of the four curricular components—production, criticism, art history, and aesthetics—in their school districts. It was, until it was terminated in 1998, the largest and most consistent and enduring source of foundation support that the field of art education has ever enjoyed. Federal projects come and go with political winds, but the Getty was a constant source of support and encouragement.

There are a couple of problems that the Getty Education Institute ran into in trying to assume a leadership role. First, some members of the field of art education resented its "intrusion" in their professional options, even though there was no legal mandate by the Getty or anybody else to do art education any one way. In addition, there was concern among some that the Getty DBAE program was not sufficiently multicultural, that it paid too little attention to women, and that it focused its efforts too heavily on criticism in art history and neglected the creative side of art education.

Although some of these complaints might have some grounding, many of the complainers, from my perspective, were motivated by a desire to find a place in the sun—and what better way to do it than for David to fight Goliath?

Despite the loudness of the criticism from a relatively small group of individuals, the Getty Education Institute did make some modifications in its

language. It also made modifications in the kinds of materials that it sponsored and made available to students. Multicultural materials became more available.

A second issue that caused a problem was the absence of curriculum materials. Discipline-Based Art Education is an extremely demanding approach to the teaching of art for any teacher, even one well-trained in art. It is especially daunting for elementary school teachers, who often have little or no background in art. Thus, the expectation that they deal with the creative side of art education, with its history and social situatedness, with the development of the critical capacities of students to see visual work, and in addition, deal, when appropriate, with aesthetic matters is quite a challenge. It is especially a challenge when teachers are expected to author their own curriculum.

The Getty Education Institute did not develop a body of curriculum materials to operationalize its conception of art education—and for good reason. First, the policy at the Education Institute was to encourage different forms of curricula rather than to promote or even suggest that there was one way to do art education, the Getty way. Thus, what was being advocated was a model, a conception, a set of values, a perspective, but it did not provide curriculum materials for teachers to use.

In retrospect, this was probably a mistake. What we could have done—and I say "we" since I feel that I was a part of it—was to sponsor curriculum development in various forms directed to the achievement of the values that DBAE embraced. We did not, and I think it made it very difficult for teachers to actually use the approach as well as we had hoped.

Nevertheless, the overall conception of DBAE is alive and well in the policy papers, the curriculum frameworks, and in our view of what needs attention in the arts in American schools. Virtually every state in the union has an approach to art education that has been influenced by DBAE, and this influence is due to the resources, the energy, and the vision of the Getty Center for Education in the Arts/Getty Education Institute as it functioned from 1984 to 1998.

I Can See Clearly Now: Possible Sequelae of the Reviewing Education and the Arts Project (REAP)

CONSTANCE BUMGARNER GEE
Vanderbilt University

The Reviewing Education and the Arts Project (REAP) discredited the hyperpublicized claim that the study of the arts results in higher academic achievement. Where "near transfer" (or "potentially casual relationships") was plausible, it was duly noted. Yet the designs and conclusions of the vast majority of studies analyzed by the REAP team were determined to be fundamentally flawed, primarily because of the lack of control groups and significant student self-selection biases. Two major recommendations emerged from those findings: (1) arts education policymaking and funding should not be based on the effect (or non-effect) of arts learning on other academic subjects; and (2) research should continue into if and how arts learning can transfer to other subjects.

The REAP recommendations seem counterintuitive from my perspective, but then, I am concerned primarily with the health, excellence, and efficacy of arts education. Research into how the mind works, and how and when transfer occurs, is intellectually fascinating and undoubtedly important for teaching and learning across the disciplines. If better quality, more accessible arts education is your professional priority, however, it is not difficult to think of numerous other areas related directly to the practice, assessment, content, and context of arts teaching and learning where research resources could be utilized for more specific and immediate results. REAP advises strongly against the use of research on transfer as a basis for arts education policymaking. It is my opinion that the arts education community needs to focus its research in areas where policymaking can and ought to be based. Furthermore, those who fund arts education research and those who advocate for a more central role for the arts in the school curriculum ought to focus their resources and efforts in areas that support—as the REAP report states—"the important and unique kinds of learning that arise from the study of the arts."

