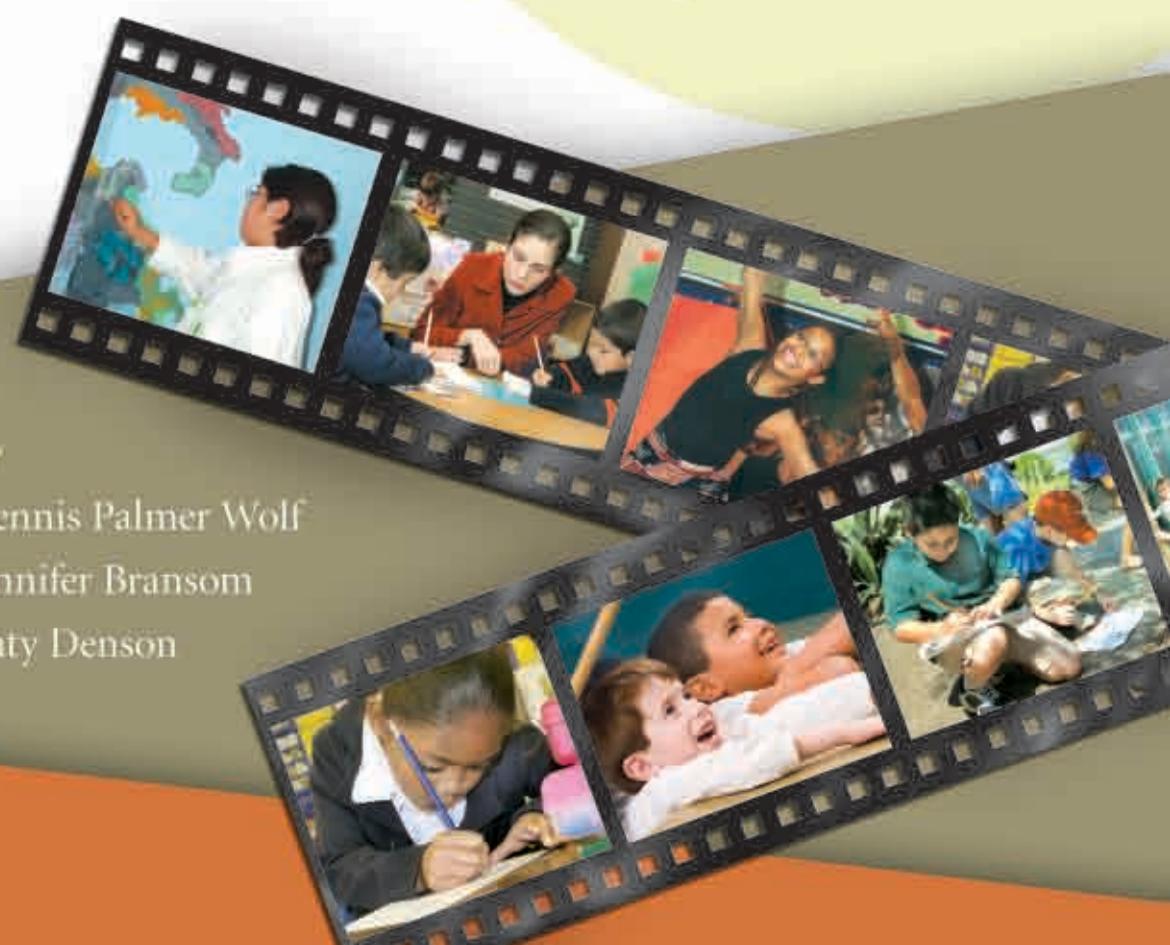




More than Measuring:

Program Evaluation
as an Opportunity
to Build the Capacity
of Communities



By
Dennis Palmer Wolf
Jennifer Bransom
Katy Denson

More than
Measuring

More than Measuring:

Program Evaluation
as an Opportunity
to Build the Capacity
of Communities

By

Dennis Palmer Wolf

Jennifer Bransom

Katy Denson

Big Thought

Dallas, TX

Research for this book and publication were supported by _____.
This book is also available online at _____.
Additional copies are available (\$15.00 postpaid) from _____.

© 2007 Big Thought

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without permission in writing from the publisher, except by a reviewer, who may quote brief passages in a review.

ISBN: 0-9769704-1-4.

COPY EDITOR: Patsy Fortney

PRODUCTION: Amy Gerald

COVER AND LAYOUT DESIGN: Wiz Bang Design, Inc.

First Printing

ABOUT THE AUTHORS

DENNIS PALMER WOLF is the director of the Opportunity and Accountability Initiative at the Annenberg Institute for School Reform at Brown University. Wolf heads a team that examines excellence and equity issues related to learning opportunities and outcomes for students in K–12 urban systems. Prior to coming to the Institute, she taught at the Harvard Graduate School of Education and codirected the Harvard Institute for School Leadership. She received an EdD and EdM from Harvard University. Her areas of interest include standards, assessment, school reform and cultural policy for youth. She has served three terms as a member of the National Assessment Governing Board.

JENNIFER BRANSON is the director of accountability at Big Thought. In collaboration with lead researcher, Dr. Dennis Palmer Wolf, Bransom directed a four-year longitudinal study assessing how teaching and learning change when arts and cultural programming is integrated into mandated curricula. Most recently, she collaborated with a team of national consultants and supervised 23 researchers to conduct policy leader interviews and a citywide census of art learning opportunities in Dallas. Bransom received an MA in arts management with an emphasis in education and public relations from American University.

KATY DENSON is recently retired after 13 years with the Dallas Independent School District (Dallas ISD). While working in the Department of Evaluation and Accountability, she evaluated reading programs and Title I. Denson is an adjunct professor at Texas A&M University—Commerce. Denson received a PhD from the University of North Texas. Her areas of interest include statistics, multiple regression, experimental design, structural equation modeling, measurement, and multivariate statistics.

Table of Contents

ABOUT THE AUTHORS	X
ACKNOWLEDGMENTS	X
FOREWORD — <i>by Cyrus E. Driver</i>	X
Introduction	X
Part I EVALUATING FOR SPECIFIC COMMUNITIES AND PROGRAMS	X
DESIGN PRINCIPLE 1: TAILOR THE EVALUATION TO THE CONTEXT.	X
– THE CONTEXT OF GROWING ACCOUNTABILITY	X
– THE CONTEXT OF ARTS EDUCATION	X
– DALLAS AS A CONTEXT FOR EVALUATION.	X
• The Context of a Concerned Cultural Community	X
• The Context of ArtsPartners as an Evolving Partnership	X
• The Context of a Growing Consensus about the Importance of a Large-Scale Evaluation	X
DESIGN PRINCIPLE 2: CREATE COMMUNITY-WIDE INVESTMENT.	X
Part II TRANSLATING DESIGNS INTO REALITY	X
DESIGN PRINCIPLE 3: ENGAGE STAKEHOLDERS IN KEY DECISIONS EARLY.	X
– COMMITTING TO A LONGITUDINAL STUDY.	X
– ESTABLISHING COMMONLY VALUED OUTCOMES: FINDING SHARED OUTCOMES	X
– DEVELOPING A CORE SET OF RESEARCH QUESTIONS.	X
– CONSTRUCTING THE SAMPLE FOR THE STUDY: CAREFUL CONTROLS.	X
DESIGN PRINCIPLE 4: ENHANCE THE CAPACITY OF ALL PARTICIPANTS	X
– THE MANAGING PARTNER.	X
– PARTICIPATING TEACHERS	X
– A NETWORK OF COMMUNITY RESEARCHERS.	X
– PARTICIPATING STUDENTS	X

Part III MEASURING OUTCOMES AND LEARNING FROM FINDINGS	X
DESIGN PRINCIPLE 5: PLAN FOR MIDCOURSE CORRECTIONS.	X
DESIGN PRINCIPLE 6: GRAPPLE WITH UNEVEN FINDINGS.	X
– EVIDENCE FROM STANDARDIZED READING SCORES.	X
– EVIDENCE FROM STANDARDIZED WRITING SCORES	X
– EVIDENCE FROM CLASSROOM WRITING SAMPLES: STRONG VS. WEAK ARTSPARTNERS INTEGRATION.	X
DESIGN PRINCIPLE 7: STAY ALERT TO SURPRISES	X
DESIGN PRINCIPLE 8: SHARE AND USE THE FINDINGS	X
Part IV EVALUATIONS THAT BUILD CAPACITY: LESSONS LEARNED	X
– DESIGN PRINCIPLES FOR EVALUATIONS THAT DO MORE THAN MEASURE	X
– LESSONS FOR ORGANIZATIONS PLANNING AN EVALUATION	X
– LESSONS FOR EVALUATORS	X
– LESSONS FOR FUNDERS	X
Part V APPENDICES	X
– APPENDIX A: LEARNER BEHAVIORS CODES	X
– APPENDIX B: LEARNER BEHAVIOR CATEGORIES	X
– APPENDIX C: NWREL 6+1 TRAIT WRITING ASSESSMENT	X
– APPENDIX D: SAMPLE ARTSPARTNERS INTEGRATED CURRICULUM	X
• Idioms Lesson Plan	X
• History Reporters Lesson Plan	X
– APPENDIX E: FOR RESEARCHERS	X
– APPENDIX F: ARTSPARTNERS LONGITUDINAL STUDY PARTNERS	X
• Lead Partners	X
• School Partners	X
• Cultural Community Partners 2001-2006	X
• Funding Partners	X

ACKNOWLEDGMENTS

Dallas Independent School District

Various staff members from the Curriculum and Instruction Services, Evaluation and Accountability, and Professional Development and Staff Training departments were instrumental to this study.

Dallas Arts and Cultural Community

Without the arts and cultural organizations that contributed their expertise, time, and programming, the research that follows would not have been possible.

Dallas City's Office of Cultural Affairs (OCA)

Funding from the OCA was a constant source of support for this work.

SPECIAL THANKS TO:

Tonyamas Moore

Charlene Payne

Anndee Rickey

Gabrielle Stanco

Thomas Wolf

FOREWORD

by Cyrus Driver

During the past 20 years, the most visible forms of educational assessment and evaluation have tended to focus on a limited number of measures, most notably student reading and math test scores. However, as the challenges and possibilities of a new century come into focus, a growing movement is afoot to ensure that the arts—as well as subjects such as civic education and the sciences—regain a central place in public education.

The growth of this movement has been fueled by new technologies, changing demographics, and rising concerns about the future of the United States in a globalizing economy. Technologies such as digital recording and photography, and virtual community spaces such as MySpace are transforming communication and providing new vehicles for artistic expression to millions of young people. Scholars and activists are recognizing the powerful role that art can play in communicating different social and cultural perspectives, and its potential to promote greater understanding across a diverse U.S. democracy. Leaders from across the political spectrum are recognizing that the 20-year standards movement has led to a narrowing of curricula and instructional practices, often at the expense of arts instruction that can provide students with the knowledge and skills they need to succeed in a 21st-century U.S. economy.

How is this nascent movement to redefine quality public education playing out on the ground? How are arts education advocates first “making the case” for the centrality of the arts in a public school education, and how are they building understanding and commitment among their communities? *More than Measuring* provides the story of how these questions were answered in Dallas, the 12th largest public school system in the United States, with over 163,000 students and 6,000 teachers.

Since 1997 the city’s lead organization on arts education, Big Thought, has worked with over 50 arts and culture organizations and their teaching artists to integrate the arts into the instructional practice of thousands of Dallas public school teachers. While many of the city’s civic, cultural, and educational leaders intuitively believed in the power of the arts in schools, many others demanded solid evidence, and everyone wanted to know the concrete ways that arts integration was making a difference in children’s lives.

This book offers a powerful story of how a large and diverse group of stakeholders was able to forge common goals regarding integrated arts education, including what should be evaluated and how the goals and activities should be assessed. They learned how they would define success and how they would use evaluation to improve their program. They learned the complexities and nuances of useful and well-designed evaluations, including the use of multiple methods—from test scores, to reviews of student work, to statements in students' own voices regarding program value. And throughout this process, the city's cultural, educational, and civic leaders came to a much clearer understanding of their own commitments regarding the education of Dallas students, including the centrality of the arts to the educational experiences of every child.

Today, Dallas is on the cutting edge of the movement to redefine high-quality education in the United States. A large, multiracial city with a large number of low-income families, Dallas is the type of place where public school reform is often seen as the most challenging. Yet Big Thought and its partners have gained remarkable, sustained commitments from the school district and civic and cultural leaders to deepen and extend the arts integration work. They also have received substantial private philanthropic commitments in recent years in recognition of the remarkable scale and quality of their arts integration programs.

It is for all of these reasons that I invite you to read this wonderful book. *More than Measuring* vividly presents the role that evaluation can play in making the case for arts integration, and in building common ground among civic leaders, parents, scholars, and others who have the capacity to return the arts to a central place in U.S. public education.

Cyrus E. Driver

Deputy Director

Education, Sexuality, Religion
Ford Foundation

February 2007

Introduction

All across the United States, communities are hard at work figuring out how to raise their children to become thoughtful and contributing adults. This involves improving how well children learn in schools. But families—and children themselves—want more than reading and mathematics proficiency. They want their children to flourish—to be safe, healthy, and considerate with a strong work ethic. However, to stay relevant in a world in which technology and information are changing at mach speed, young people will need more than effort and character to achieve this goal. Their capacity to earn a living will increasingly depend on their ability to think and work in innovative ways. This means that creativity is no longer for the gifted and talented—it is a basic skill.¹

In fact, the educational excellence challenge of this century is to organize learning for innovation. The equity challenge is to guarantee that gender, economic status, race, and native language cease to predict who will invent a vaccine, write a prize-winning play, or engineer a major breakthrough in technology. But realizing this vision depends on all children, throughout their education, being cultivated as potential inventors, entrepreneurs, and artists. It also depends on children having repeated opportunities to imagine, experiment, refine, polish, and edit. This is a tall order—it will take schools, libraries, museums, parks, and community learning centers working together. It will take new and redesigned resources. It will demand vision, will, and a fierce pursuit of quality programs.

Like any original work of art or science, learning how to educate for innovation requires experimentation, which produces errors, and necessitates recoveries and refinement. There must be many prototypes, drafts, and field trials. This report tells the story of how one community, Dallas, Texas, sought to pair education and creativity for lasting benefits. The programmatic result of this vision is almost a decade's worth of cultural programming for all elementary school students and professional development experiences for all elementary

¹ National Center on Education and the Economy (NCEE), *Tough Choices or Tough Times: The Report of the New Commission on the Skills of the American Workforce* (Washington, DC: NCEE, 2006).



classroom teachers in the Dallas Independent School District (Dallas ISD) through a citywide partnership known as Dallas ArtsPartners (ArtsPartners). However, just as important as the results of the partnership are the means and methods used to build, refine, and retool the partnership over the past nine years to ensure that it remains viable and relevant. Therefore, the focus of this report is the partnership’s commitment to the ongoing evaluation of what worked and why. In a process that lasted nearly five years from planning to final results, program designers, district administrators, classroom teachers, and students worked together alongside a national team to conduct research and build learning that was plowed back into the partnership. Thus, at the heart of this report is an examination of the role that program evaluation can play in building the capacity of a community to design, implement, and continuously refine the opportunities that all children need to flourish.

**WINNING AND UNDERSTANDING:
TWO VIEWS OF PROGRAM EVALUATION**

In 1998 more than 50 cultural organizations began serving the Dallas ISD elementary schools throughout the city via a coordinated program of services known as ArtsPartners. Only three years later, a major longitudinal evaluation of ArtsPartners’ impact began. At the end of the first year of the evaluation, project staff, cultural providers, board members, and evaluators met to examine the data. When the graph below flashed onto the screen, an audience member asked,

The stars on the graph, they show that when kids wrote during the ArtsPartners programs, their writing beat what they did in the regular school curriculum? Basically, the stars show that ArtsPartners won, right?

Fourth-Grade Spring 2002 Literacy Group Means
Classroom Curriculum (CL) vs.
Classroom + ArtsPartners Curriculum (CL+AP)

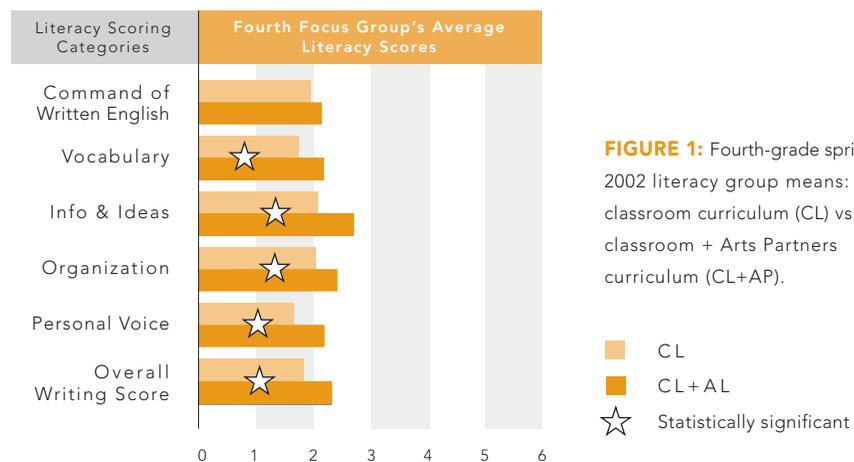


FIGURE 1: Fourth-grade spring 2002 literacy group means: classroom curriculum (CL) vs. classroom + Arts Partners curriculum (CL+AP).