An equally important, although little-discussed, implication of REAP is the frequent distortion of legitimate research findings and the widespread promotion of and reliance on "advosearch" (i.e., advocacy-instigated research) by arts organizations and advocacy groups courting political support and influence, and seeking public and private funding opportunities. Discussion

of current "uses and abuses" of research findings specific to the field of arts education could have been a valuable part of the REAP-focused conference, *Beyond the Soundbite: What the Research Actually Shows About Arts Education and Academic Outcomes*. At one point during the conference, REAP Project Manager Lois Hetland remarked that we need to be more "vigilant" about the manner in which our own and our colleagues' research findings are interpreted and portrayed. One conferee whose research was reported on in *Champions of Change* noted that the authors of the studies included in the publication were not given final approval of the executive summary. Apparently, some of the authors were unhappy with the advocacy-style tone and tenor of the executive summary and press releases. John Harland, head of the northern office of the United Kingdom's National Foundation for Education Research, mentioned later that NFER-affiliated researchers would not have allowed an unsatisfactory or inaccurate executive summary or press release to be written and distributed by research sponsors. Harland added, "With all our sponsors, whether or not they have copyrights, we insist that the researchers have the ultimate right to interpret the data—otherwise our independence is compromised. The final responsibility for framing executive summaries and even press releases is ours." Serious consideration of ways and means for individual researchers and the arts education profession to better oversee the marketing of research findings by research sponsors would be a stride toward smarter uses of arts education funding, more responsible program development, and more credible advocacy efforts. The unfortunate fact that the media may contort even the most meticulously composed press release does not absolve the research community from its responsibility to do everything within its power to ensure as accurate a portrayal as possible of research findings and conclusions. The REAP findings are a sobering tonic for the spinning, blurry vision of advocacy-driven education agendas run amuck. We must not overlook the opportunity to discuss the ethical, political, and fiscal ramifications of those findings.

The invitational letter to serve as a commentator at the *Beyond the Soundbite* conference stated: "We are interested not only in your reactions to the REAP findings, but also in your views on how best to determine and study the sequelae of quality arts education." Although I assumed the word sequelae had something to do with sequel, I wanted to be certain so as to be able to respond most dutifully to my hosts' charge. Turning to my trusty *Random House Dictionary* for assistance, I was surprised to find *sequela* (the plural of which is *sequelae*) defined as "an abnormal condition resulting from a previous disease." *Stedman's Medical Dictionary* defines *sequela* in somewhat more neutral terms as "a condition following as a consequence of a disease." What

fun to be asked to contemplate the pathologically based conditions of *quality* arts education!

I realize some of you may be thinking (and perhaps rightly so): Don't be such a semantic fanatic, obviously in this case the word *sequelae* means "consequences," as in "how best to determine and study the consequences of quality arts education." While that may indeed be the proper interpretation, it makes the challenge less compelling because good arts educators can already identify the consequences of a quality arts education. Studying the consequences is only slightly more interesting for many of us; I personally would rather enhance and intensify the consequences, and of course spread them around as much as possible.

Several questions spring to mind as I return my attention to the first definition I encountered of *sequelae,* the definition offered by *Random House*. What are the *sequelae*—i.e., the conditions, the modes or state of being—of quality arts education? What was the disease that led to these sequelae? How might something of "quality" emerge from disease? It occurs to me that the duteous conferee might approach those questions from any number of positions and dispositions—biological, spiritual, material, political, metaphysical. For this initial and very brief (I promise) exploration into my hosts' requested inquiry topic, I choose to ponder the previous questions in a quasi-metaphysical-spiritual manner. After all, really good art often seems to dwell in those realms. Besides, art education master Elliot Eisner recently exhorted the art education community to return to "first principles"—to base teaching and learning on the basic logic of the field giving "pride of place to those unique contributions that only the arts make possible." (The concept of first principles is of course very metaphysical.) Additionally, who is to say what does and does not qualify as "spiritual"—which pretty much means I can say whatever I want since no one can actually "prove" me wrong.