■ CL
■ CL+AL
☆ Statistically significant

This was an honest question from a citizen-leader who is a strong advocate for the ArtsPartners program. His question reflects a widespread belief about evaluation: It’s about what wins. Although it is true that evaluations are about results, we will argue throughout this report that evaluations can and should be more than measuring sticks. Specifically, we suggest that evaluation is also a process for developing:

- An understanding of why a program works and what can be done to improve it
- A keen sense for what can go wrong with a program and how to correct it
- The capacity of a managing partner to reflect on the work as a whole and its effects
- A community-wide investment in the process of improving a program both on the micro (i.e., individual programs) and macro (i.e., citywide partnership program) level

In highlighting the different opportunities for learning, this report focuses on the evaluation of a specific partnership program, ArtsPartners, studied between 2000 and 2005. Readers who work in other settings, with other kinds of programs, under different conditions could easily ask, “What does this have to do with our work?” We are hopeful that the broad principles for building human capacity, the frank discussion of successes and failures, and the examination of lessons learned presented here will stimulate other civic partnerships to think about evaluation less in terms of documenting wins or losses, and more in terms of a process in which stakeholders collaborate to examine and improve programs.

The story of the ArtsPartners evaluation contains several chapters, each of which centers on one or more design principles for conducting evaluations in ways that build the capacity of communities to design and improve programs for children and youth.

- **PART I** describes the importance of understanding the contexts that the evaluation must weave together to be authentic and engage community support.
 - Design Principle 1: Tailor the Evaluation to the Context
 - Design Principle 2: Create Community-Wide Investment



- **PART II** examines how evaluations can be designed—from the outset—to enhance the capacity of many different stakeholders to make key decisions.
 - Design Principle 3: Engage Stakeholders in Key Decisions Early
 - Design Principle 4: Enhance the Capacity of All Participants
- **PART III** explores how to use multiple sources of evidence in developing a full understanding of when a program works and when it falls short.
 - Design Principle 5: Plan for Midcourse Corrections
 - Design Principle 6: Grapple with Uneven Findings
 - Design Principle 7: Stay Alert to Surprises
 - Design Principle 8: Share and Use the Findings
- **PART IV** summarizes the lessons learned for different groups of stakeholders.
 - Lessons for Organizations Planning an Evaluation
 - Lessons for Evaluators
 - Lessons for Funders
- **THE APPENDICES** offer more detail on the research tools and processes of the ArtsPartners evaluation.

A vision is not just a picture of what could be; it is an appeal to our better selves, a call to become something more.

— Rosabeth Moss Kanter²

The decision to use the ArtsPartners evaluation as an illustration of wider issues in the field of program evaluation—rather than as an end in itself—is not accidental. There is a growing sense that communities and organizations need more than a single approach to assess their investments.³ This is especially true at turning points, when multiple stakeholders are considering major commitments or new directions and when there is an opportunity to open that decision-making process to a more inclusive group. The evaluation of ArtsPartners occurred at such a moment and allowed for that wider collaboration and discussion. At the center of this discussion was the proposition that imagination and innovation are basic elements of any child’s education.

The ArtsPartners evaluation demonstrates how public and private resources can be focused on accomplishing one of the most urgent tasks facing the United States: how to create a culture of innovation in which everyone can participate and benefit. Of course, this effort has to ensure strong literacy and mathematics learning. But the ArtsPartners program also insists that a “basic” education should include learning to imagine and to invent, as well as the understanding that information can be created, not just consumed.

Finally, the ArtsPartners evaluation represents the kind of cross-sector collaboration that cities will need the courage and will to build if they are to make a major difference in the lives of urban children and youth.

² Quoted in IDRA presentation, Public Education Network conference, Washington, D.C., November 12-14, 2006.

³ The approach used in the ArtsPartners evaluation drew on a wide body of work that addresses this need for multiple approaches. This includes the work of evaluation specialists such as Eliot Eisner, David Fetterman, E. G. Guba, and Y. S. Lincoln.



Part I

EVALUATING FOR SPECIFIC COMMUNITIES AND PROGRAMS



DESIGN PRINCIPLE 1:

TAILOR THE EVALUATION TO THE CONTEXT

If we think of evaluations as being simply about measuring broadly valued outcomes such as academic achievement or school attendance, then an evaluation of after-school programs in Minneapolis could easily be replicated and used to evaluate arts and cultural learning in Dallas. However, if an evaluation is also meant to build the capacity of a specific community to design and implement programs that address its own unique challenges, then the process must be tuned to the contexts in which the results will be received, interpreted, and implemented. These contexts include national debates and concerns that shape local efforts, the needs and issues of the specific community, stakeholders' understanding of what the program needs to accomplish, and the particular moment in the history of the program or organization commissioning the evaluation.

THE CONTEXT OF GROWING ACCOUNTABILITY

As a nation, we like facts. When we make major investments in programs, we want “hard” evidence of their effectiveness: more individuals served, more effective use of dollars, more rapid cures, higher rates of employment, etc. An investment's moral or intrinsic worth is no longer enough to engage public or private philanthropy. During the War on Poverty in the 1960s and 1970s socially valued investments in human capital did not put an end to chronic problems of unemployment, achievement gaps in education, and racial segregation. Thus, policy makers—as well as many private foundations—began asking for scientifically collected and tested evidence of the effectiveness of programs.⁴ For the last quarter century, both government agencies and foundations have begun demanding rigorously collected data that provide statistical evidence of results, even when interventions operate in very complex human situations, take effect in

⁴ The W. H. Kellogg Foundation. Handbook on Evaluation. www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf

In this era of high accountability for basic academic skills, the arts are often seen as an extra or a frill. Throughout my time as a researcher I witnessed children labeled as failing or struggling find new entry points to literacy skills through the arts. It made me realize what a mistake it is to reserve the arts for children we see as gifted or for the weeks after testing. I am back in a classroom teaching and committed to making arts a constant.

— Sheri Vasinda, COMMUNITY RESEARCHER

varied ways that may not be easily measured quantitatively, and accomplish their ends over extended periods of time. The consequence is that both public programs and private philanthropic organizations frequently require that programs be designed around “evidence-based practice” to be eligible for funding.

In Dallas, the demand for accountability in educational programs has been particularly strong since the 1970s. Basic skills testing has been in effect in Texas since 1979 with the explicit goal of ascertaining which schools and districts are making effective use of the public investment in education. In 1994 the Texas legislature decided that student test scores would become the foundation of the state’s educational accountability system. In 2001, as a part of this accountability program, the Texas School Performance Review audited the Dallas ISD. A major finding was that there was “inadequate focus on education” and a “lack of accountability.”⁵ Beginning in 2003 the statewide Student Success Initiative mandated that every third-grade student had to pass the Texas Assessment of Knowledge and Skills (TAKS) test in order to move to fourth grade.

⁵ On August 24, 2000, the Dallas Independent School District (Dallas ISD) Board of Trustees asked the Texas Comptroller to conduct a performance review of the district’s operations. A copy of this review, published in June 2001, can be found at www.cpa.state.tx.us/tspr/dallas/toc.htm.

THE CONTEXT OF ARTS EDUCATION

From the 1970s through the 1990s, as school budgets were cut, many arts programs disappeared from public education. During those lean years, some committed arts practitioners argued that the arts could play a major role in supporting and amplifying other forms of learning. Although there is considerable experiential evidence that the arts have the power to engage students in learning, as well as other purposeful research along these lines,⁶ the exact nature of this chemistry remains far from an exact science. As a field, arts education has a rudimentary—but incomplete—understanding of which arts disciplines are effective in supporting learning in other disciplines.⁷ Also, researchers and practitioners of arts education are still in the process of learning what kinds of instruction, or how much experience, students need for a particular cultural experience to make a significant contribution to learning. Studies of learning outside of classroom settings are still charting how much time and what kinds of activities students need to deepen their understanding of important concepts such as experimentation, hypotheses, and the research process.⁸

Thus, as the evaluation for ArtsPartners was being planned, a much more demanding discussion about the effects of arts integration was occurring nationally. This broader discussion, and the looming evaluation, created an opportunity for the staff and collaborating cultural partners of ArtsPartners to ask the following questions:

- When does the addition of music, theater, creative writing, visual art, or dance increase children’s learning?
- What exactly does each of these experiences add?
- What dosage is necessary to make a difference?
- What level of quality is necessary before children’s learning is significantly affected?

It became clear that one of the features of the evaluation would have to be time to think. Evaluators would have to engage cultural partners—both individual artists and institutions—in reflecting on exactly what they could affect and how

⁶ James. S. Catterall, Richard Chapleau, and John Iwanaga, “Involvement in the Arts and Human Development,” Shirley B. Heath with Adema Roach, “Imaginative Actuality,” in *Champions of Change: The Impact of the Arts on Learning*, ed. Edward B. Fiske (Washington, DC: Arts Education Partnership and President’s Committee of the Arts and the Humanities, 1999), 1-18 and 19-34.

⁷ E. Winner and L. Hetland, “The Arts and Academic Achievement: What the Evidence Shows, Project Zero’s Project REAP (Reviewing Education and the Arts Project),” *The Journal of Aesthetic Education* (University of Illinois Press) 34, nos. 3/4 (Fall/Winter, 2000).

⁸ J. Falk and L. Dierking, *Learning from Museums: Visitor Experiences and the Making of Meaning* (Walnut Creek, CA: Alta Mira Press, 2000).

that could be reliably measured. An even more challenging goal for the evaluation process was to create a shared agreement across cultural partners and disciplines about the following:

- The kinds of learning outcomes to which a range of integrated arts experiences could all contribute
- What would constitute evidence that ArtsPartners experiences had supported learning in other academic areas (e.g., reading, writing, mathematics)
- The different effects that might occur immediately, in the short term, or over a longer period of time
- The variety of qualitative evidence that would shed light on why and how effects occurred

DALLAS AS A CONTEXT FOR EVALUATION

In the mid-1990s, Dallas had a rich infrastructure of arts organizations and facilities. Its fine art museum and symphony orchestra were leaders in their respective fields. Multiple theater and dance organizations and scores of small and mid-sized organizations representing a range of arts disciplines were distributed around the community. However, many in the city believed that more could and should be done. For example, many cultural facilities were clustered downtown and provided space for large performing arts organizations, while small and mid-sized organizations could not find or afford performance space. Many moderate to poor neighborhoods had little or no arts facilities outside of school auditoriums and arts classrooms. Much needed to be done to support neighborhood and community cultural development; partnership opportunities with the Dallas Library System and the Park Department were not fully utilized. Finally, as in many cities, there was a dearth of activities that supported cross-cultural and diverse programs.⁹

At the same time, through budget cuts and an increasing focus on purely academic outcomes, music and arts instruction had virtually disappeared from many Dallas ISD schools. This was particularly poignant given the fact that the district serves the poor and working-class children of Dallas, whose major chance for formal and sustained arts instruction is public education. In 2000, when this evaluation was in the planning process, only 160,000 of 217,500 Dallas children were enrolled in the public schools (see Figure 2). By and large, white, affluent, and highly educated families had abandoned public education in Dallas for private

⁹ Community Cultural Plan for the City of Dallas (February 2002), www.dallascityhall.com/pdf/Bond/CulturalPlan.pdf

and parochial schools. The net result was that exactly those families with the time and resources to advocate for arts education had left the scene. Classrooms were virtually emptied of children who had had consistent involvement with the arts. In essence, the public demand for arts education had fled the public schools.

	DALLAS ISD 2000		CITY OF DALLAS 2000	
	Number	Percent	Number	Percent
Persons between 6 and 18 years	162,188		217,510	18.3
RACE OR ETHNICITY				
White	12,640	7.9	83,741	38.5
African American	57,440	35.9	56,335	25.9
Hispanic or Latino	87,200	54.5	77,434	35.6
OTHER CHARACTERISTICS				
Language other than English spoken at home	77,753	47.9	80,696	37.1
Qualify for free/reduced lunch	105,428	65.0		

FIGURE 2: Comparison of the City of Dallas with Dallas ISD.

The Context of a Concerned Cultural Community

The Dallas cultural community faced a shrinking and homogeneous audience in a city that should boast a lively, multilingual, and multicultural arts community. In 1997 the city’s advisory board for the arts, the Dallas Cultural Affairs Commission, called for a more formal assessment of the state of arts education opportunities. It engaged an organization named Big Thought (known at that time as Young Audiences of Dallas) to gather data about the availability and accessibility of educational outreach programming offered by the city’s arts and cultural institutions.

Despite the difficulty of collecting disparate data on student participation in arts and cultural learning, the staff at Big Thought was able to demonstrate that serious inequities existed. Some public schools were rich in arts learning, while others were virtually starved. A charmed quarter of the district schools planned for and secured multiple performances, residencies, master classes, and field trips, while an estimated 75 percent of the schools received few if any services. The bottom line was that a majority of Dallas students were finishing high school without anything approaching an arts education. There was little sequential instruction except in the specialty arts classes in middle and high schools. Most students never attended a school-sponsored cultural field trip or live professional performance.

Believing in the benefits of arts programs for all children—not just those attending the most advantaged schools—the Dallas Cultural Affairs Commission advocated for systemic change. Determined to make a difference within a year, the commission created a citywide partnership with Dallas ISD, the cultural community, and Big Thought to bring arts opportunities to all young people in the city. In 1998 ArtsPartners came into being. Almost immediately, the stakeholders recognized a need for evaluation. Thus, they asked the question, How could an evaluation reveal the program’s ability (or failure) to meet its goals without jeopardizing its commitment to provide arts and cultural education for every child?

The Context of ArtsPartners as an Evolving Partnership

ArtsPartners was a partnership program designed to guarantee the equitable and high-quality delivery of educational services from more than 50 diverse cultural institutions to participating public elementary schools. Despite the name ArtsPartners, the program providers represented a wide array of organizations: historic sites, science museums, arboreta, nature centers, and the zoo as well as theaters, art museums, dance companies, and orchestras. Joining them were businesses and local foundations that made contributions of both cash and in-kind services (e.g., public relations and use of facilities).

ArtsPartners’ immediate mission was to build an effective infrastructure that would ensure that children and teachers throughout the city had the chance to learn about and participate in the cultural life of their city. These opportunities could come in many forms: school performances, field trips, artist residencies, master classes, workshops, and guided tours. All of the services would incorporate firsthand experience with creative professionals (e.g., dancers, instrumentalists, scientists, historians, journalists, archaeologists).

As ArtsPartners’ managing partner, Big Thought was responsible for leveraging the resources of all partners and raising private sector dollars. Dallas ISD contributed money to pay for direct services to students and authorized time for classroom teachers to meet with ArtsPartners integration specialists (staff and

It does enhance what you’re doing in the classroom... the things you learned that ArtsPartners wants you to do may be some of the things that you don’t normally do but you can do..., I continue to use these things in some other things that I teach.”

— Mr. Bryan Robinson, SIXTH-GRADE TEACHER AT CASA VIEW ELEMENTARY SCHOOL



cultural professionals). During this time (hour-long sessions with each grade level at all schools), specialists worked closely with teachers to design lesson plans infusing arts and cultural programming into a variety of subjects.

Throughout the period of the evaluation (2001-2005), the scope of the program increased dramatically. By 2002-2003 ArtsPartners offered arts and cultural learning to each of Dallas’ 101,000 public elementary school students, along with aligned professional development to 6,000 general classroom teachers. By centralizing the financial tracking of district money for cultural programming as well as professional supports, ArtsPartners was able to equalize what had once been a free-market vendor system that favored the most organized and well-resourced schools. At the same time, the model of arts integration was straining to apply equally to theater residencies and science inquiry. These concrete successes and expansions suggested a series of questions for the evaluation:

- Given the sheer scope of the program, is it possible to deliver services of enough quality and intensity to make a measurable difference?
- Given the admirable diversity of services (e.g., visits to the planetarium, writer residencies, tours of historic sites), is there a common core of experiences and possible outcomes that could be examined across schools and grades?

- Given the diverse interests of the different stakeholders (e.g., the school district, major cultural institutions, smaller cultural arts groups, funders), can they agree on a set of outcomes that can be tracked in the evaluation?

Beyond these questions, there were still others. The program was deeply rooted in a set of highly motivating and strongly held beliefs, many of which were still far from being evidence-based. These beliefs included the following:

- There are genuine and important overlaps between the academic learning of elementary school children and the ways of working common to professional artists, writers, curators, and designers.
- Some of these overlaps lie in habits of mind (e.g., intense engagement, close observation, empathy, revision); others lie in shared concepts (e.g., pattern, character, metaphor).
- Even young children can learn information or develop understandings in the context of a cultural activity and then transfer that insight to another field such as social studies or mathematics.
- Teachers can learn the techniques (e.g., engaging students as inventors, having students share observations and reflections) from artists and cultural providers in relatively short exposures.
- Teachers can then transfer these techniques to new contexts, enlivening and enriching their instruction.

The Context of a Growing Consensus about the Importance of a Large-Scale Evaluation

From the program's inception, the major stakeholders—the school district, the city, the cultural community, and Big Thought—were clear about the need for evidence of the impact on teaching and learning. Ongoing significant investment was unlikely without clear results. In a first step, Dr. Donna Bearden, executive director for special projects evaluation in the Dallas ISD's Division of Evaluation and Accountability, authored a report on the academic impact of the ArtsPartners program. The report compared reading and math standardized test pass rates from 13 schools receiving ArtsPartners services to those of the overall district and to those of a demographically matched set of schools not receiving ArtsPartners services. In both reading and math, children in the ArtsPartners schools outperformed the comparison and district groups by the third year of implementation. The report concluded that "ArtsPartners is a win-win program for all involved: the schools, the arts and cultural organizations, and the City of Dallas."

Bearden's report confirmed the efficacy of ArtsPartners as a partnership serving multiple needs in the community. It also stirred interest in the program and assured the district that it was acting responsibly by making additional investments in ArtsPartners. Like all initial studies, Bearden's report raised a next generation of questions:

- Were the ArtsPartners schools a special group of sites, led by strong principals and staffed by teachers who were willing to go beyond the call of duty?
- Would other measures of student achievement (e.g., samples of student work, interviews, observations of children's behavior) confirm these findings?
- Would a study of the same students (rather than schools) yield similar findings over time?
- What was driving this difference? Where did this difference in performance come from?

Dr. Bearden's initial report was received with much enthusiasm. Stakeholders were encouraged by the positive trends reported, but there was a desire to know what lay behind the findings.

In thinking about what they wanted from a follow-up study, the network of partners that constituted ArtsPartners agreed that it should do more than simply replicate—on a larger scale and with greater rigor—what had already been done. In an all-day session, Big Thought staff, cultural providers, Dallas ISD staff, and potential funders developed a list of what they wanted from the study:

- Of greatest interest was assessing the impact of the program on students.
- School board members and city leaders sought evidence about academic learning—especially evidence that spoke to the accountability demands they faced (e.g., raising standardized test scores). At the same time, there was an equal appetite for tracking additional outcomes related directly to children's learning about arts and culture, such as the ability to experiment and ask questions or to create work that was expressive and individual.

We had been saying that the program had an impact on kids. We had seen it with our own eyes in classrooms. But there came a point when we knew, as an organization, that we couldn't go on saying, "Just trust us. Art is good for kids."

— Gina Thorsen, SENIOR DIRECTOR, BIG THOUGHT

- The stakeholders wanted any evaluation to have an open format that allowed for feedback and refinements throughout the process. Described as a “smart” research study, it should be designed to evolve as it progressed. Results and lessons learned should be shared along the way, rather than awaiting a final report.



DESIGN PRINCIPLE 2:

CREATE COMMUNITY-WIDE INVESTMENT

Big Thought was new to the field of large-scale evaluation. Thus, the staff, along with their stakeholders, decided that a critical ingredient in the evaluation was a research partnership that featured strong ties both to the school district and to nationally known researchers and institutions.

From the beginning the integrity of the partnership that constituted ArtsPartners was critical. In terms of evaluation, the contributions of Dallas ISD were particularly important; district input came from staff at all levels and from most departments, including professional development and staff training, curriculum and instruction, evaluation and accountability, community affairs, and communications. This combination of high-level and broad district support ensured that principals and teachers participating in the study would carry out the lessons and work with the highest level of fidelity. It also ensured a central office investment in the study that has persisted across three general superintendents. At the level of implementation, Dallas ISD’s Division of Evaluation and Accountability approved the ethical and research protocols that ArtsPartners proposed. They also reviewed the proposed research questions, methodology, tools, and protocols. This shared planning created a foundation for collaboration that would last throughout the study.

For credibility, and also for the purposes of learning, the community of stakeholders wanted an external evaluator who would be willing to conduct the evaluation both as a research study and as a sustained conversation among the stakeholders about the current state and the possible future of the program. A special assessment committee, with representatives from different stakeholder groups, identified a list of research candidates whose approach to evaluation was consistent with their goals. Once the assessment committee had a candidate, they proposed an audition: a daylong session designed to help wrestle with the open questions about the focus and methods of the study. The process of selecting a candidate, designing the audition, and participating in the discussion was a first episode of learning for ArtsPartners stakeholders.

During the course of the evaluation, the principal evaluator moved to the Annenberg Institute for School Reform at Brown University. This resulted in widening the evaluation partnership even further. Because of the institute’s investment in improving urban districts, the Dallas work became part of a national conversation about the kinds of bold civic partnerships it would take to close the achievement gap in the United States, particularly in urban communities such as Dallas.¹⁰

Thus, from the outset the evaluation partnership’s focus on research design was balanced by an equally strong desire to build the capacity of the organization, its cultural partners, the school district, and the wider community, as well as the evaluators themselves. Long after the outside evaluator had come and gone, Big Thought staff, the cultural providers, and the schools that used ArtsPartners needed to be able to make ongoing judgments about current quality and what could be improved. Done right, the evaluation was intended to be a “school” for making this possible.

When the assessment committee asked if I would “audition” with them, the time was ripe. For years my own work had been about trying to change how schools and districts conducted student assessment. I had focused on assessment “as an episode of learning” that could enhance students’ sense of themselves as competent learners.¹¹ I had worked on using procedures from the arts—such as portfolios—as levers for changing instruction. The invitation to work with ArtsPartners was a chance to think about evaluation in a similar way. Personally, it was an opportunity to combine my skills in qualitative analysis with the quantitative tools that the district would bring to bear. But it was the idea of the interplay of so many perspectives that was irresistible.

—Dennie Palmer Wolf, PRINCIPAL EVALUATOR,
ANNENBERG INSTITUTE FOR SCHOOL REFORM

¹⁰ R. Rothman, ed., *City Schools* (Cambridge, MA: Harvard Education Press, in press).

¹¹ D. Wolf and N. Pistone, *Taking Full Measure: Lessons in Assessment from the Arts* (New York: The College Entrance Examination Board, 1992); D. Wolf and J. B. Baron, eds., *The National Society for the Study of Education XXV Yearbook: Performance-Based Student Assessment* (Chicago: University of Chicago Press, 1996); D. Wolf, “Assessment as an Episode of Learning,” in *Construction vs. Choice in Cognitive Measurement*, eds. R. Bennett and W. Ward (Hillsdale, NJ: Lawrence Erlbaum Associates, 1992).



Big Thought and the ArtsPartners program came of age at a time when there was growing emphasis on collecting hard, or scientific, evidence about program effects. This meant that an evaluation of ArtsPartners had to be rigorously designed to find out how the program contributed to student achievement. At the same time, given its roots in the arts, stakeholders wanted to ensure that any evaluation investigated whether the program affected more than basic academic skills.

Locally, the partners realized that, although they certainly wanted a “stamp of approval,” they also wanted to do the following:

- Build a network of partnerships that would support and eventually act on the findings of the evaluation. Rather than an audit, they sought something more like an ongoing conversation about the effectiveness of their work and how they could improve it.
- Build the capacity of Big Thought, the managing partner, to act as an ongoing assessor of the quality of programs, the fidelity of implementation, and the outcomes for participants.
- Provide participants (students, teachers, research staff, program providers) with important opportunities to focus their goals, expand their skills, and “take the plunge” into rigorous assessment.
- Create a climate or culture, both candid and respectful, in which the ongoing improvement of programs is the norm.

For the evaluation to be effective, its design, conduct, and range of results had to do more than just respond to these specific issues; it also had to reward and sustain this level of curiosity and reflection.

We entered into evaluation for “have to” reasons. We knew we had to demonstrate that the dollars we were getting from the district and the community were paying off. But the process of conducting the longitudinal evaluation changed that—not overnight or easily—but steadily. Evaluation became not just how Big Thought measured programmatic impact, but more important, it became a tool to design programs that would have impact. This transformation from having to do an evaluation to needing to do an evaluation is huge.

—Gigi Antoni, EXECUTIVE DIRECTOR, BIG THOUGHT

Part II

TRANSLATING DESIGNS INTO REALITY

The expression “the map is not the territory” could easily be the motto of any evaluation that is attuned to a particular community and program and invested in building human capacity in that context. This section details how the original hopes of the stakeholders in the ArtsPartners evaluation were translated into reality. The designers of the evaluation wanted one that would yield rigorous results at the same time that it engaged, learned from, and gave back to many different stakeholders—students, classroom teachers, cultural providers, and the staff of Big Thought. They also wanted the evaluation to engage stakeholders in key decisions while enhancing the capacity of all stakeholders to offer effective programs that made a significant difference in the lives of children.



DESIGN PRINCIPLE 3:

ENGAGE STAKEHOLDERS IN KEY DECISIONS EARLY

An important part of the early phase of the evaluation was building a broad partnership that included the national evaluation team, the staff at Big Thought, the school district, funders, and a representative set of arts and cultural providers. This group became a working partnership by wrestling with key decisions that set the direction for the study and the ways partners would collaborate and communicate throughout the process.

COMMITTING TO A LONGITUDINAL STUDY

The quickest and most efficient way to approach the evaluation would have been to conduct a one-year study by collecting data from students enrolled in grades 1 through 6. Then, assumptions could have been made about how students learn about creating new works and ideas from kindergarten through sixth grade.

In order to learn how the effects of the program accrued across multiple years, researchers and the major stakeholders made the costlier and more difficult to execute decision to conduct a study that would follow the same children over

time (i.e., a longitudinal study).¹² This choice was made by considering a variety of needs and opportunities:

- Although looking at different groups of students enrolled in a series of grades can suggest how the program might affect a child over time, the only real way to understand this is to watch the same children growing up with ArtsPartners as an integral part of their learning.
- A longitudinal study would also allow the time to develop an understanding of the multiple effects the program might have. For instance, the immediate effects on student behavior, the intermediate effects on learning within the same year, and possibly whether students continued to be affected even as they moved into middle school (where there was no similar programming).
- By working together over a more sustained period of time, the partners would have the opportunity to look at the findings from each year of the study and jointly develop the habit of building reflection and improvement into the annual cycles of planning and implementation.

However, given the urgency of developing findings to share with the wider community and to use in strengthening programs, a modified longitudinal study was developed. Rather than follow one group of students throughout elementary school, the study would follow two groups: one beginning in first grade (the grade 1



¹² J. D. Singer and J. B. Willet, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* (New York: Oxford Press, 2003).

cohort) and one beginning in fourth grade (the grade 4 cohort). In this way, all elementary grades would be covered. (Eventually, those same children were followed on selected measures past the end of the study to find out whether any of the effects endured.)

Given the longitudinal scope of the study, and thus a delay of several years before the final results would be determined, the cultural partners and researchers decided to create an annual event to share findings to date and surprises in the data. This event also challenged the partners to actively reflect on the information shared. The point was to build the habit of being accountable, even as the findings emerged.

ESTABLISHING COMMONLY VALUED OUTCOMES: FINDING SHARED OUTCOMES

The growing demand for accountability at the state and national levels was challenging for ArtsPartners. The program had been founded and designed on the twin principles that all children deserve access to arts education and that infusing the arts throughout the curriculum would help children to understand the role of invention and imagination in learning.

Honoring these principles while being asked to embrace a scientifically rigorous evaluation raised the following concerns in both Big Thought staff and cultural providers:

- Would state achievement tests become the sole yardstick for measuring the success of the program?
- Would the program's emphasis on creativity be pushed aside?
- If the evaluation emphasized "hard" evidence, what would happen to the experimentation and invention that had been the hallmarks of ArtsPartners programs?
- Would the pressure to demonstrate outcomes valued by the schools end up making theater, dance, music, and art "workbooks" for basic skills?
- Would the demand for scientifically rigorous evaluation designs consume human and fiscal resources that were once dedicated to programs for children?

These questions made an excellent starting point for the evaluation. By putting them on the table, everyone was engaged in an open discussion of commonly valued outcomes and shared commitments about how to work together. Cultural educators wanted the study to consider what performances, residencies, and field trips added to students' learning. Classroom educators were adamant

that the study should help to identify the best teaching and curriculum practices for conveying core academic learning. Their interests came together in a basic question: What is the difference in student outcomes between a well-taught and well-resourced classroom lesson and a similar lesson infused with ArtsPartners programming?

But what student outcomes would be measured? The study needed a common focus so the separate experiences offered by a variety of cultural partners could be combined into a substantial treatment with measurable effects. However, many ArtsPartners providers favored—or were most familiar with—measuring outcomes specific to their programs such as singing and improvisation in music, or knowing about outer space following a visit to the planetarium. This forced all stakeholders to think hard about where cultural and academic learning intersected in substantial ways.

Teachers shared that effective writing instruction includes more than attention to phonics and spelling. It demands that information be delivered in compelling ways to engage the reader. They talked about how “more than basic” literacy requires the generation of new ideas and powerful forms of expression (e.g., word choice or the use of figures of speech). Similarly, powerful reading instruction involves more than cracking the alphabetic code; it includes helping young readers infer what is going on, to make connections, and to visualize events or ideas in their “mind’s eye.”

These complex ideas about literacy resonated with cultural educators and suggested a point of intersection. Cultural providers discussed how investigation of early literacy has demonstrated the contributions of activities such as story reading, play, and dramatization.¹³ They also discussed research regarding children’s spontaneous exploration of literacy in which they invent the rules that inform

They [ArtsPartners] give us [teachers] everything we need.
And we’re always prepared, because we go through it
all beforehand and how we could implement this and that.
Also, the resource people, the artists, have really been
great to work with.

— Dawn Jiles, THIRD-GRADE TEACHER AT CASA VIEW ELEMENTARY SCHOOL

¹³ John Ainley and Marianne Fleming, *Literacy Advance Research Project: Learning to Read in the Early Primary Years* (Melbourne, Australia: Catholic Education Commission of Victoria, and the Australian Council for Educational Research [ACER], for the Department of Education, Science and Training, 2000).

how adults make lists, write books, create maps, and so forth.¹⁴ Learning how to use symbols (e.g., letters, numbers, drawings) was also considered a critical skill common to both academic and cultural learning. For example, a student must use the same skill to read textbooks as he uses to navigate an exhibit using labels and signage. Similarly, maps and graphs are as likely to show up at a historic farm as in a social studies class. And finally, students have to interpret theater and dance, just as they have to interpret the data in a science experiment. Thus, after much discussion, the partners settled on a broad definition of literacy, the generation of ideas and their expression, as the right overlap between familiar forms of classroom learning (e.g., speaking and writing) and imaginative learning (e.g., drama, song writing, stories, poetry—even experimentation, design, and research).

This view of literacy is very similar to one that informs the work of the famous preschools of Reggio Emilia, Italy. There the educators embrace the idea of the “hundred languages of children,” believing that, like speech and writing, music, drawing, puppetry, gardening, imaginative play, and dance are powerful ways of capturing, expressing, and sharing experiences.¹⁵ Similarly, in recent international studies, reading literacy is defined as more than the ability to read texts. It is seen as an essential tool for achieving life goals, developing one’s knowledge and potential, and participating in society.¹⁶

With a decision made about where to investigate the intersection of their work, teachers and cultural providers began to put together working theories about what they would each have to contribute for classroom and ArtsPartners learning to support one another. They suggested that arts and cultural experiences are powerful allies to other forms of learning when they do the following:

- Provide rich and stimulating content
- Model how to translate personal experience into compelling messages in new and effective ways
- Provide the supports for helping a full range of students—including those who are acquiring English and those with learning disabilities—to engage in this kind of activity in increasingly competent ways
- Actively demonstrate how these skills and insights could translate to other areas

¹⁴ C. Genishi and A. H. Dyson, *Language Assessment in the Early Years* (Norwood, CT: Ablex Publishing, 1984); C. Genishi and A. H. Dyson, eds., *The Need for Story: Cultural Diversity in the Classroom and Community* (Urbana, IL: National Council of Teachers of English, 1994); John Willinsky, ed., *The New Literacy: Redefining Reading and Writing in the Schools* (New York: Routledge, Chapman, and Hall, 1990); Dennie Wolf and A. M. White, “Charting the Course of Student Growth,” *Educational Leadership* 57, no. 5 (Alexandria, VA: Association for Supervision and Curriculum Development).

¹⁵ Carolyn Edwards, Lella Gandini, and George Forman eds., *The Hundred Languages of Children: The Reggio Emilia, Approach—Advanced Reflections* (Norwood, CT: Ablex Publishing, 1998).

¹⁶ Organization for Economic Co-operation and Development (OECD), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (Paris: OECD, 2003).

Thus, teachers, cultural providers, and researchers began a long conversation about quality that was an important thread throughout the life of the study.

The products of this conversation were joint lessons codesigned by teachers and their cultural partners to help young readers and writers develop enriched and meaningful forms of expression. In first grade, classroom teachers ensured that students had strong foundational skills such as phonics and the conventions of written English. Docents in the Dallas Museum of Art helped children expand those skills through looking at and talking about the visual images they discovered in the galleries. In many cases the combination was stunning:

When they began to put words on paper, they understood that it wasn't just something academic, that it is actually communicating. Before we write, we always draw. I think that...when we started adding details to writing, they understood it right away, because they understood putting details in their drawing.

— Ms. Wood, FIRST-GRADE TEACHER AT WALNUT HILL ELEMENTARY SCHOOL

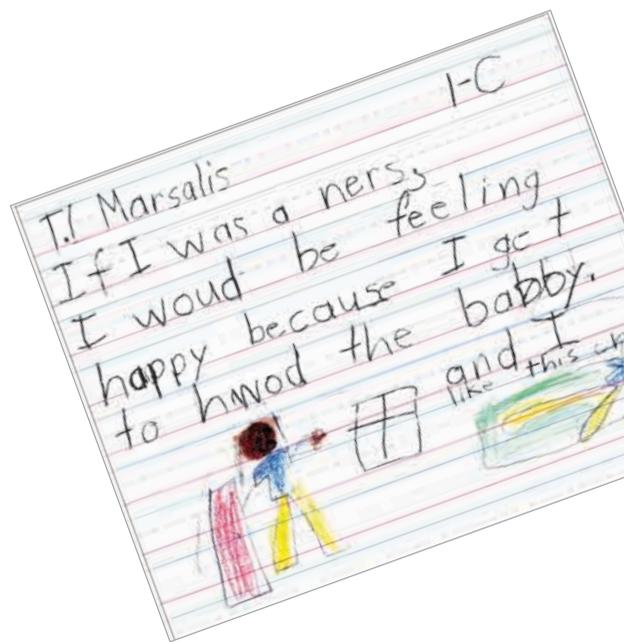
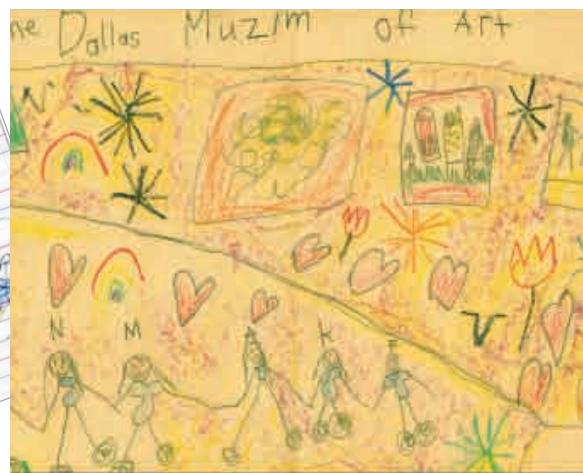


FIGURE 3: A six-year-old's use of multiple languages to depict her experience at the Dallas Museum of Art.