What are the sequelae or conditions of quality arts education? To begin with, "quality" arts education may well be a sort of abnormality in that it is probably more the exception than the rule. The conditions of quality arts education are well-prepared, dedicated teachers; engaged students; and significant knowledge about and in-depth experience in one or more art form by students and teachers. Other conditions are well-funded school districts and well-educated, attentive parents. Occasionally, quality arts education sequelae are found in low-income areas among students whose parents are poorly educated and/or less than attentive, but more often than not, solid fiscal support of school arts programming and strong parental support of teachers are noticeable sequelae of quality arts education. I submit that the disease that led to the above conditions/sequelae was a *dis-ease* with powerlessness, ignorance,

and mediocrity. How might something of "quality" emerge from disease? Through the attainment of awareness, enlightenment, understanding, Sauls become Pauls when the scales fall from our eyes, allowing us to see with absolute clarity—to recognize the truth, the heart, the first principle in all that we encounter.

I can now return with exuberance to my hosts' request to share my perspective on how best to determine and study the sequelae of quality arts education. Pondering how best to prepare and support arts teachers, to engage students, and to determine the quality and type of learning that has occurred is a very worthwhile use of an arts educator's time. We can best study those sequelae by refusing to allow the opacity of hype and confusion to dull our vision; by keeping faith in the importance of teaching the next generation about music, theater, dance, and visual arts; and by steadily and honestly working toward the development of more effective ways to help others experience what is unique, useful, and enticing about various art forms and works of art. It is my sincere hope that the sequelae of REAP will be the utter discouragement of the justification of arts education on the basis of improved skills in other subject areas. Arts educators have made enormous progress in curricula development over the past two decades; our thinking about the content and purposes of arts education has expanded and been refined. It is essential for arts education professionals to continue to analyze, evaluate, and build upon their accomplishments in curricula development. It is important that we participate actively and forcefully in policy development affecting our research agendas, and the content and practice of pre-kindergarten-through-university arts education.

Recently, it seems arts educators have taken several steps backward as we struggle with the most fundamental of questions—"Why?" Why is it that we do what we do? Why should anyone else care? Although it is important to ask such questions from time to time, it is equally important to remember that the answers lie within our own hearts and minds. They are not to be found in the "talking points" of an advocacy campaign or in a collection of SAT scores.

I would like to make one additional observation about the *Beyond the Soundbite* conference. It was enormously instructive to be surrounded by people "outside" the arts education field discussing the purpose, consequences, and future research directions of arts education. Hearing physicist Gordon Shaw proclaim that his main, almost sole concern is improving math skills because "math is key" in enabling "kids, especially low-SES kids, to get high-tech jobs" was a real eye-opener for someone who can't imagine wanting a high-tech job. I think all of us could benefit from more frequent out-of-field experiences. It makes you think, well, more metaphysically.

A Personal and Non-Academic Report on the Getty Center Reviewing Education and the Arts Project (REAP) Conference

PATTERSON SIMS
Museum of Modern Art, New York

Initial Concerns

My delight in being included in the Reviewing Education and the Arts Project (REAP) conference was mixed with anxiety. This non-academic is never sure if he should attend academic conferences. The stated topic itself and those I knew who were organizing and planning to attend the conference conferred a sense that I had a rightful place there, but I still found my position unsettling amid a congregation of research-based academics. Los Angeles itself presents the same love/hate, need/repulsion dichotomy as academia does for me, so I knew it would be a few intense days that would pass in summits and depths of mind and emotion. The lure of the Getty's famously efficient and elegant way of conducting such gatherings made me know that my periodic intellectual discomfort was going to be utterly balanced with constant creature comfort.

The articles and documentation that was sent in advance of the conference were formidably rigorous and made their points with meticulous clarity. For the visual arts, the area of arts education that I am professionally attached to, it was depressingly clear that academic outcomes cannot be proven. This was a message that I did not want to see exhaustively confirmed. Knowing that in the areas of drama and music, positive correlations were somewhat more provable filled me with some envy; I experienced the swiftly fleeting conviction that perhaps I had allied myself within the wrong professional field. I reminded myself that my choice of the visual arts and contemporary visual art had been almost a calling, a quasi-religious combination of obsessive need and unalloyed pleasure. It stemmed from childhood learning difficulties that made me unable to read until fourth grade. Time, emotional growth, and the intervention of a savvy teacher, who bribed me with clusters of the beloved postcards that I had been diligently collecting, brought me to reading and the start of a semblance of academic proficiency. The combination of the postcards' quaint images of summer holidays and the short and sweet banality of their messages made letters become words, connected thoughts, and gave me reading's gift of being ensconced within other minds.

From the start, it was clear that Ellen Winner had done her homework. She and Lois Hetland had set themselves a Herculean task (so that was why we were put in the Getty's Herculaneum Conference Room for all our meetings). Their use of meta-analysis and Robert Rosenthal's patient and enthralling description of this tool the first morning helped me enter a mirrored kingdom of data and mathematical formulae. The methodology made my contrasting professional milieu of inquiry-based multiple narratives and interpretations seem trivial, amorphous, and weightless. What also became apparent was that for many of the speakers and respondents (and perhaps this justifies the personal approach of this summary), to "respond" meant to repeatedly tell the group about their own latest research and findings, no matter how disconnected to our stated conference subject they might be.

The Museum of Modern Art and Visual Thinking Curriculum

As the proceedings advanced, I began to experience regret that this research had been done. It might, I fear, rationalize the exclusion of visual art from kindergarten through twelve school life. I knew that two researchers from Harvard's' Project Zero, one of them, Shari Tishman, present at the conference, had found in an assessment commissioned by the Museum of Modern Art some evidence that the museum's inquiry-based Visual Thinking Curriculum (VTC) had positive results for cognition and verbal skills. VTC's sequenced, observation-based set of open-ended questions and, in the last five years, limited amounts of hard information and insight has thrived at MoMA in the last decade. Its essence was that careful attention to key icons of modern art from the museum's own definitive holding (and in a recent project in Caracas, Venezuela, with selected works from a semiprivate collection and a public one in Venezuela) raised students' skills of observation, analysis, and articulation. In the aftermath of the Project Zero group's VTC assessment, VTC is undergoing revisions that will take the students' and teachers' needs and feedback into much greater account. A carefully reformulated and edited version of our printed materials and teacher resources for collections and exhibitions will also provide the platform for professional development opportunities for teachers and an ongoing forum for feedback and communication.

The Museum has become even more committed to the ways that VTC's skill-based components can spill over into other areas of the students' curriculum. In its essential form, VTC is delivered to fourth- and fifth-grade students and teachers by full-time and freelance museum educators and, more broadly, to other grades in single or multipart sessions in the schools, at the museum, and in the last year, through video conferencing. It is vital that MoMA's education programs expand from a local to a global public.

Additional Observations

The argument resonated at the conference that school athletics does not add to the ability of students to score higher on standardized tests, and yet few protest its inclusion or validity for American schools. The need to justify results struck many of us as shortsighted and distinctly short-term. The other issue that was raised on the second day of the conference was the realization that, in many cases, very little was known about the underpinnings of the myriad of research papers and reports from which the otherwise exhaustive REAP analysis was done. The REAP project seemed to be very scrupulous about not looking at the "research" done by advocacy groups, feeling that positive results emanate from positive criteria. The reports and studies were perforce analyzed as data, yet their net was cast so very wide that it appeared that it could not capture any "inside" or definitive information about the way the original research projects had been set up. As MoMA Education learned, as the Project Zero assessment was started, the terms and conditions of research were critical: what you prove depends on control groups, the information you are given, and precisely what you set out to prove.

A Practical Suggestion for the Getty

In the final forum on "Future Research Directions: Building on REAP," ever the non-academic, I made a practical proposal, and I have since expanded it somewhat. These suggestions are based on my understanding that Getty Education is at a crossroads. My previous awareness of the Getty's extensive arts education program was that it was reformative and specific, so what I have suggested here is very practical and basic.

My proposal was not based on the prior, extensive, and "discipline-based" Getty Education model as I imperfectly understand it, but a program that would be pursued by the Getty Museum for its own objectives and service to an immediate and regional community. The proposal would have significant implications for the Getty Center and its public identity and is conceived as a social service to the city of Los Angeles and its immediate metropolitan area. My suggestion is not content-based, but delivery-centered.