Being playful makes me feel like going to the park. the way I look, being happy makes me feel like I'm all differen kind of colnos on the inside.

BEFORE: Image at left shows one first-grader's writings before an ArtsPartners unit of study.
AFTER: Image at right shows the same first-grader's writing one week after an ArtsPartners unit of study that included a trip to the Dallas Museum of Art, a hands-on arts experience in the classroom, and teacher-led lessons discussing how art conveys emotion. Notice how much richer the student's expression of ideas and words are.

DEVELOPING A CORE SET OF RESEARCH QUESTIONS

Based on this first round of joint lessons, the research team (outside evaluators, Big Thought staff, and Dallas ISD classroom educators and the cultural community) worked together to develop a small set of focused research questions. A first and major decision was to place student learning at the center of the evaluation. This would be complemented by modest work with teachers and the staff of cultural organizations. But first and foremost, people wanted to understand what was—and what wasn't—occurring for students.

Three questions were primary:

- Do ArtsPartners experiences have immediate effects on participating students? Specifically, does the same student behave differently in ongoing classroom instruction than during episodes of arts and cultural learning?
- Does ArtsPartners learning contribute to students' achievement in school? Specifically, is students' literacy enhanced by the addition of arts and cultural learning? What is the nature of these effects?
- Do the effects of ArtsPartners programs last?

CONSTRUCTING THE SAMPLE FOR THE STUDY: CAREFUL CONTROLS

To test the effects of ArtsPartners rigorously, evaluations would have to examine the differences between student performance in classrooms where teachers collaborated with cultural partners to design integrated lessons and student performance in classrooms where teachers selected cultural programming without the benefit of collaboration or guided planning with Big Thought staff. With the help of the staff of Dallas ISD, researchers and Big Thought staff selected four treatment schools whose collective population matched the demographics of the district as a whole in terms of economics, gender, ethnicity, race, first language, and academic performance. The team then chose control schools that matched the demographics of the treatment schools. Throughout the study all eight schools were given the same provision of ArtsPartners experiences—the funds, information, and logistical support to provide their students with programming from 50+ cultural providers.

To be clear: control schools received money from ArtsPartners to purchase arts and cultural experiences for students. What they did not receive was the program's professional development component. Unlike classroom teachers at the treatment schools, control teachers did not meet with ArtsPartners integration specialists (staff and cultural community professionals) to design lesson plans that infused arts and cultural programming to enrich learning in a variety of subjects.

Thus, the difference between the two groups of schools was not in the numbers of opportunities available, but in the way those opportunities were implemented. Thus, any results would have to do with differences in teaching and learning—rather than the sheer excitement and novelty of field trips, special visitors, etc.

To facilitate reliable and valid statistical analyses, four comparison groups were created¹⁷ as shown in Figure 4.

FOCUS STUDENTS¹⁸. These 64 children (eight per cohort per school) received the most intensive experience. In addition to participating with their classmates in ArtsPartners programming twice a year, these Focus students were observed during and interviewed after the implementation of each classroom and ArtsPartners unit of study. By design, Focus students at each school remained together in the same class for all four data collection years.

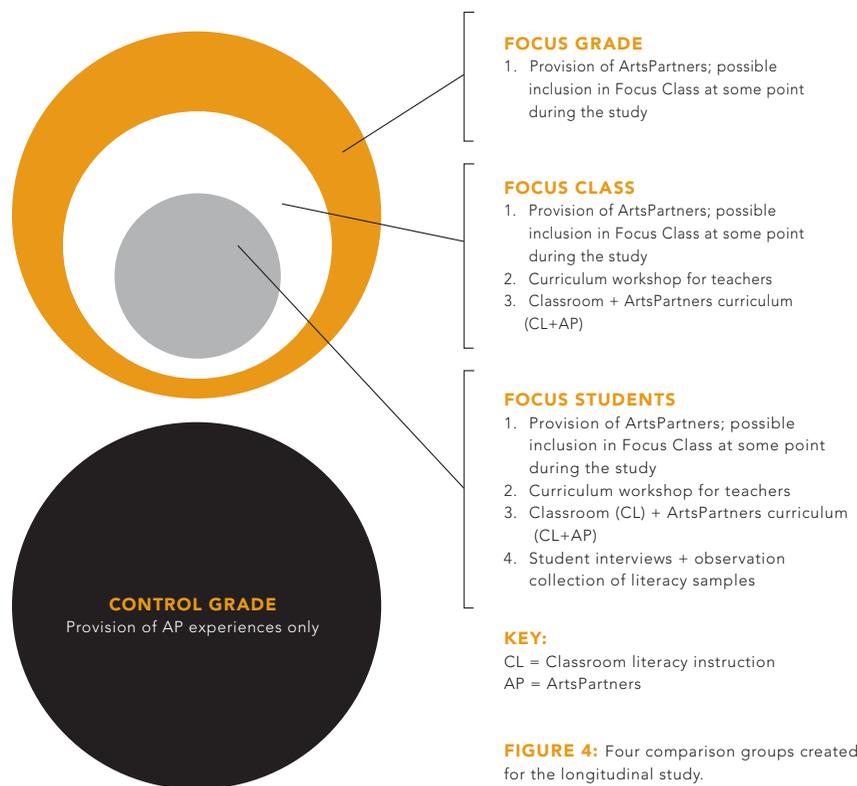


FIGURE 4: Four comparison groups created for the longitudinal study.

¹⁷ All students who were enrolled in one of the eight study schools (four treatment and four control) at any time during the four data collection years of the study (2001-2002, 2002-2003, 2003-2004, and 2004-2005) were coded by year into one of these groups. Therefore, group membership changed slightly each year.

¹⁸ Teachers completed a survey for each child who had familial permission to participate. The survey asked for students' status on talented and gifted designation, special needs designation, free or reduced lunch, ethnicity, and learner level (learner level was a designation of low, medium, or high based on the teacher's knowledge of each student's previous academic performance). Using this information, the researchers selected a diverse set of students representing all the factors.

FOCUS CLASS STUDENTS. These children were enrolled in the same classroom as the Focus students. Focus Class students received the same ArtsPartners programming as Focus students, but were not observed or interviewed individually.

FOCUS GRADE STUDENTS. Focus Grade students were enrolled in one of the four treatment schools at the same grade level as the Focus Class. These students received the same ArtsPartners experiences (i.e., arts and cultural programming) as Focus Class students did, but their teachers were not involved in curriculum design and discussion, and the students were not interviewed.

CONTROL GRADE STUDENTS. Control Grade students were enrolled in the control schools at the same grade level as the Focus Grade students for that year. These control schools chose their own ArtsPartners experiences, but they did not receive the ArtsPartners professional development component.



DESIGN PRINCIPLE 4: ENHANCE THE CAPACITY OF ALL PARTICIPANTS

Developing the sample populations for the study might seem like a purely mechanical part of the study. However, in trying to ensure that a representative group of children was involved, Big Thought staff as well as members of the cultural community came face to face with how heterogeneous Dallas classrooms are in terms of ability, first language, and ethnicity. At the same time they learned how uniformly poor most children in the district are. From this came a respect for the work classroom teachers do day in and day out in providing children with fundamental concepts and information. What began as a piece of technical work was a critical first step in building a partnership. ArtsPartners stakeholders could no longer base their work on simplistic “good/bad” comparisons such as the following:

Classroom Teacher = Task Master **ArtsPartners Teacher** = Creator, Inventor
Classroom Content = Basic Skills **ArtsPartners Content** = Higher-Order Skills

Creating mutual respect was vital to building an increased understanding of the complementary nature of both classroom and ArtsPartners learning. This understanding was essential to winning the trust of classroom teachers—they had to know that they were not viewed as “bit players” compared to artists. Rather, their work was seen as the foundation on which cultural experiences build and to which they add value.

This points out how an evaluation is more than a design challenge. Any evaluation is also an intensely human process, much like a long conversation or a play that evolves over time. Viewed in this way, even the most technical aspects

of an evaluation are occasions for learning. Unconventional as it may be, the capacity-building mandate of the ArtsPartners evaluation demanded that for every bit of data collected, an equal amount of respect, insight, development, or help must be contributed to whoever was involved—principal, teacher, artist, writer, researcher, or student.

THE MANAGING PARTNER

When the evaluation began, Big Thought had a staff of 18, none of whom had assessment as a part of their job description. Five years later, the organization now has a director of assessment, a full-time program manager, a robust partnership with the Dallas ISD’s Division of Evaluation and Accountability, and an ongoing relationship with six consultants who extended the expertise of the organization. In many ways, the growth of this capacity for planning and evaluating came about through the longitudinal study of ArtsPartners.

It was clear that the work of the evaluation would need to be coordinated and run by Big Thought, even as it was “steered” by national evaluators. This collaboration of national expertise with local knowledge also allowed two Big Thought staff members to bring insights and skills from other disciplines. One was a science teacher and curriculum developer; the other was an experienced museum educator, with a keen sense for communicating complex ideas in direct and powerful ways. Both had worked extensively with teachers and had a deep appreciation for children’s thinking.

PARTICIPATING TEACHERS

Prior to the start of each semester during the study, the research team worked with the teachers from that year’s Focus Grades to design curriculum modules that focused on a “big idea” at the intersection between classroom learning and what cultural providers had to offer. Teachers first agreed on the learning objective they wanted to use as a focus.¹⁹ Then research staff and national evaluators worked with one or two teachers to draft a module or unit in which students could explore that topic. During this time teachers and researchers selected a cultural partner whose work could inform and strengthen the concepts and literacy skills that formed the core of the proposed learning unit.

¹⁹ Two examples: Fourth-grade teachers were adamant that the spring semester curriculum focus on writing because the state test for fourth-grade writing took place the last week in February. Fifth-grade teachers asked that their fall curriculum give students background knowledge that would support the Beyond the Stars unit in their language arts curriculum.

Each unit had the following characteristics:

- It was focused on key content from the Texas state standards (Texas Essential Knowledge & Skills; TEKS).
- It was designed to provide in-depth literacy learning that included:
 - A range of different types of text
 - Reading materials that supported the development of sophisticated reading comprehension skills (e.g., analysis, interpretation, application)
 - Opportunities for students to learn how to write connected texts (e.g., letters, reports, stories in which they could discover and express original ideas)²⁰
 - Strategies native to one or more forms of cultural learning (e.g., dramatization to help children imagine how people might have lived in a historic house)

In each unit children had two opportunities to develop their literacy skills. The first was during ongoing classroom literacy instruction (CL). The second was an ArtsPartners unit of study designed to provide students with new forms of world knowledge, collaboration, and motivation (CL+AP). In both, children generated the same type of writing (e.g., biography, song lyrics, persuasive writing).

The teachers were very important informants and codesigners of the process. Each year a teacher in the study served as part of the design team that selected and developed the curriculum. Then, throughout the year, additional teachers helped researchers interpret what they observed in classrooms. This level of involvement meant that teachers had a deep understanding of the ArtsPartners curricula and could carry them out with fidelity.

Figure 5 provides an overview of the preparation and classroom teaching and data collection during procedures employed each semester of the study. Two ArtsPartners lesson plans are available in Appendix D.

²⁰ In particular, these features are highlighted in *Preventing Reading Difficulties in Young Children*, edited by Catherine E. Snow, M. Susan Burns, and Peg Griffin (Committee on the Prevention of Reading Difficulties in Young Children, National Research Council, 1998). The report’s emphasis on the powerful role of background knowledge, rich vocabulary development, and the active processing of texts has helped cultural partners see the many connections between literacy learning and the experiences they offer (e.g., trips to historical sites, chances to apply science vocabulary to experiments at a science center, musical scores, dance notation, timelines, and so on, as examples of texts to be interpreted).

Procedures	Lead Party	Materials generated or collected
Design of Classroom (CL) and Classroom + ArtsPartners Curriculum (CL+AP)		
When: Prior to semester <ul style="list-style-type: none"> Determine “big idea” (desired skills and knowledge outcomes) Determine cultural program best suited to achieve “big idea” in the CL+AP lessons Write lessons 	Classroom teacher Classroom teacher and ArtsPartners (AP) educators	Model curriculum
Curriculum Workshop		
When: First month of semester <ul style="list-style-type: none"> Model CL lessons with teachers Experience arts/cultural program selected for CL+AP lessons Model CL+AP lessons 	Big Thought staff ArtsPartners educators Big Thought staff	Teacher survey
CL Curriculum Implementation in Schools		
When: Two to three weeks after curriculum workshop <ul style="list-style-type: none"> Prelesson(s) introducing concepts and cultural program Culminating lesson with writing outcome 	Classroom teacher Classroom teacher	CL+AP prewriting observation
CL+AP Curriculum Implementation in Schools		
Where: School When: One or two weeks after CL curriculum <ul style="list-style-type: none"> Prelesson(s) introducing concepts and cultural program Cultural program Culminating lesson with writing outcome 	Classroom teacher ArtsPartners educator Classroom teacher	CL+AP writing observations and writing samples

Key: CL = Classroom Literacy Instruction AP = ArtsPartners

FIGURE 5: Overview of Preparation and Data Collection

A NETWORK OF COMMUNITY RESEARCHERS

Running the study required many more person-hours than the two staff positions at Big Thought could provide. Rather than strain the budget, researchers decided to build a major training component into the work. The motivations for this were more than fiscal. A key to the design of this study was to build organizational and community capacity while delivering the necessary research product. The belief was that research collected with a community would be much more likely to be understood, owned by, and acted on by that community.

ArtsPartners recruited widely in Dallas and the surrounding areas for teachers, retired teachers, graduate students, and staff members from participating cultural institutions to participate in the study as field researchers. It was particularly important

that they be interested and experienced in working with Dallas students, including those who were still in the process of learning to express their ideas in English.

Through ongoing training sessions, these field researchers received the equivalent of several research courses. They learned the fundamentals of educational research, including how to do the following:

- Work respectfully within a school environment
- Interview using a protocol (i.e., asking students to expand and explain their answers without leading them on)
- Conduct classroom observations
- Analyze student writing samples
- Reach inter-rater reliability for coding the data

Every semester, each researcher was part of a two-person team responsible for collecting comprehensive data at one school. The research team observed two students for both the classroom lessons and the lesson infused with ArtsPartners programming at “their” school. Researchers were also encouraged to participate in the ArtsPartners experience (e.g., field trip, residency, in-class program).

Pairing researchers and schools provided the research team with a deeper understanding of the ArtsPartners program, strengthened researchers’ intuitions about where to look for effects, and established close relationships between students and researchers that were critical to the quality of observations and interviews.

Once the curriculum was complete, researchers interviewed eight Focus students individually from their designated schools. After each interview, researchers also completed interview impression forms to highlight notable points from the interview. Teachers were also interviewed to provide their general impressions of the entire lesson process as well as their perspective of how each of the Focus students responded or participated throughout.

In the training sessions, researchers and classroom teachers often brought very different perspectives to the work of coding the data. Sometimes researchers would want to credit children with all kinds of insight and intentions. Often the teachers would challenge that view, knowing the strict rules used for scoring on state tests. The back and forth made both sides more thoughtful.

—Jennifer Bransom, DIRECTOR OF PROGRAM ACCOUNTABILITY, BIG THOUGHT

I enjoyed the training and the work with kids in schools, but actually my favorite part was when we worked on developing the coding together. Always before, I had been told, “Do this.” But in this case, we watched tapes, talked about what we saw, and argued through what was worth trying to capture through our coding. That part was great—we were a part of the research process.

— Janet Morrison, FIELD RESEARCHER, ARTSPARTNERS

PARTICIPATING STUDENTS

In addition to discovering the hoped-for benefits from arts and cultural programs, the evaluation was also designed to “give back” to participating students. Following each of the units of study, the Focus students were interviewed about their experiences, the works they created, and whether what they learned spilled out into other portions of their lives. The interview had two major segments:

- **BIOGRAPHY OF A WORK.** In this segment, students explained work they created in the classroom and during an ArtsPartners experience. (These works varied from song lyrics to a program for a dance performance.) Their conversations with interviewers included questions about the source of their ideas, resources they used, difficulties, revisions, and so forth. The purpose was to explore how children described themselves as learners in these two learning environments. Students understood the interviews as opportunities to think aloud and often pushed themselves to share emerging thoughts. For example, below a fourth-grade student describes how she wants to continue exploring the connection between music and writing:

I want to read more books about how—how they make sounds in music. The important thing is about what are you going to do in your story, and everything else. When you got everything you wanted in your brain—in your mind, and wanted to put it in your story, that means that you’re going to be a great writer.

- **DIARY DAY.** Interviewers asked students to narrate their activities during a recent day, asking about who was present and what they did. The purpose of Diary Day was to see (1) whether students made explicit, or implicit, connections to their recent ArtsPartners experience and (2) whether, over time, students would report a higher incidence of formal and informal activities that reflected a growing investment in the arts, and cultural activities or informal investments in imaginative activity. In this example, we hear how a sixth-grade student convinced his mom to take him shopping for a particular author’s work after working with an ArtsPartners provider, The Writer’s Garret.

I told my mom if she can, like, take me to the bookstores... I asked her if she can take me, because I said I was learning about Tim Seibles and some other um, authors.

Student interviews were designed to help evaluators and program staff get a “child’s eye view” of the ArtsPartners program. They were also intended to provide insight into why some experiences had significant effects on children’s ability and willingness to express original ideas in their writing. At the same time, these interviews were also rare chances for students to practice talking about themselves and their work to interested adults.