The proposal is based upon the Getty Museum acquiring a major downtown Los Angeles building and handsomely converting it as a center for professional development for teachers, for student and family art understanding and art-making classes, and a small (approximately six thousand square foot) art exhibition center. The building should be in Los Angeles's urban core near the two Museum of Contemporary Art (MoCA) facilities, the Disney Symphony Hall, and their ten-block environs. It should not be a new structure, but a rehabilitated one, preferably a former light industrial or office

building of some architectural distinction that might not otherwise have been saved. It would contain numerous flexible, simple classrooms—open box, fully wired spaces for activities that reflect the changing needs for the appreciation, understanding, and hands-on attractions of art. It should have lots of adjacent, free parking space. This facility would be connected electronically by Internet and video-conferencing and also in other lively and on-going ways, including bus service, with the Getty Museum and the Center campus as well as the soon-to-be reopened Getty Villa Museum in Malibu.

This new Getty Education facility would be aimed at teachers, their students, and the families and adult companions of those students. The exhibition space would show works from the Getty collections in ways aimed to make connections with teachers, students, and families and build upon the increasingly effective dedicated learning spaces and exhibitions at the Getty Museum. The downtown facility would be set up in collaboration with Los Angeles and surrounding school districts. Proactive and not immediately proscriptive, it would make available numerous special hands-on art-making and appreciation classes for students and families and courses of instruction for teachers, from daylong sessions to summer teacher institutes. From its full round of activities and interactions, numerous research and study projects might be launched, where a variety of learning models can be tested using visual art at their core.

The educational content of all of these activities would be developed by the professional education staff, the Los Angeles schools' administrative hierarchy, the teachers and students themselves, and outside consultants as decided by the core Los Angeles city and Getty Museum program collaborators. These programs would enhance the appreciation of the Getty's art collection, bring the Getty to a central and easily accessible location, and directly ally the Getty's artistic resources to the overall learning and curriculum needs of the area's teachers and students, as well as their families.

Following the Reviewing Education and the Arts Project (REAP)

SAMUEL HOPE
National Office for Arts Accreditation

At present, arts education is enjoying significant public attention. A phenomenon of this kind has many causes. Several powerful generators are: (1) the formulation and successful promulgation of national standards for kindergarten through twelve dance, music, theater, and visual arts education in the context of general education reform; (2) a return to the arts, almost in desperation, in situations where schools are in deep trouble and other improvement efforts have failed; (3) a definition of arts education extended to such breadth that the term or a putative synonym is used promotionally by an ever-increasing number of organizations and institutions; and (4) publicity about research connecting or correlating study or experience of an arts discipline with brain and mind development, especially in children.

The Reviewing Education and the Arts Project (REAP) focuses primarily on the last of these. It uses the specific technique of meta-analysis to take a specific type of snapshot of a specific body of work. The picture REAP provides is comprehensive. By summarizing and correlating previous studies, it offers a view that has never been seen before.

This picture does not show us everything. It does not contain all that is known and understood. It is not taken from all possible perspectives. It cannot be interpreted or extrapolated to tell the whole story. But it does document comprehensively what is known from a certain body of research while avoiding a connection with the desperate advocacy of arts justification. It warns against coercive illusions rather than trying to produce them. This achievement renders the findings dynamic and useful in perpetuity.

Given the characteristics and achievements of the REAP project, two kinds of continuing efforts seem obvious: research itself and the implications of findings for policy. The project demonstrates many knowledge gaps that need to be filled. For example, the REAP analysis tells us virtually nothing about the content and quality of the arts instruction or experience that produced or did not produce certain results. REAP could not have provided this information because the studies it reviewed did not. Thus, present REAP findings give us only the barest glimmerings of what might make a difference. Longitudinal studies, experiments that change variables, and comparisons among levels of capability in artistic and other areas seem to be real possibilities for future

studies. The REAP results indicate that a new level of sophistication in choices of research topics and in project design could fill present gaps and open new productive research questions.

The REAP results also indicate the importance of pursuing how study of the arts develops the artistic mind. Issues of transfer are extremely interesting, but one route to understanding more about transfer may be through understanding more about how the arts teach what we know they teach.