Part III

MEASURING OUTCOMES AND LEARNING FROM FINDINGS

ArtsPartners researchers selected and designed a diverse set of outcome measures in order to gather both qualitative and quantitative evidence to answer the three key research questions that had been defined. The following figure summarizes these questions and the measures used to answer them:

QUESTIONS	QUANTITATIVE	QUALITATIVE
<p>Do ArtsPartners experiences have immediate effects on participating students? Specifically, does the same student behave differently in ongoing classroom instruction than during episodes of arts and cultural learning?</p>	Classroom observations (coded for frequency and intensity of learner behaviors exhibited by participating students)	<p>Student interview data</p> <p>Classroom observation data</p>
<p>Does ArtsPartners learning contribute to students' achievement in school? Specifically, is students' academic literacy enhanced by the addition of arts and cultural learning? What is the nature of these effects?</p>	<p>State assessment test scores in reading and writing during the years of the study</p> <p>Writing samples (scored on six traits of writing development)</p>	<p>Student interview data</p> <p>Classroom observation data</p>
<p>Do the effects of ArtsPartners programs last?</p>	State assessment test scores in reading and writing in years after the study	

FIGURE 6: Research methodologies used in the ArtsPartners longitudinal study

The following section takes the research work that was just outlined and reexamines it in light of how these processes—which are often confined to “experts”—can provide additional opportunities for capacity building for the many different individuals who help to design, carry out, and interpret the evaluation. In the case of each type of data (immediate classroom behaviors, contributions to academic achievement, etc.), examining the design and data allowed stakeholders to grow their understanding of ArtsPartners programs, how students and teachers use them, and how the programs could be strengthened.

➤ DESIGN PRINCIPLE 5: PLAN FOR MIDCOURSE CORRECTIONS

Having made a decision to investigate the intersection of their work through the lens of literacy, teachers and cultural providers began to put together working theories about what they would each have to contribute to support one another. Both cultural providers and teachers suggested that there would be marked differences between children’s behavior during a well-taught and well-resourced classroom lesson (CL) and a similar lesson infused with ArtsPartners programming (CL+AP). To explore and test this theory, researchers observed students in both CL and CL+AP settings and coded the frequency and intensity of over 20 learner behaviors (see the figure in Appendix A).

LEARNER BEHAVIORS. These behaviors characterize the many ways in which students demonstrate, explore, and acquire new knowledge and skills. Examples of learner behaviors include asking questions, studying what other children are doing or making, revising one’s own work, experimenting with materials, and seeking additional resources.

During the first years of the study, field researchers observed children as they created their final products (e.g., songs, stories), expecting that following the CL+AP lessons there would be much more conversation, peer consultation, editing, and so forth. That prediction was totally wrong; no matter the content of the lessons, or the work being produced, children tightly controlled their actions when writing in the classroom. Later, in conversations with teachers, researchers realized that this was due to the need to practice the constrained behavior expected of students during the administration of state standardized writing tests. Thus, researchers had to go back to the drawing board, discussing and reviewing their notes and videotapes. In stepping back, they realized that there were differences in learner behaviors—but they showed up only in the more freewheeling prelessons

when teachers and cultural providers worked with children to introduce concepts and explore ideas. This realization called for a major midcourse correction in data collection procedures.

Figures 7 and 8 show sample observation notes and learner behavior codes assigned to the same student in a CL prelesson and a CL+AP prelesson exactly seven days apart. In the first lesson (Figure 7), the student followed the teacher working on the overhead projector and sought opportunities to participate by reading aloud (RHR: raises hand to respond) and answering questions (VP: verbal participation).

In the CL+AP lesson (Figure 8), a theater artist asked the students to show what they had learned about pioneer living by working in groups and using only their bodies to express their knowledge. The observed child participated verbally (VP) and answered questions (AN). But, she also discussed (CL: collaborating) various ideas with her peers (P) and then physically became (PE: play/experimentation) different objects in a pioneer home, such as a table. Then, with pride she showed and explained (SW) her work to others in and outside her group.

Student Observation Form

* Place a Star next to events that you believe are “aha” moments (special occurrences not to be missed in data analysis)

TIME - CONT	LB CODE	EVENT	NOTES “overall impressions”
6	OT	reads logic book & looking @ worksheet	
7		reads overhead	
8		follows overhead & w/	
		reads worksheet	
		looks @ teacher	
		looks @ T	
		writes	
		adds details to table	
		raises hand to read	
9	OT	reads aloud w/ pencil on each word	
		writes notes on worksheet - working ahead	
10		reads aloud	
		looks @ teacher	
		writes	
		looks @ T	
		looks @ work	
11	OT	reads aloud w/ pencil on each word & reads aloud (no inside voice)	
		reads aloud	
		“It’s Corpus Christi”	
		reads “response” Texas	
12	OT	looks @ T	
		writes in worksheet	
		writes	

Observation Date: 9/21 School: Hogg Classroom Teacher: Ms. Dzierz Grade: 4th Observer Initials: JAB

Student Name: _____ 1st Segment or 2nd Segment Student ID#: 30184 Lesson Type: AP / CL

Lesson Description: CL #1

Classroom Setup: Whole Class Big Groups Small Groups (6 or less) Individual Semester: Fall Spring

FIGURE 7: Student observation form completed during a classroom lesson (CL).

Student Observation Form

* Place a Star next to events that you believe are "aha" moments (special occurrences not to be missed in data analysis)

TIME - CONT	LB CODE	EVENT	NOTES "overall impressions"
8	OT	listening	
9	AN	"Mom + Dad ... help each other" - AN	
10	OT	listening	
	VP	MOVES to front to see screen	
	VP	CE "yes!"	
	VP	CE "better"	
	CL	talking to neighbors	
	CL	gives ideas	
	PE	helps screen w/	
11	CL	talking to peer	
12	RV/PE	talking to peer	
	CL	no possibility of special class	
	CL	talks to peers	
	CL	peer talks to her	
	CL	she talks back	
13	VP	"A table"	
	VP	"Made it"	
	VP	"No" - CE	
14	PE	tells p. what others are doing	
	OT	makes table again	
15	OT	listening	
	IM	makes table again	

Observation Date: 9/30/04 School: Hogg Classroom Teacher: Ms. Dorate Grade: 4th Observer Initials: SD

Student Name: _____ 1st Segment or 2nd Segment Student ID#: 30184 Lesson Type: AP/ CL

Lesson Description: Ms. Dorate

Classroom Setup: Whole Class Big Groups Small Groups (6 or less) Individual Semester: Fall / Spring

FIGURE 8: Student observation form completed during an ArtsPartners infused classroom lesson (CL+AP).

Figure 9 displays the contrasting patterns of four major categories of learner behaviors (see Appendix B) for codes by categories in CL and CL+AP prelessons, across the entire population of Focus students. This demonstrates how the addition of ArtsPartners programming stimulates significantly more and different kinds of learner behaviors.

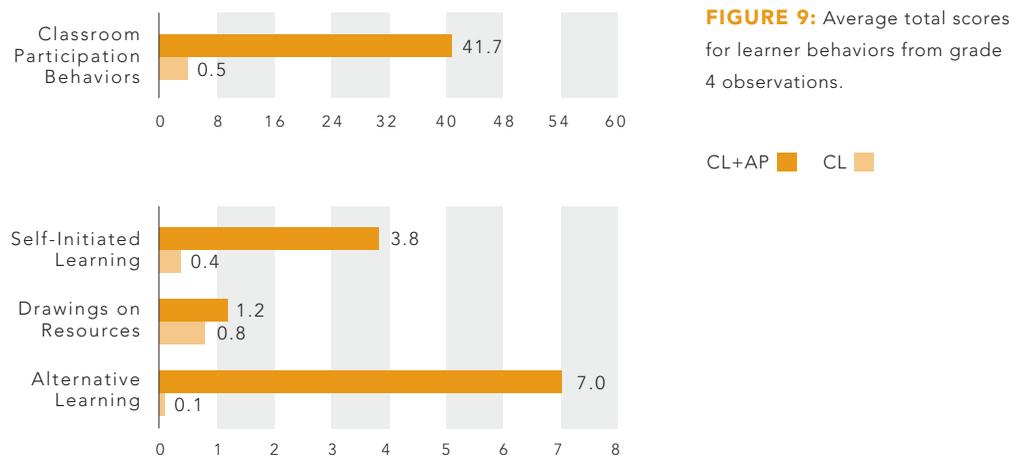


FIGURE 9: Average total scores for learner behaviors from grade 4 observations.

DESIGN PRINCIPLE 6: GRAPPLE WITH UNEVEN FINDINGS

Although no one would deny that data showing students' increased engagement in the learning process were "nice," the critical issue, for the program's longevity, was whether there was evidence of a relationship between arts and cultural learning and the skills that are at the heart of public education—skills such as reading and writing. Thus, it was imperative to look at measures of students' literacy achievement to ascertain whether the patterns of learner behaviors were linked to substantial differences in students' literacy achievement. These measures of literacy achievement included the following:

- STANDARDIZED READING SCORES.** State criterion-referenced tests were used to assess reading skills during the course of the study.²¹ These scores were used to assess yearly academic achievement as well as growth across the years of the study.²²
- STANDARDIZED WRITING SCORES.** The state criterion-referenced writing tests are given in fourth and seventh grade. They contain both multiple-choice items and a written composition, scored on focus and coherence, organization, development of ideas, voice, and conventions.²³ Scores were used periodically to assess writing achievement.
- CLASSROOM WRITING SAMPLES.** The TAKS test as a measure of writing skill is unlike much of the other writing that children do. Although the test is untimed, there is little time for thoughtful revision, none of which involves collaboration with a teacher or other students. Another measure of writing ability was needed to help researchers understand the effects of arts programming on students' daily writing activities. To complement the TAKS scores in composition, researchers also collected and scored a set of classroom (CL) and classroom and ArtsPartners (CL+AP) writing samples. These were scored using a six-trait writing scale developed by the Northwest Regional Educational Laboratory (NWREL). Using this assessment rubric, researchers scored literacy samples for six traits: Ideas/Content, Organization, Voice, Word Choice, Sentence Fluency, and Conventions with possible scores ranging from 1 to 5.²⁴ (see Appendix 5)

²¹ These were the Texas Assessment of Academic Skills (TAAS) Reading, given in 2001 and 2002, or Texas Assessment of Knowledge and Skills (TAKS), given beginning 2003 to the present.

²² Students' percentage correct of total possible items were computed for each testing period (one per year).

²³ Scores on the written composition range from 0 to 4, with scores of 2, 3, or 4 considered as "Pass." A 0 is given when the composition has a nonscorable response. A designation of "Pass" or "Fail" on the entire writing test is determined from the aggregate of the multiple-choice items and the written composition.

²⁴ A total score was also computed for analysis purposes.

ArtsPartners programs affected some, but not all aspects of literacy. These measures yielded a complex pattern of findings across semesters within single years and across the successive years of the evaluation. On the other hand, the quality and content of specific ArtsPartners curricula had major effects on whether student learning was enriched. Thus, throughout the evaluation process both researchers and program staff had to grapple with the hard—but very important—work of developing a much more nuanced understanding of what and when the programs contributed to students’ learning.

EVIDENCE FROM STANDARDIZED READING SCORES

The state criterion-referenced TAKS Reading test was used as one outcome measure of whether students’ literacy achievement was affected by their participation in ArtsPartners.

Grade 1 Cohort: TAKS Reading Scores

Although the grade 1 cohort began the study in 2001-2002, the students did not take the TAKS Reading test for the first time until they were in third grade (2004). In their results at third grade, Focus students averaged more correct answers (10 percent higher) than their Focus Grade peers. When they took the test again in 2005, Focus students performed better (78 percent correct answers) than their Focus Grade and Control Grade peers (71 percent correct).

In 2006 Focus and Control schools received ArtsPartners programming; however, no student received the intensive treatment (i.e., team-built lesson plans, observations, and interviews) that was part of the longitudinal study. During this first year after the study concluded, the grade 5 TAKS Reading administration showed that students from the Focus schools (Focus, 78 percent; Focus Grade, 77 percent) had a greater average percentage of answers correct than the Control Grade students (72 percent)(see Figure 10). This difference between the Focus and Focus Grade students and the Control Grade students was statistically significant.

Grade 4 Cohort: TAKS Reading Scores

There are six years of data for the grade 4 cohort, beginning with a “pretest” score in 2001. At that time, all groups were in grade 3 and scored similarly (between 75 and 79 percent). For the next five years, Focus students had an average percentage of correct answers higher than all other comparison groups, even though none

of the students received ArtsPartners programming of any kind after the 2004 administration (because they were in middle school). By the grade 8 TAKS Reading administration in 2006, Focus and Focus Grade students answered 91 percent and 88 percent of the answers correctly, respectively.

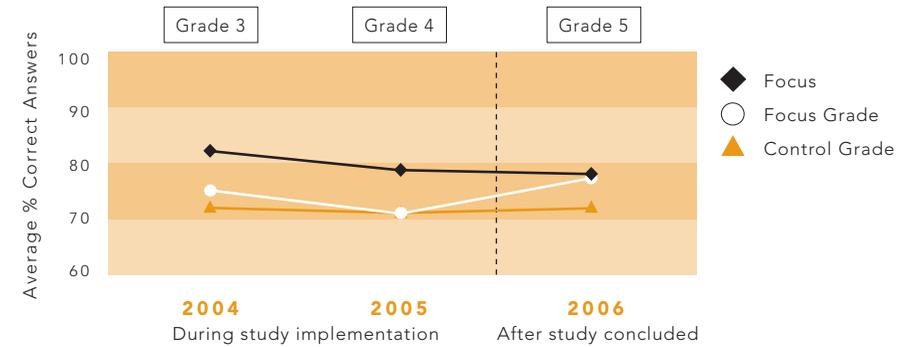


FIGURE 10: Average percentage of TAKS Reading questions answered correctly by grade 1 cohort comparison groups across three years.

Control Grade students began a steady increase in the percentage of answers correct in 2004 (65 percent), which continued through 2006 (87 percent), when their scores were almost as high as those of Focus Grade students. Although there were no statistically significant differences among the groups in 2005 and 2006, 13 Focus and Focus Grade students (9 percent) and 6 Control Grade students (4 percent) answered 100 percent of their grade 8 TAKS Reading questions correctly (see Figure 11).

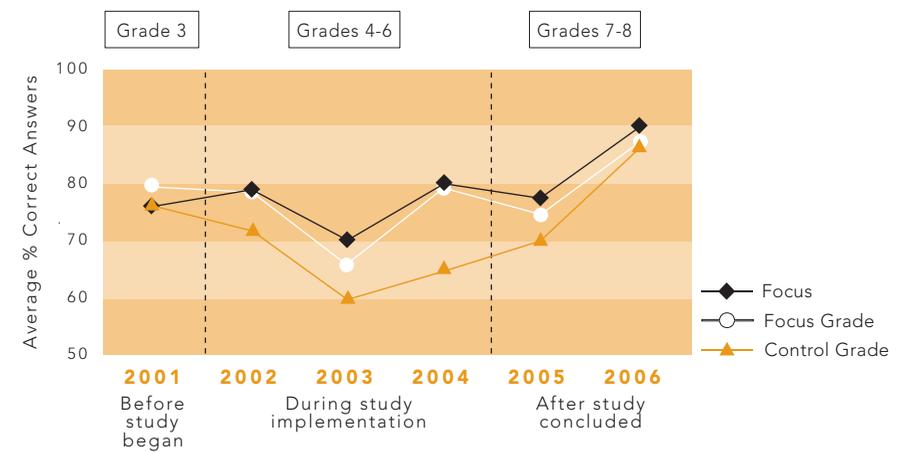


FIGURE 11: Average percentage TAKS Reading questions answered correctly for grade 4 cohort comparison groups across six years.

In the grade 1 cohort the research revealed that Focus and Focus Grade students maintained their advanced reading skills even after the more intensive treatment related to the study (team-designed curriculum, classroom observations, and interviews) was withdrawn. In fact, the Focus Grade students actually advanced their skills after the study concluded, with only the standard ArtsPartners treatment in place.

Even more encouraging were the data collected from the grade 4 cohort. In children’s school lives the transition from elementary to middle school is possibly the most difficult developmentally. Moreover, national data often show that student performance sinks from grades 4 to 8. Thus, the fact that the Focus and Focus Grade students were able to retain the skills they developed through ArtsPartners programming was well worth noting.

EVIDENCE FROM STANDARDIZED WRITING SCORES

Grade 1 Cohort: TAKS Writing Scores

When the grade 1 cohort was in the fourth grade, they took the TAKS Writing test for the first time. By spring 2005, students at the Focus schools (Focus, Focus Class, and Focus Grade) had experienced four years of CL+AP lessons—each including a writing component. Differences in written composition scores were statistically significant because a greater percentage of Focus students scored a 2 (45 percent) or a 3 (48 percent), while the other comparison groups had greater percentages scoring a 2 (65 percent to 67 percent)(see Figure 12).

The grade 4 writing prompt taken from the 2006 released TAKS test was: Write a composition about your favorite place to go. When this very simple prompt is compared with the writing assignment that was part of the History Reporters Lesson (see Appendix D), it is no wonder that 100 percent of the Focus students received a passing score on their composition.

When multiple-choice writing items were combined with the composition score for a Pass/Fail designation, 93 percent of Focus students passed the writing test, in comparison to the Focus Class (86 percent), Focus Grade (89 percent) and Control Grade (90 percent) students.

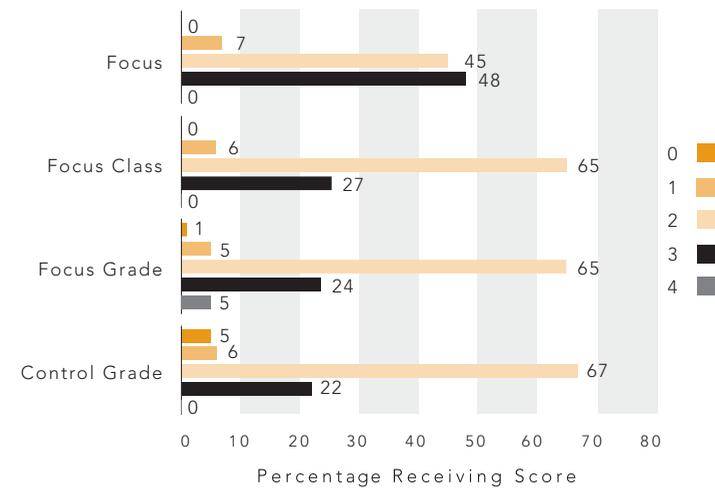


FIGURE 12: Percentage of original grade 1 cohort students receiving various grade 4 2005 TAKS Writing Composition scores.

Grade 4 Cohort: TAKS Writing Scores

When the grade 4 cohort was in the seventh grade, they took the TAKS Writing test for the second time. Now that these students were in middle school, they no longer received ArtsPartners programming. There were no statistically significant differences in written composition scores among the groups (see Figure 13).

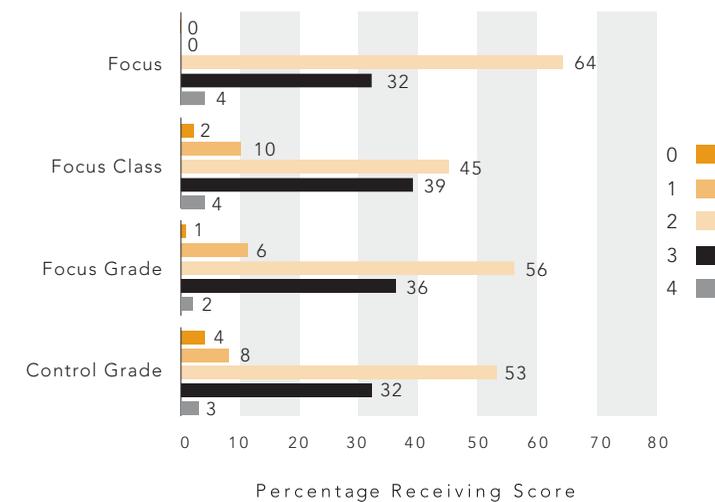


FIGURE 13: Percentage of original grade 4 cohort students receiving various grade 7 2005 TAKS Writing Composition scores.

Although there was little difference in students' composition score, a look at the overall passing rates tells a different story (see Figure 14). Even though Focus students had received no further ArtsPartners intervention during the 2004-2005 school year, a statistically significantly higher percentage (93 percent) passed as compared to the other student groups.

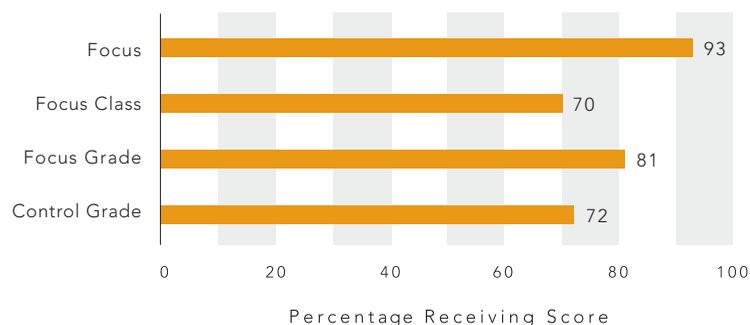


FIGURE 14: Percentage of original grade 4 cohort students passing the grade 7 2005 TAKS Writing test.

With the grade 1 cohort we see how homogeneous data can be when all students are taking the test for the first time. Even so, there were still some promising indicators that the ArtsPartners program might support skills students needed for this test—Focus, Focus Class, and Focus Grade students all outperformed the Control Grade students.

The data from the seventh-grade administration of the test to the grade 4 cohort continues to indicate encouraging linkages between ArtsPartners and writing skills. Although there was little difference in students' composition scores, Focus student, as well as Focus Grade students, outperformed Control Grade students on passing rates.

**EVIDENCE FROM CLASSROOM WRITING SAMPLES:
STRONG VS. WEAK ARTSPARTNERS INTEGRATION**

The NWREL 6+1 Trait scale for scoring student writing is more sensitive to the dimensions of literacy that ArtsPartners instruction might foster than the TAKS scoring rubric.²⁵ Using NWREL, trained evaluators scored literacy samples for six traits: Ideas/Content, Organization, Voice, Word Choice, Sentence Fluency, and Conventions, with possible scores ranging from 1 to 5. A total score was also computed. (see Appendix C for more information.) Additional differences were found when the Focus students' classroom (as opposed to test-based) writing samples were analyzed. Three of the six traits were consistently stronger in the CL+AP writing samples: Ideas/Content, Word Choice, and Voice. These were clearly areas in which ArtsPartners programs were adding value to classroom writing instruction.

The team of field researchers was comprised of elementary-level classroom teachers and past scorers of the TAKS writing section. This ensured that the literacy pieces were scored by people with an understanding of the average level at which students were performing on this type of lesson. They were trained in the NWREL system by two Big Thought staff members who received NWREL's 6+1 Trait writing certification at a regional workshop.

Because this was a longitudinal study, researchers could assess growth and differences in the six-trait ratings of students' writing samples over time and in the context of different types of ArtsPartners curricula. Figure 15 illustrates that improvement is not a simple straight line pointing upward. It also indicates that ArtsPartners' integrated curriculum did not always boost the level of student writing.

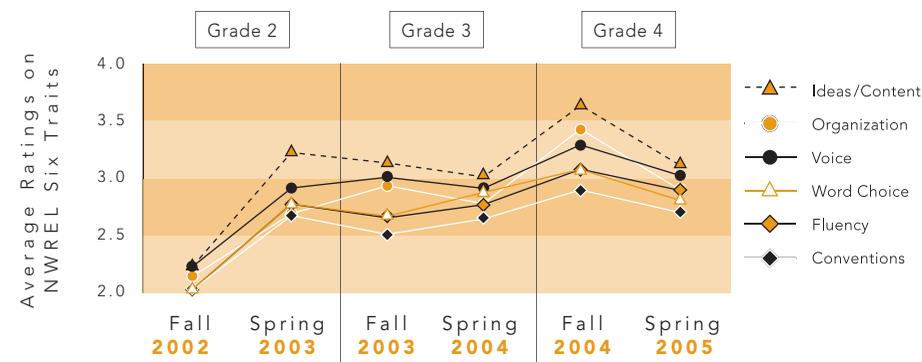


FIGURE 15: Trait ratings of CL+AP writing samples for grade 1 cohort Focus students by grade and semester.

²⁵ The 6+1 Trait writing assessment, produced by the Northwest Regional Educational Laboratory, can be found at www.nwrel.org/assessment/scoring.php?odelay=3&d=1.

However, Figure 15 does show considerable growth from the first grade 2 rating in the fall of 2002 to the second rating in the spring of 2003, then to the fifth rating in the fall of 2004. Differences in ratings for these three time periods were responsible for the majority of the statistical differences found in the overall changes in students' work.

Faced with the zigzag pattern of scores, researchers, cultural providers, and classroom teachers had to reevaluate the curricula that generated such dramatically different effects. This reflective thinking about what works, why it works, and in what context it works best was the fertile soil that produced an important realization. Carefully designed curriculum does not always equate to an effective curriculum. When an integrated curriculum fails to find and focus on the genuine intersection between classroom skills development and a cultural program's intrinsic value, it will fail to deliver strong outcomes. Likewise, if lessons are constructed to fit researchers' plans but do not make sense to the cultural partner and/or the teacher, then the implementation can fall flat and fail to provide added value.

Grade 1 Cohort: Weak ArtsPartners Integration

During the fall of 2002, grade 1 cohort students were in second grade. As part of their language arts studies, they explored the use of idioms as a writing convention. To support this study, teachers invited a storyteller to their rooms to work with the students and then took the students to a Dallas Children's Theater production of *Amelia Bedelia*, a children's book that recounts the adventures of a household helper who creates havoc by interpreting familiar idioms way too literally.

A subsequent comparison of the CL and CL+AP writing samples showed no significant difference on the six-trait scores. In fact, average ratings were almost the same for every trait. More humbling still, the ratings were higher for the CL than for the CL+AP samples (see Figure 16).

Appendix D contains a copy of the curriculum. Although it is well written, the lesson tries to serve too many masters. The genuine

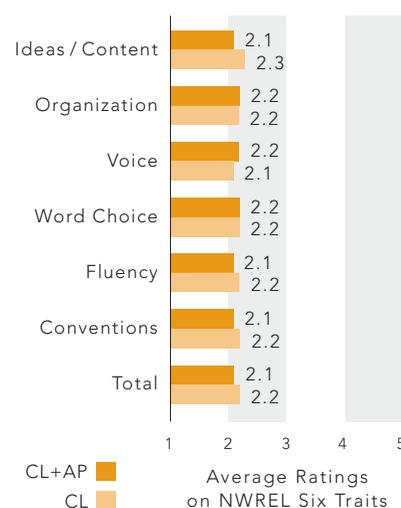


FIGURE 16: Trait ratings of CL and CL+AP writing samples for grade 2 Focus students, fall 2002.

intersection among theater, storytelling, and literacy development is missing.

This, however, did not diminish the positive impact of the play. One student with a learning disability who had struggled with the give and take of earlier interviews, was able to retell virtually the entire play—a milestone for her. At another school, a second-grader who was still depending on Spanish in her classroom and at home sent the researcher who worked at the school a simple thank-you card written entirely in English. Even though she did all her other assignments in Spanish, she chose English for this purpose, knowing it was the researcher's own first language.

The problem was with the learning module. Teachers and researchers saw possible connections that were not compelling to students. In addition, even the literacy concept was too narrowly conceived. It focuses on idioms rather than on the larger, more generative concept of figurative language. In hindsight, researchers and teachers have suggested that the theater production could have been put to much better use if it had been used to show students how a story comes to life on stage, how to develop a character, or how to write lively dialogue. But the findings demonstrate that the power and humor of the play were misspent. Focusing on the genuine intersection between classroom and culturally enriched learning is a delicate business. Connections cannot be concocted. For learning to occur there has to be a clear and meaningful synergy among students, teachers, and cultural providers alike.

Grade 1 Cohort: Strong ArtsPartners Integration

By contrast, in the fall of 2004 the ArtsPartners curriculum was a very effective complement to classroom learning. Students toured the Dallas Arboretum's *Texas Pioneer Adventure* and studied pioneer homes. Then, back in the classroom, a theater artist helped students notice, perform, and think about what it might have been like to live in those structures. Throughout this process, the students used a journal in which they recorded interesting details and specific vocabulary that could create pictures in their readers' minds of what life was like inside a covered wagon or a sod house.

The CL and CL+AP writing samples collected during this unit of study showed statistically significant differences in the ratings of Ideas/Content and Voice. In addition, the ratings in CL+AP settings were higher than those in CL settings for all traits except Conventions, where ratings were the same for both settings (see Figure 17). Appendix D contains a copy of the History Reporters lessons.

The effects of ArtsPartners experiences on student writing varied markedly from semester to semester, suggesting that some of the partnerships between classroom and cultural learning were considerably more effective than others. Field researchers, working with teachers, were very helpful in figuring out why this might have been the case. The findings, as well as the ensuing discussions, were important opportunities for staff and cultural providers to learn from results.

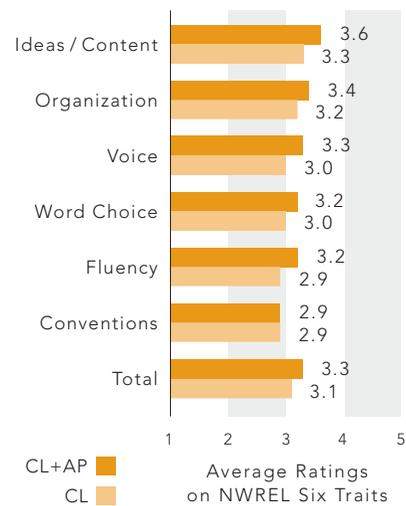


FIGURE 17: Trait ratings of CL and CL+AP writing samples for grade 4 Focus students, fall 2004.

DESIGN PRINCIPLE 7: STAY ALERT TO SURPRISES

As the study progressed, participants and funders asked, “Do the effects of ArtsPartners programs last, or do they ‘wither away’ when students are no longer enrolled in classrooms where the program is running?” This wondering became an additional research question when an outside funder asked Big Thought to provide the answer.

TAKS Reading and Writing scores were analyzed for the grade 4 cohort, who had matriculated into middle school where ArtsPartners was not part of the curriculum. Although there were not significant lasting effects for all students who had participated at the treatment schools, the Focus students—the subset of children who had been observed closely and interviewed twice annually—showed a sustained effect, remaining the highest-scoring students followed into middle school. This finding was not just a speck, but a pot of gold, opening up many new possibilities regarding the importance of student conversation with teachers or artists about their work (see Figure 11).

This is an example of the importance of staying alert to surprises in the data. Originally, the observations and interviews were not intended as one of the ingredients that might enhance student achievement. The purpose of these qualitative tools was to help field researchers and program staff understand when and how the program worked (or fell short of expectations). The discovery about Focus students was totally unanticipated. Although this clearly requires more thought and study, the finding suggests that some combination of feeling special,

forming a relationship with interviewers, and having the opportunity to reflect aloud with a skilled questioner makes a significant difference to students as learners.

Although this “pot of gold” will take more investigation to understand, it may point to a whole new dimension for Big Thought’s programmatic work with ArtsPartners and beyond. It raises a number of questions, for example: If this kind of discussion between researchers and students were to become a regular part of programming, would a program’s effect on the students be enhanced? Plowing such surprises back into practice is vital. For instance, it is now up to Big Thought’s staff to consider whether they want to make interviewing a more regular part of their work. Is this feasible? Are the gains worth the effort? Are there some populations of students for whom it might be especially important?

DESIGN PRINCIPLE 8: SHARE AND USE THE FINDINGS

Although ideas and learning were being shared informally among the researchers’ networks and colleagues, formal opportunities to learn from and discuss the study were also created. Feedback and reflection forums were created with multiple groups of people attending in the hope that stakeholders representing different perspectives could learn from the questions and discussions raised in such a mixed setting. In the course of conducting these forums, researchers realized that each stakeholder group wanted more information specific to its distinct viewpoint.

Stakeholders’ basic questions were the same: What did you learn? And, What do we do next? However, each stakeholder group was asking from a particular point of view. Investors and policy makers wanted to know how the program worked and whether support should continue. Program developers, artists, and cultural providers wanted to know which programmatic pieces were most successful, which needed to be developed or refined for greater effect, and what unforeseen challenges might be limiting the program’s impact. Cultural partners, teachers, and principals wanted to understand how their work was reflected in the student outcomes and how ArtsPartners could add value to their work.

Thus, additional forums were designed for specific audiences and focused on their concerns. In presentations before the city council and roundtable discussions at schools, researchers shared what they were learning and led discussions about how stakeholders would use the information. To have its full effect, an evaluation needs to be digested and discussed. The end of the process is not the day the final document is delivered; it is the day the findings have informed the most basic ways of doing business.



Part IV

EVALUATIONS THAT BUILD CAPACITY: LESSONS LEARNED

Many specifics of the ArtsPartners evaluation were unique to a given time, place, and group of organizations. But other aspects have wider implications and point to certain lessons from which much can be learned about using an evaluation to build capacity. These lessons begin with the design principles that both guided and arose from the ArtsPartners evaluation and their implications for sponsoring organizations and funders.

DESIGN PRINCIPLES FOR EVALUATIONS THAT DO MORE THAN MEASURE

The first responsibility of an evaluation is to take stock of the results of a program. However, the processes of taking stock of results and tracing those results (or lack thereof) to their sources can also build the capacity of a network of stakeholders. Conducting an evaluation in such a social and inquiring way can improve the program, the skills of participants, and the ways in which they are able to work together.

As suggested throughout the text, the ArtsPartners evaluation yields a set of principles for organizations and communities that seek to build capacity as well as measure results. These principles, discussed throughout this document, are as follows:

1. **TAILOR THE EVALUATION TO THE CONTEXT** of the program, the community, and the debates that prompted the call for an assessment. Strong results are results that speak to the needs and the questions of those who seek the evaluation. Only when evaluation results match the needs and questions of consumers will those consumers understand, value, and use the findings.

I noticed that the students learned a lot of new words. It broadened their vocabulary. It enriched their experience. It has helped them to be more expressive and more creative.

—Ms. Giles, THIRD-GRADE TEACHER AT MARSALIS ELEMENTARY SCHOOL

2. **CREATE COMMUNITY-WIDE INVESTMENT.** Building a network of partnerships among all stakeholders creates community-wide investment. Creating capacity in the organization, participants, funders, and evaluators produces a climate or culture of respect, in which the ongoing improvement of programs is the norm.
3. **ENGAGE STAKEHOLDERS IN KEY DECISIONS EARLY.** This process should include discussions of the purposes, the scale (time, dollars, and the amount of human effort that will be involved), and the design of the evaluation, as well as the kinds of claims for the program that the evaluation will—and won't—support.
4. **ENHANCE THE CAPACITY OF ALL PARTICIPANTS.** In different ways and to varying degrees, an evaluation is an opportunity for all participants to learn and improve the quality of their work together. This learning can happen as they decide on the goals and strategies of the evaluation, and on their definitions of quality. In the complex and difficult decision-making process of creating and undertaking an evaluation, perspectives are often stretched and ideas become tight and focused. The broader understanding that results strengthens the work that all stakeholders contribute individually and collectively to the community.
5. **PLAN FOR MIDCOURSE CORRECTIONS.** The process of data collection and analysis often uncovers some poor choices or aspects of the evaluation that need redirection. That process, if conducted publicly and candidly, can strengthen the evaluation and build the sense that it is a genuine inquiry open to change and input.
6. **GRAPPLE WITH UNEVEN FINDINGS.** The process of analyzing the data is likely to reveal a complex story. Some expected findings may not show up, or they may show up unevenly. Some hoped-for effects may not be statistically significant. Results may occur for some groups of participants but not for others. Although disappointing, uneven findings can be very instructive. They are an invitation to ask, Why are the results not what we expected? What does this say about the program?
7. **STAY ALERT TO SURPRISES.** The process of data analysis can also turn up surprises. These surprises may point to important new directions for the program or for the evaluation, or the need to change some deeply held beliefs about a program and its effects. Not all findings will be positive. Some may indicate deficiencies in the program design or less-than-impressive outcomes. Sometimes results look good, but when analyzed

statistically, no significant differences are found. In every case, specks of gold can usually be extracted from the data to improve the program or point to new directions for research.