Arts content, then, seems to be an overarching theme for developing next steps. Research is starting to show what creators and performers already know as they bring intellectual technique to their craft: content makes an enormous difference. Content includes narrative or other features that can be put into words—this play is about love in a context of political conflict; the dancers defied gravity. But perhaps more important for psychological studies, it also includes the kind of content expressible only in the art form itself. Whether primitive choices—green because it is the most relaxing color, for example—or deep structural considerations—how much repetition of what types in the first thirty seconds of this string quartet to establish a basis for the listener to follow the rest of a four-minute movement—such formulations of non-verbal content make art as a mode of thought and works of art what they are. Slight changes of nuance or pace can produce large perceptual shifts. Anyone attending a professional orchestra rehearsal can hear this for themselves immediately. The same principle applies in all the arts disciplines. Research designs that avoid such content variables may find an increasing number of connections among brain, mind, the arts disciplines, and other disciplines, but they are unlikely to find what makes the connections function or how they work. If the REAP project can foster a turn toward content variables, it will have made a major contribution.

The REAP study connects its findings to policy issues. It delivers a stern warning about claiming more powers for the arts as a whole or for the arts disciplines than the evidence shows. Who will hear this warning made in an advocacy culture that lives by total positives? The REAP project takes a courageous and wise position. Neither the arts nor arts education are fundamentally jeopardized by REAP's findings themselves or by its cautionary message. Survival for the arts is not the issue, but rather the extent to which the highest achievements in them are accessible; not whether rigorous arts instruction will be delivered as a part of basic education, but to whom and in what numbers. The connection between psychological studies in the arts and educational policy is important. Unfortunately, the main reason seems to be that today, many people do not understand or accept the importance of arts study on any other terms. This means extra care with information, developed from all kinds of

research lest, when the years conquer the hours, the arts are left in a worse public values position than they were before.

Psychologists continue to make important contributions to arts education. Today, an increasing number are devoting themselves to arts-related studies. REAP demonstrates the importance and power of looking comprehensively and methodically over a wide body of psychological research and drawing summary conclusions. It also demonstrates the need for those who use these findings to take care lest they become ultimately destructive to the larger goal of full access to arts education. The REAP study gives psychologists, teachers of the arts, and decision-makers new grounds for connecting research, teaching, and yes, advocacy in newly productive ways. Although psychological studies are only a part of the whole, they are critically important. Continuing and intensifying such work is a critical element of future progress for teaching and learning, and thus for the arts more broadly. Following the REAP project with new specific and comprehensive studies of equal rigor and honesty can make a significant difference, especially if pursued in concert with those working in and for arts education programs that seek to be as serious about arts learning as REAP is about psychological research, and as forthright about what they accomplish as REAP is about its findings.

VII.

Research in Arts Education: Directions for the Future

ELLEN WINNER AND LOIS HETLAND

REAP uncovered mixed support for the claim that the arts boost academic achievement. We were able to document three areas with clear causal links between the arts and non-arts, cognitive outcomes, two areas with weak support for a causal link, and five areas in which no reliable causal link has yet been demonstrated. While our findings did not reveal a great deal of support for the claim that arts study transfers to non-arts outcomes, it is certainly possible that future researchers will discover more ways in which studying the arts transfers to some valued non-arts cognitive ability.

In addition, there is still existing research that remains to be synthesized. When we began this project, we speculated that there might be but a handful of studies to sift through, and we wondered whether two of us were not one too many for this project. But we soon learned that there were hundreds and hundreds of studies examining the relationship between some form of arts study and some form of academic outcome. And before we knew it, we were making decisions about what we had to rule out because we had only three years of funding. Ultimately we decided to focus on studies that examined the results reported here. But clearly, there is work we could not address. Because we carried out meta-analyses, we could only include studies with quantifiable outcomes. Hence, for example, we could not include qualitative case studies, however valuable and informative such studies might be. We also decided that we could not study social and motivational outcomes of arts education. Nor could we study all cognitive outcomes. For example, there are studies of the role of music in foreign language learning, as well as the role of music study in affecting creative thinking, that have not been reviewed by REAP.

Thus, our work should not be used to conclude that researchers should stop looking for transfer from the arts to non-arts academic areas. We believe that there is much work to be done about the question of transfer from arts to non-arts domains. Our quarrel is only with using transfer as the primary *justification* for the arts. There is merit in examining the possibility of transfer, both because when transfer is possible, teachers in all disciplines can exploit it, and because a demonstration of transfer helps us understand more about the mind.

An Agenda for Research

Unfortunately, all too often researchers, practitioners, and advocates simply assert all the wonderful things that the arts can do—from engendering perseverance to training critical judgment. Arts may or may not accomplish such wonderful things. It is the researcher's job to find out both whether and, if so, how. We believe that links between arts and non-arts outcomes are most likely to be demonstrated when there is an explicit theoretical argument and psychological mechanism that relates an arts skill to another valued ability (as there is, for example, for the link between drama and verbal ability, since enacting a verbal text may well cause one to process that verbal text more deeply than if it is just read). While it is true that such theories are relatively easy to create (see Perkins, this volume), they nevertheless tend to sharpen both the program's intent and execution and the way it is assessed. If future well-designed studies actually demonstrate a robust and meaningful dividend from excellent arts instruction or practice, we would be gratified.

Many of the studies we found in the REAP searches were weak and not based on a theoretical reason about why transfer might occur. So for example many studies simply stated that their aim was to test whether an arts integrated curriculum boosted test scores. What was typically lacking was a theory about why such an effect might occur, whether as a result of heightened student engagement in school or actual transfer of cognitive skill. Any study testing for transfer needs first to have a theory, and then must test whether the requisite skill that may transfer is actually learned in the arts class. Only then does it make sense to check for whether students begin to use this new skill in another area. We give an example of three possible theories and how they might be tested.

The first theory has to do with confidence and engagement. Perhaps the arts enhance academic achievement, but only for a certain type of student, and only when the arts are used as engaging entry points into academic subjects (e.g., role-playing in history courses; analysis of rhythms in a proportions unit in mathematics).[1] A particular kind of student might benefit particularly from this kind of instruction—the student who is academically at risk but who

has strengths in an art form. If such students experience success in the art form linked to the academic subject, they may then believe that they can succeed in the academic subject, and their willingness to stay with the subject may grow. Increased confidence should lead to increased motivation, effort, and long-term engagement, which in turn should result in deeper understanding and higher achievement. Few experimental studies thus far have even begun to test this hypothesis. To do so comprehensively and rigorously would require comparisons of academically strong versus academically at risk students taught the same subject matter with and without the arts as entry points. Can we identify students who first experience success in the art form and subsequently go on to show heightened interest and effort in the academic subject matter? And do levels of interest and/or motivation predict later achievement in that subject matter?

The next two theories are what several of our colleagues at Project Zero call "dispositional," by which we mean that students develop habits of mind and behavior that are motivated both cognitively and emotionally.[2] In the first case, students may learn through long-term projects in arts classes to persevere and work hard on revising their work toward higher quality. Such a mechanism for transfer relates to an incremental, mastery-oriented view of learning[3] in which students become more inner-directed, self-managing, and self-invested. This kind of perseverance and insistence on high standards may then show up in history class. Transfer may be further enhanced when the strong arts instruction is bolstered by making deliberate and concerted efforts to help students identify, understand, and connect the approaches they use in arts classes to the ways they approach other subject matters. To determine whether such working habits do transfer, one must first make sure that students enrolled in serious arts classes do indeed develop perseverance and appreciation for high standards. Only then does it make sense to check whether their perseverance in history class begins to grow stronger. In the studies assessing whether arts learning boosts general academic achievement, we found no research that first demonstrated learning of the requisite skill or attitude in the arts class itself and then assessed transfer to other subject areas.

A third possible theory is also dispositional. Students in a strong arts program may learn a "disposition" to be reflective about their work and to step back and make critical judgments about what they are doing.[4] This critical judgment faculty may then be used in history class, showing up in the form of considering more nuanced interpretations based on multiple points of view, probing the ambiguity of various interpretations for common assumptions and themes, or finding novel problems to explore by interpreting complex chains of evidence about historical causes and effects. To demonstrate this

possible "transfer," the same logic applies. The researcher must first find evidence that students are in fact developing critical judgment skills in an arts class. The next step is to see, through sensitive assessment measures that go beyond factual memory, whether students begin to show more critical judgment in history class after having been in such an arts class.