8. **SHARE AND USE THE FINDINGS.** For many, the payoff of the evaluation begins the day the findings are delivered. But a long technical report—even if it is overwhelmingly positive—will not necessarily win many friends. The information has to be distilled and presented in a form that is understandable and compelling, and in a number of different formats. At the same time, organizations must resist efforts to distort the work, oversimplify, or misstate the conclusions even when there is pressure to do so. Finally, scheduling when, where, and to whom the findings will be presented is a critical component of the initial planning.

Likewise, it is important to plan how findings will be woven into the evolution of the program design. For an evaluation to have its full effect, it needs to be digested and discussed. The end of the process is not the day the final document is delivered; it is the day the findings have informed the most basic ways of doing business.

These very broad principles will play out in different ways for different participants in the evaluation process. Additional lessons—at least the major ones—for organizations, evaluators, and funders are listed in the following sections.

LESSONS FOR ORGANIZATIONS PLANNING AN EVALUATION

1. **IF YOU SEEK MORE THAN A “SCORE” FROM THE EVALUATION, BE CLEAR ABOUT THE OTHER OUTCOMES FROM THE START.** An evaluation can do more than assess the quality of a program. As the ArtsPartners example shows, an evaluation can build organizational, human, and community capacity—but only if it is designed from the outset to do so. And only if all the stakeholders endorse working in this way.
2. **FOCUS PROGRAMS SO THAT THEY CAN BE EVALUATED.** Too often organizations put evaluators into a near impossible position. If the program goals are fuzzy, data have been collected haphazardly or not at all, and expectations for program outcomes are totally out of scale with the activities and intervention, it is unrealistic to expect an effective or persuasive evaluation.

ArtsPartners began with a broad (some would say loose) proposition that arts integration promotes learning. Before the evaluation could be useful, that broad proposition had to be focused and tightened to become: “When cultural learning experiences are effectively designed and delivered

in a planned for and sustained way, elementary school children can gain important understandings about the expression of ideas and experiences that will inform their reading and writing capacities.”

3. **KEEP YOUR EXPECTATIONS REALISTIC.** The most common mistake organizations make is to expect too much from their evaluations. Often this is an outgrowth of promising too much to funders and to other audiences who are looking for unrealistic results in a very short time frame. Particularly in the complex world of urban schools (where factors ranging from public health, the lack of affordable housing, or good jobs for parents affect how well children learn), it can be extremely difficult to isolate and claim substantial effect, particularly for programs in which the content and the approaches vary from artist to artist or from site to site.
4. **INCLUDE EVALUATION IN YOUR INITIAL PROGRAM PLANNING AND BUDGET.** The best evaluations are those that are planned from the earliest stages of initial program design. Knowing that a program will be evaluated sharpens the thinking about goals, desired outcomes, and activities. Similarly, because it is often difficult to raise money for evaluations, incorporating a line item for evaluation directly into the program budget will help ensure that an organization does not have to patch together adequate (or even inadequate) funding later.
5. **FIND AN EVALUATOR WHO SHARES YOUR GOALS AND ASPIRATIONS FOR THE EVALUATION BUT CAN MAINTAIN OBJECTIVITY AND INDEPENDENCE.** Choosing an evaluator solely on the basis of reputation is not enough. The individual (or team) should have a demonstrable track record of work that is consistent with the expectations you have established. Some evaluators may favor one approach, and others another. Finding the right match is crucial. Also important is finding someone who will be able to maintain objectivity throughout, especially when the work is controversial and the various interest groups are vocal.
6. **REMEMBER YOUR ORGANIZATION’S ROLE AND RESPONSIBILITIES.** It is definitely not the case that once you hire an evaluator, your job is done. In fact, a good evaluator will probably substantially increase your workload and that of your organization. Information must be gathered, meetings must be organized, questions must be answered, documents must be distributed, and much more. This is one reason it is so important to get mutual expectations spelled out right from the start (before a contract is signed) and to be sure you can hold up your end of the bargain.

7. **INVOLVE AS MANY PEOPLE AS POSSIBLE, INCLUDING THOSE WHOSE ACCEPTANCE OF THE RESULTS ARE IMPORTANT.** Organizations often establish a group to oversee the planning and even the oversight of an evaluation. The group should include key staff, other stakeholders, and, when possible, those whose acceptance of the findings will be critical. Sometimes it is not possible to include a specific individual with a stake in the results of the evaluation (e.g., a city council member or a national funder) on the evaluation committee. In this case, the committee would do well to seek out that individual and brief him or her on the process, the evaluator, and other details. Doing so can often lead to greater interest in and acceptance of the results.
8. **BUILD ONGOING RESEARCH AND EVALUATION CAPACITY WITHIN YOUR ORGANIZATION.** As many organizations are learning today, research and evaluation are increasingly becoming part of the cost of doing business in a world that is hungry for accountability. Thus, building expertise and capacity in your organization is simply a good business decision. By collaborating with external evaluators, organizations can create meaningful mentorship opportunities for staff who can then carry out day-to-day evaluation activities. In addition to saving money by pairing local and national evaluators, this collaboration also infuses more local knowledge and understanding into external evaluators’ work.

LESSONS FOR EVALUATORS

1. **CREATE A CLIMATE OF MUTUAL RESPECT.** When a community invests in a large-scale and very public evaluation, respect on the part of the evaluator should be the cornerstone of quality work. Presume good intentions. For instance, in the ArtsPartners evaluation it was important to help stakeholders balance what could have been competing interests in academic and artistic outcomes. Also, in setting up the comparisons of CL and CL+AP lessons, it was important not to pit teachers and cultural providers against each other by asking who did a better job. Instead, the question was framed in terms of the complementarity of these two types of experiences.
2. **INCLUDE TIME AND STAFF FOR CAPACITY BUILDING IN BUDGETING AND PLANNING.** An evaluation designed to build organizational and community capacity is not a speedy process. To build that capacity, evaluators need time for training, discussion, practice of and retraining of stakeholders. Some of that time may be needed for explanations and instruction on

evaluation and statistical analysis procedures. Moreover, this is not a one-time investment. Community-based research teams require constant rebuilding as people's lives shift and change. The payoff is great when you create a way of doing business in which people are continually asking, "Why?" and "How well?" However, time is money and it is vital to be clear that this approach comes with associated costs both in dollars and in staff time for the sponsoring organizations.

3. **PLAN FOR MIDCOURSE CORRECTIONS AND UNEVEN FINDINGS AS PART OF THE PROCESS FROM THE START.** If an evaluation is going to be carried out jointly, with ongoing participation by stakeholders, everyone needs to understand the nature of the inquiry from the start. Ongoing evaluation of small segments of data throughout the longitudinal study allowed program and evaluation designers to make meaningful midcourse corrections that were not envisioned at the outset. When stakeholders were disappointed at the uneven results of observations of students' writing, teachers and researchers concluded that the real differences in student behavior were seen when arts integration was occurring, not during writing time. Analyses of this unplanned setting helped evaluators more finely assess just where in the educational process arts integration was having the greatest effect.
4. **DIG FOR DEEPER MEANINGS TO UNEXPECTED FINDINGS AND SEEK OUT SURPRISES AS THE EVALUATION PROCEEDS.** It should be part of the evaluator's job to dig for deeper meanings when findings are not what were expected and to lead stakeholders to look at the findings in other ways. As the evaluation proceeds, evaluators can look ahead and suggest further questions that could be answered that may bring about unexpected findings. In the ArtsPartners evaluation, Focus student interviews were a strategy designed to help researchers understand where and why the ArtsPartners curriculum made a difference. However, later unplanned analysis showed that Focus students continued to outperform their peers on large-scale measures of achievement as much as two years after the study ended. It turns out that the interviewing strategy may have been one of the most powerful treatments in the ArtsPartners study. As a result, program designers began to consider how one-on-one discussion of a student's work with an adult could be built into the program on a large-scale design.
5. **BE RESPECTFUL OF A COMMUNITY'S RESOURCES.** Big Thought invested considerable resources in following the Focus students. The idea was to be able to understand—in a deep way—where and why ArtsPartners worked. As

a part of that effort, researchers developed a number of exploratory parts to the interview—only some of which yielded findings. Perhaps a better use of the available resources would have been to restrict the work with Focus students to a few well-established measures. This would have allowed for larger numbers of Focus students to be followed. With larger numbers enrolled, researchers might have been able to develop a more detailed picture of how ArtsPartners effects take hold.

LESSONS FOR FUNDERS

1. **BE CLEAR ABOUT WHY YOU ARE REQUESTING AN EVALUATION.** Some funders require an evaluation to demonstrate that the programs they fund are accountable. The results are received, looked at in a cursory way, and filed. In other cases, funders believe that only the organization should learn from evaluations; thus they do not pay attention to the results. In still other cases, future funding decisions hang in the balance. Being clear with grantees is the best policy.
2. **LOOK FOR CAPACITY-BUILDING BENEFITS FROM THE EVALUATION.** Throughout this story of the ArtsPartners evaluation, there have been countless examples of how the program, the organization, and the community benefited in ways that went well beyond the evaluation itself. Funders should encourage such capacity-building opportunities. In some cases, this could even extend to helping organizations build a more permanent research and evaluation capacity internally.
3. **USE THE "LABORATORY" RATHER THAN THE "REPORT CARD" METAPHOR.** Grantees' fears of receiving negative findings from an evaluation can be palpable if funders imply that they are expecting a "report card" on the quality of the programs. Funders need to repeatedly stress that their philanthropic work and the work of their partner organizations is like a laboratory in which some experiments will succeed and others will fail. One can learn from both the positive and the negative findings.
4. **HELP ORGANIZATIONS UNDERSTAND THE ROLE OF OUTSIDE EVALUATION.** Organizations can do much themselves to assess their work. They can design surveys, collect data, conduct interviews, and so forth. Ultimately, however, someone from the outside needs to look at this information objectively and independently and decide what it means. Unfortunately, many organizations, their staff, and their sites experience an evaluation as an audit. They put

their best foot forward and hide their difficulties, with the result that the evaluation outcomes are only partially meaningful.

5. **BE REALISTIC IN WHAT YOU ASK FOR.** Many funders expect far too much from an organization conducting an evaluation. Those expectations may be out of scale with the capacity of the organization, the budget, or the scope of the program. On the other hand, organizations themselves will often overpromise. A focused funding partner will nip such promises in the bud before they cause trouble.
6. **DON'T EXPECT TOO MUCH TOO SOON.** Funders often hope that major change can happen in short time cycles, after minimal interventions and small monetary investments. Yet change generally takes significant time, effort, and money. It often requires repeated treatments at frequent intervals. The common funding cycle of three years is often too short to see meaningful results in almost any program. Looking to an evaluation to find those results can be unrealistic and harmful. In the case of ArtsPartners, some of the most important results were not clear until two years after the study had officially ended.



7. **IF YOU WANT A GOOD EVALUATION, ESPECIALLY ONE THAT BUILDS THE CAPACITY OF AN ORGANIZATION OR A NETWORK OF STAKEHOLDERS, BE WILLING TO PAY FOR IT.** It is not unusual for an in-depth evaluation to cost between 5 and 15 percent of a program's total budget—particularly once staff time is taken into account. Yet, it is the rare funder that will pay for an in-depth evaluation. If a funder is going to invest in a program, it should also plan to invest in high-quality evaluation. Without such investments, it is hard to see how to build a valid understanding of what works. The longitudinal evaluation of ArtsPartners cost approximately \$250,000 a year, when everything was accounted for. To make this possible, a number of funders had to agree to share in its costs.
8. **PROVIDE TECHNICAL ASSISTANCE.** For many organizations, a major evaluation can be a turning point in their organizational growth. This is particularly true as organizations expand their scale or mission. But many small to mid-sized organizations may never have participated in more than a cursory evaluation. The staff may be inexperienced and not understand the potential for building capacity at the program or organizational level. They may not know what they are looking for in an evaluator, how to write a request for proposal, or how to interview potential evaluators. If funders expect good evaluations, they should provide the necessary technical assistance to their grantees, either singly or in groups.
9. **DISSEMINATE THE FINDINGS.** Increasingly, funders are treating evaluations as important deliverables—not just as devices for internal monitoring and accountability. Understanding the conditions under which programs have a positive impact is difficult work given the complexities of the real world of classrooms, schools, and districts. It will take more than solo missions to build the knowledge needed. When evaluations contain important lessons about how to design, implement, and measure effects, they should be shared. Foundation annual reports, website or special publications such as this one are wonderful vehicles for dissemination and should be regarded as valuable learning tools for the field.

Part V

APPENDICES

APPENDIX A: LEARNER BEHAVIORS CODES

	VERBAL BEHAVIORS		ACTIONS/GESTURES
QT	Asks substantive question of teacher or artist <ul style="list-style-type: none"> Asks a content question (e.g., Did the Alamo burn down?) Asks a quality question (e.g., Is my ending good?) 	SK	Seeks out other work to look at <ul style="list-style-type: none"> Uses peer or teacher work as example Makes effort to get more or different materials (e.g., looks up at overhead, bulletin board; reaches into desk/backpack; gets up to get book, reference papers)
QP	Asks substantive question of peer <ul style="list-style-type: none"> Asks a content question (e.g., Did the Alamo burn down?) Asks a quality question (e.g., Is my ending good?) 	O	Closely observes what another does or makes <ul style="list-style-type: none"> Looks for a full two seconds
SW	Shows/demonstrates work, explaining it <ul style="list-style-type: none"> Shares or reads paper in class Presents work (e.g., plays a song created or acts out a skit developed) 	IM	Imitates as a way of learning <ul style="list-style-type: none"> Makes movement based on another's movement, such as an artist or peer (e.g., watches an actor stand like a warrior, then poses the same way)
RF	Reflects on work <ul style="list-style-type: none"> Talks about what he or she is doing or working on Demonstrates meaningful self-talk (e.g., I'm gonna do my plan, then write about my favorite animal) 	RW	Reviews work <ul style="list-style-type: none"> Stops to look at, ponder, read over work
EV	Evaluates own or others' work <ul style="list-style-type: none"> Makes value judgments (e.g., "I/That was good/bad") 	RV	Revises work <ul style="list-style-type: none"> Erases, starts over, edits work (e.g., adds a period when reading over first draft of story)
AP	Asks to participate, takes a turn, answers (Note: Nonverbal behavior is RHI.)	H	Helps another student
AH	Asks for help <ul style="list-style-type: none"> Asks how to do something, to be shown or taught (e.g., What should I write next? How do you spell <i>Alamo</i>?) (Note: AH behavior has a short duration; for longer interactions, consider CL.) 	PE	Plays or experiments <ul style="list-style-type: none"> Expresses innovation, generates new ideas, extends past current information Can also be a verbal behavior
AN	Offers a unique answer in response to a question <ul style="list-style-type: none"> Gets another AN for every unique answer Gets an AN even if the teacher doesn't hear or respond (as long as it is still an on-topic answer to a question) (Note: Distinction between AN and VP is that ANs tend to be answers not widely shared or known, thus unique.) 	RHR	Raises hand to answer question <ul style="list-style-type: none"> Makes a physical gesture (each time a hand goes up, it gets coded, even if the student is just switching hands) If student keeps hand up across two-minute interval, he or she gets second code of RHR

FIGURE 18: Codes for learner behaviors (continued on next page)



	VERBAL BEHAVIORS		ACTIONS/GESTURES
VP	Verbal participation <ul style="list-style-type: none"> • Shares widely known or expected answer chorally with many classmates; also chimes in with an idea • May be an action also 	RHI	Raises hand to initiate question <ul style="list-style-type: none"> • Makes a physical gesture (each time a hand goes up it gets coded, even if the student is just switching hands) • If students keeps hand up across two-minute interval, he or she gets second code of RHI
CL	Collaborates with another student or students <ul style="list-style-type: none"> • Shares ideas in either or both directions • Usually thought of as “peer-to-peer feedback” 	OT	Performs on-task behavior <ul style="list-style-type: none"> • Works diligently on the assignment
CR	Clarifies own response <ul style="list-style-type: none"> • Revises own answer so it is more correct, better worded (e.g., when asked to explain what he or she has said, NOT A NEW QUESTION) (Note: Clarifying another students’ answer is AN.) 		
WITH MUSICIANS			
CL	Collaborates <ul style="list-style-type: none"> • New CL for each stop or start or new element added (e.g., new instrument) 	PE	Plays or experiments <ul style="list-style-type: none"> • Expresses innovation, generates new ideas, extends past current information • Can also be a verbal behavior
VP	Verbal participation <ul style="list-style-type: none"> • One code per verse sung (stop/start) EXCEPT call-and-response songs (each time student responds is a new code) • Counts beats for drummer 	IM	Imitates as a way of learning <ul style="list-style-type: none"> • Imitates rhythm by saying it aloud or showing it on hands, kinesthetically
SW	Shows/demonstrates work, explaining it <ul style="list-style-type: none"> • Presents work (e.g., plays a song created) 		
WITH THEATER ARTISTS			
VP/IM	Verbal participation AND imitates as a way of learning <ul style="list-style-type: none"> • Makes a movement with the rest of the class to imitate the actor (e.g., says “mind” and points fingers at head) 	IM	Imitates as a way of learning <ul style="list-style-type: none"> • Makes movement based on an actor’s or another student’s movement (first time only, then OT)
RF	Reflects on work <ul style="list-style-type: none"> • Says what he or she is doing while doing it (e.g., makes or acts out a scraping motion while saying “I’m scraping”) 	OT	Performs on-task behavior <ul style="list-style-type: none"> • Performs any movement an actor specifically prescribes or any repetitions of the same movement
EV	Evaluates own actions <ul style="list-style-type: none"> • For example, finishes acting a scene, bows, and says out loud, “They loved me!” 	PE	Plays or experiments <ul style="list-style-type: none"> • Any movement or action the student generates on his or her own (first time only, then OT)
CL	Collaborates with another student or students <ul style="list-style-type: none"> • Works or talks meaningfully with a peer 	RV	Revises work <ul style="list-style-type: none"> • Edits a position or acts to better it (e.g., tries to be a tepee by extending arms, then decides it would be better if arms were bent) • Action must be different and meaningful
SW	Shows/demonstrates work, explaining it <ul style="list-style-type: none"> • Presents work (e.g., acts out a skit developed) 		

APPENDIX B: LEARNER BEHAVIOR CATEGORIES

Four overarching learner behavior categories (distinguished by capital letters) were created from 21 observed and coded individual learner behaviors: Classroom Participation Behaviors, Self-Initiated Learning, Drawing on Resources, and Using Alternate Learning. The individual learner behaviors that contributed to each learner behavior category are listed in Figure 19.