Researchers should also be aware of a possible indirect link between strong arts programs and strong academic achievement. It is conceivable that schools that treat the arts seriously institute other kinds of innovations that are favorable to academic learning. For instance, these schools may become more inquiry-oriented, more project-based, more demanding of high standards, and more focused on processes that lead to excellence. Teachers who are drawn to schools that take arts seriously may, on average, be more committed, more knowledgeable, or more skilled in actively engaging students in serious, ongoing inquiry. Educators and policy makers need to understand what comes along with the arts.

To discover factors that contribute to academic achievement when arts are emphasized, researchers need to carry out ethnographic studies of exemplary schools that grant the arts a serious role in the curriculum. What kinds of innovations have been made in these schools to foster excellence? If certain innovations are always found in schools that grant the arts a serious role, this finding could account for why schools with serious arts programs have high academic performance. It is important that we not confound innovative teaching with arts programs. When we find schools that have both and that report high academic achievement, we should not simply assume that the arts are playing a causal role. Rather, we need to define and articulate the pathways from instruction in arts through various potential moderating factors to higher achievement. Only experimental study where variables are disentangled through rigorous design and analysis can allow us to draw such conclusions.

Beyond Research: A Rationale for Arts in Education

While an exploration of transfer may reveal interesting relationships between artistic and other forms of learning, the primary rationale for the arts in education should not rest on transfer. It is our deeply felt conviction that we need to distinguish core from bonus justifications for teaching the arts. As mentioned above in Winner (this volume), core justifications are the central reasons—learning in the disciplines of the arts themselves; bonus reasons are the side effects—enhanced learning outside of the arts, which may or may not occur. We believe that the two most important core reasons for studying the arts are to enable children to appreciate some of the greatest of human cre-

ations, and to enable children to have enough skill in an art form that they can actually express themselves in this art form. The arts are the only disciplines in which recognizing and expressing deep personal feelings are at the heart of the enterprise. The goal of education should be to help our children learn to appreciate the greatest of human creations, and these include the arts no less than the sciences and humanities.

When arts education is justified on the basis of bonus effects, the arts are put in a vulnerable position, for three reasons: (1) some of the bonus claims are not well supported by the evidence; (2) if the only valued goal of arts education is academic enhancement, why not just teach academic skills directly, rather than indirectly through the arts?; and (3) allowing bonus reasons to drive arts education may lead arts educators to stray from the heart of the arts and teach the arts in inauthentic ways simply to boost test scores. We have seen researchers turn strings of music notations into math problems and show that this kind of music education helps math skills. While this may help math skills, it is not music education. And we have read of dance programs in which children learn to put their bodies into the shapes of letters. While this helps children learn letters, it should not be confused with serious dance education.

The arts are not just handmaidens to other disciplines (though they may be vehicles that enhance learning in other subjects). Nor are they merely modes of self-expression (though they are that), nor just entertainment (though they are that as well). Instead, the arts are serious modes of inquiry and discovery in their own rights, and they are based on symbol systems no less complex, and no more transparent, than the symbol systems of mathematics and language. This view was developed in detail by the philosopher Nelson Goodman, founder of our research center, Harvard Project Zero.

Of course, we do not know for sure what is the best education for children to ensure that they will grow up to lead productive and happy lives, but let's bet on history. The arts have been around longer than the sciences; cultures are judged on the basis of their arts; and most cultures and most historical eras have not doubted the importance of studying the arts. Let's assume, then, that the arts should be a part of every child's education and treat the arts as seriously as we treat mathematics or reading or history or biology. Let's remember why societies have always included the arts in every child's education. The reason is simple. The arts are a fundamentally important part of culture, and an education without them is an impoverished education leading to an impoverished society. Studying the arts should not have to be justified in terms of anything else. The arts are as important as the sciences: they are time-honored ways of learning, knowing, and expressing.

Endnotes

1. Gardner, H. (1999). *The Disciplined Mind: What All Students Should Understand*. New York: Simon and Schuster.
2. Perkins, D. N., Jay, E., and Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly* 39 (1), 1–21.
3. Dweck, C. S. and Elliott, E. S. (1983). Achievement motivation. In Mussen, P. H. (ed.). *Handbook of Child Psychology* (4th ed.), 14, 643–692.
4. Perkins, D. N. (1994). *The Intelligent Eye: Learning to Think by Looking At Art*. Occasional Papers 4. The Getty Center for Education in the Arts, Santa Monica, CA: The J. Paul Getty Trust.