CATEGORIES	INDIVIDUAL LEARNER BEHAVIORS
Classroom Participation Behavior	Reviews own work Revises own work Answers substantive question Raises hand to answer a question Clarifies own response Participates verbally
Self-Initiated Learning	Shows work Reflects on own work verbally Asks to participate Evaluates own or another’s work Collaborates with others
Drawing on Resources	Questions teacher Questions peers Helps another student Asks for help (verbal) Seeks help (nonverbal) Seeks out another’s work or gets more materials Closely observes another Raises hand to ask a question
Using Alternate Learning	Imitates to learn Plays or experiments Chooses among options

FIGURE 19: Learner behavior categories.

APPENDIX C:
NWREL 6+1 TRAIT WRITING ASSESSMENT

Trained evaluators scored literacy samples using the Northwest Regional Educational Laboratory's (NWREL) 6+1 Trait writing assessment. The six traits are Ideas/Content, Organization, Voice, Word Choice, Sentence Fluency, and Conventions (see Figure 20), with possible scores ranging from 1 to 5 (see Figure 21). A total score is also computed.

TRAIT	DESCRIPTION
Ideas/Content	Ideas/content are the purpose, the theme, the main idea, and the important and interesting details that support it. The topic should be neither too broad nor too narrow, and the message should be clear.
Organization	The organization is the internal structure of the composition, the way the ideas are developed. A well-organized composition has a meaningful beginning; the topic is developed logically; transitions are smooth; and the conclusion brings a sense of resolution.
Voice	When a composition has voice, we feel the writer's unique personality coming through the written words, and we feel a personal connection to the writer. Voice is the heart and soul of a composition.
Word Choice	Word choice is the use of rich, vivid, precise language that not only informs but also moves the reader. Good word choice is characterized not so much by the use of "big" words as by the skillful use of simple words to leave a more lasting impression with the reader.
Sentence Fluency	Sentence fluency is the rhythm and flow of the composition. The length and structure of sentences are varied so that they read smoothly and enhance the interest of the reader. A composition with sentence fluency is pleasing when read aloud.
Conventions	Conventions are the mechanics of language—spelling, punctuation, capitalization, grammar, and usage. A composition with good conventions has been proofread and edited with care.

Source: Taken from "The Write Direction," Dallas ISD writing curriculum, August 2002.

FIGURE 20: The six traits of writing.

SCORE	RATING	DESCRIPTION
1	Not yet	A bare beginning; writer not yet showing any control
2	Emerging	Need for revision outweighs strengths; isolated moments hint at what the writer has in mind
3	Developing	Strengths and need for revision are about equal; about halfway home
4	Effective	On balance, the strengths outweigh the weaknesses; a small amount of revision is needed
5	Strong	Shows control and skill in this trait; many strengths present

Source: Northwest Regional Educational Laboratory website (www.nwrel.org).

FIGURE 21: 6+1 Trait writing scoring continuum.²⁶

²⁶ The Dallas Independent School District uses the 6+1 Trait scoring system as part of its developmental writing curriculum, The Write Direction. Although it was not in use until 2002, many teachers had been exposed to it through reading and language arts staff development efforts.

APPENDIX D: SAMPLE ARTSPARTNERS INTEGRATED CURRICULUM



ArtsPartners

Idioms Lesson Plan

Dallas Children's Theater—
Amelia Bedelia

Young Audiences of North
Texas—Dan Gibson, Storyteller
Second Grade, Fall 2002

Big Idea: Students will develop an understanding and enjoyment of nonliteral (figurative) language, exaggeration, and humor. Then, they will be able to use these forms of language to enhance their own writing and communication. Finally, students will discuss the costs and limits of only being able to think and speak literally, as well as the benefits and pleasures of being able to communicate both literally and figuratively.

Day 1

Objective: The students will be able to identify and understand the usage of idioms.

Materials:

1. Chart paper and marker
2. List of example idioms (ArtsPartners will supply)
3. *Teach Us, Amelia Bedelia* book (ArtsPartners will supply)

Questions:

1. What is an idiom?
2. What are idioms used for?
3. Can you think of an idiom you've heard?

Activity 1 (10 minutes): Ahead of time, write the following idiom on chart paper: raining cats and dogs. Read the idiom aloud with your students. Ask them whether it can really rain cats and dogs.

Ask students to close their eyes and imagine what "raining cats and dogs" would look like. Let students predict what the idiom means and share their predictions. Explain that the phrase "raining cats and dogs" is an idiom that generally means that it is raining very hard. Explain that an idiom is an expression that cannot be understood from the individual meanings of its words. It is a phrase that carries a colorful, sometimes exaggerated meaning that helps a listener understand what the speaker means by using humor, images, and exaggeration.

Activity 2 (15 minutes): Give more examples of idioms from the list provided, and ask the students if they can tell what each idiom means (in the conventional, rather than literal, sense). (*Teacher note:* Idioms are figures of speech that have nonliteral meaning. Thus the challenge is to alert students to the humor and beauty of idioms, and their imaginative quality.) Ask the students to give their own examples of idioms. Students can offer ones that they have heard, and then create their own idioms. Ask what makes an effective/good idiom. Be sure to discuss humor, common words used in new ways, words that make a funny picture in the mind's eye, and so on. Explain that an idiom is a creative use of words that people share as "insider" language. It's a fun and very expressive way to talk!

Activity 3 (45-55 minutes): Introduce the book *Teach Us, Amelia Bedelia*. Explain how the main character, Amelia Bedelia, does not understand idioms and interprets them literally. The result is that she does exactly what the words in the idiom say to do. Also discuss why this makes her such a funny and amusing character. (*Teacher note:* Begin to read the book on day 1 and continue to read on day 2.)

Day 2

Objective: The students will be able to identify idioms, understand their usage, and predict their meanings. They will understand the costs and limits of thinking literally.

Activity 4 (25-35 minutes): Continue to read the book, *Teach Us, Amelia Bedelia*. When you are halfway through the book, stop and ask students to point out any idioms that they have heard so far. (If the students have any difficulties, go back and point out the idioms to the students and talk about how Amelia interpreted them and what they were intended to mean.) Emphasize that idioms are unique and creative ways to say something very basic, so they make our writing and our talking more interesting and imaginative. Finish the book, repeating the same process of identifying the idioms in the second half as needed.

Activity 5 (20 minutes): Discuss *Teach Us, Amelia Bedelia* with the students. Ask for reactions and questions from the students concerning the book and the idioms. Then tell the students that they will be attending a theatrical performance starring Amelia Bedelia! Remind students of the type of character Amelia Bedelia is, and ask

them what they expect to see at the play.

Write down their expectations on chart paper. This will be the foundation for a "What do you know? What do you want to know? What did you learn?" (KWL) chart that you will use throughout the lesson cycle. Tell students to pay close attention to the performance so that they can discuss it when they return to class. Finally, discuss appropriate theater behavior with the students (i.e., clapping at the end of the show, being good listeners, and not responding to the characters on stage unless they are prompted).

Day 3

Objective: The students will be able to identify idioms, understand their usage, and predict their meanings as they hear them in a spoken (theater) format. They will understand the costs and limits of thinking literally as they watch and listen to a character who can only use and understand literal language.

ArtsPartners Program
Amelia Bedelia
Presented by the
Dallas Children's Theater
(90 minutes)

Amelia Bedelia, the charming but literal-minded housekeeper, does everything you say exactly the way you say it. She merrily upsets the household by "dressing the chicken"—in clothes! Or "changing the towels"—by cutting them into little pieces! Amelia Bedelia's well-meaning efforts leave the audience laughing, her employers fuming, and everyone else very, very confused!

Activity 6 (15 minutes): Discuss the Amelia Bedelia performance with the students. Refer to the work you have done on idioms and ask the students what they learned from the theater performance.

Then, tell students that Dan Gibson, a storyteller, will come to visit their class. Discuss the idea of how people use words and their voices to conjure images, make mental pictures, and create a mood by asking the following:

- What are some ways people use their voice in telling stories?
- What qualities make someone good at telling stories?

Explain that Gibson will use language skills to help the students create their own stories. Discuss some things they may see when the storyteller visits. How might this experience be the same or different from the Amelia Bedelia performance? Again, discuss appropriate behavior for students working with a storyteller (e.g., they are welcome to ask Gibson questions directly).

Day 4

ArtsPartners Program
Dan Gibson, Storyteller
Presented by Young Audiences
(45 minutes)

Gibson tells thought-provoking, entertaining stories that make students laugh, stimulate their thinking, stir their imagination, or offer a new perspective. Old-time banjo music and sounds of the mountain dulcimer add musical textures to his programs.

Objective: The students will extend their understanding of idioms and figurative language to other kinds of expressive language as they occur in the work of a skilled storyteller.

Activity 7 (15-20 minutes): Discuss Dan Gibson's performance with the students. Refer back to the KWL chart and ask the students what they learned about using idioms. In addition, ask what other kinds of expressive language Mr. Gibson used. Have students think about Mr. Gibson's use of different kinds of voices, tones, and inflection. Record this information in the L column.

STUDENT WRITING PROMPT HANDOUT IDIOM STORY

STUDENT NAME: _____
SCHOOL: _____
DATE: _____
TEACHER: _____

In five to eight sentences, write a story about what happened when someone said "Hold your horses!" to Amelia Bedelia.

Day 5

Objective: Using a given idiom, the students will create a short story demonstrating their understanding of idioms.

Materials:

1. Chart paper or chalkboard
2. Manila paper, pencils, crayons
3. 15 different Idiom Story sheets with a different idiom at the top of each sheet

Activity 8 (10 minutes): Review with the students all the discussions and programs they have had during this unit on idioms as a kind of figurative or expressive language. Ask students to share what they have learned about idioms and write a list of what they have learned about them on the board or chart paper. Have students give examples from the play and the storytelling program to demonstrate their knowledge of idioms and how telling good stories includes using creative, imaginative language.

Activity 9 (20 minutes): Have students look at their sheet of paper and read their idiom silently. At this time the students should not share their idioms with their classmates. Explain to the students that they are to create a funny and interesting story using their idiom. The students do not have to begin their stories with the idiom, but the idioms need to be included in some part of the story. Each student's story should contain five to eight complete sentences.

Activity 10 (10 minutes): Students may begin illustrating their story on a piece of manila paper with pencils, crayons, or markers.

Activity 11 (5 minutes): Allow the students to share their finished stories with the class.



History Reporters Lesson Plan

Dallas Arboretum—Texas Pioneer Adventure Young Audiences of North Texas—Theater Improvisation with Sarah Weeks Fourth Grade, Fall 2004

Big Idea: Students will use observation and inference as they explore pioneer life by looking at photos, touring homes at the arboretum, and acting out scenes with Young Audiences of North Texas theater artist Sarah Weeks. Then, they will write an informational essay that shares something new that they have learned about pioneer life, using the details and inferences they have collected.

Integration of Social Studies and Language Arts Curriculum:

A visit to the arboretum's Pioneer Village will allow students to enter, observe, and think about various settings. Students will find themselves sitting in a covered wagon, standing in pioneer homes, and walking through gardens attached to the homes. During the tour, students will be accompanied by docents at the Dallas Arboretum who will prompt students to experience the space, to grasp what life could have been like in that setting, and to develop the vocabulary and images to convey what life under such different conditions must have been like (e.g., problems or challenges the pioneers had and the solutions they found).

This tour will give students the opportunity to make inferences about what life was like for the people who lived in each dwelling. Afterwards, Ms. Weeks will reinforce the insights gained through improvisation. She will then help the students translate their impressions into informative essays in which students organize their insights into life in a different time and place (setting). Ms. Weeks will help students select the most important details, choose effective vocabulary, and personalize their observations to create a slice of life in pioneer Texas.

COMPREHENSION SKILLS

TEKS 4.10 A, H
Making Inferences

Explain that when readers make inferences, they use information from the text along with personal experiences or knowledge to understand the total picture about a character or an event: The inference is not explicitly stated in the text. Have students use the following clues from the story to make an inference about Anna:

- Anna thought when Caleb was born, "He was homely and plain, and had a terrible holler and horrible smell."
- She went to bed that night thinking about "how wretched he looked." She forgot to say good-night to her mother. Her mother died the next morning.

We can infer that Anna is sad because she was upset about Caleb when he was born and forgot to say goodnight to her mother. What other inferences can the students make about Anna and Caleb?

TEKS 4.121

Have students identify details that help describe the setting of the story. What details let them know the time period in which this story takes place? How does the author describe the prairie? Why is the setting important to the story?

Purpose: Students will understand that as "history reporters" they can investigate and then inform others by writing essays with interesting details and powerful vocabulary that create pictures in their readers' minds of what life was like for pioneers in Texas.

Materials:

1. History Reporter journal (provided by ArtsPartners)
2. Color pictures of pioneer homes for overhead projection
3. Color pictures of pioneer homes
4. Arboretum study guide (provided by Arboretum)
5. Explorer's Scout journal (provided by Arboretum)

Teacher note: Although this is a scripted lesson, we trust you to guide your students through it based on what you know about their skills and knowledge level. You and the docent or artist are partners. Please help them scaffold and connect with your students. You are as responsible for the integration as the docent or artist is.

Day 1 Introducing Inference

These activities should help to build toward the visit to the arboretum and increase students' ability to infer not only from text but from the objects they see.

Activity 1 (45 minutes): Using the Dallas ISD scope and sequence, read *Sarah, Plain and Tall* in *Open Court Reading* textbook and complete the Making Inferences activities in the teacher's manual (see chart above and at the top of the next page).

Also, using the social studies textbook, ask and discuss with students, What is a pioneer, and when did they live?

Day 2 Acting as a Camera

Activity 2 (5 minutes): Review the definition of *inference*. Ask students what the difference is between guessing and making inferences. Explain that making an inference means concluding from fact or

COMPREHENSION SKILLS

TEKS 4.10 A, H
Making Inferences

Explain to students that they can make inferences by taking small pieces of information about a character or story event and using these pieces to make a statement about that character or event.

Tell students that they made inferences about how Anna felt about Caleb's birth. What other inferences can they make about Anna and Caleb?

- On page 50, Caleb asks Anna to recall the songs their mother sang so he could remember her. How do you think he feels about his mother? (We can infer that Caleb is sad because he never knew his mother.)
- On page 51, tears come to Anna's eyes after Caleb asks her to remember the songs their mother sang. Why do you think she was sad? (We can infer that Anna misses her mother.)

evidence (the activity performed by a reader or interpreter in drawing conclusions). It is guessing by taking the best idea based on the facts. Tell students that now that they have practiced drawing conclusions from a story, you want them to practice drawing conclusions based on what they can observe in a photograph of a setting.

Create a master chart to capture all information that emerges during this lesson (see Master Chart below).

Activity 3 (15 minutes): Place one of the arboretum pioneer home transparencies on your overhead projector. Ask students to describe what they see. Pick out a few details (use Acting as a Camera cue card on the next page for ideas); encourage students to pick out still more. Model drawing inferences based on the evidence of those details. The emphasis has to be on using observed detail, drawing inferences from it, and then translating it into a pioneer's life (i.e., problems the pioneers had and the solutions they might have found).

Activity 4 (15 minutes): Hand out students' History Reporter journals and take out the packet of color photos

SETTING	DETAILS / OBSERVATIONS	INFERENCES / CONCLUSIONS
	<p>What information does the setting give you?</p> <p>What details jump out or catch your eye when you enter the home?</p>	<p>What can you infer about pioneers who lived and worked in these homes or gardens?</p> <p>What problems or challenges might the pioneer who lived in this house have faced?</p> <p>How might he or she have solved the problem or faced the challenge?</p>
Covered wagon		
Tepee		
Sod house		
Lincecum house		
Lindheimer house		

Master Chart: Pioneer Observations and Inferences of the Arboretum Pioneer Home

Acting as a Camera Cue Card

Big Idea: Ask carefully sequenced questions that lead to discovery.

Ask students to think of an image as a “scene from a time” that they, as reporters, need to investigate. For each image you project, ask a series of questions that spiral from the basic to the critical-thinking level. In an exercise like this, students often want to analyze images with interpretive statements without carefully inspecting all the visual details. Move to the next level of questioning ONLY when most of your students can “see” the answers to your questions. To keep engagement high, show a new image every 5 to 15 minutes or until you feel students have a satisfactory understanding of the concepts.

Level 1: Ask students questions that require them to describe the details they see as though they were a camera recording the information objectively. Ask the following:

1. Imagine you are a camera. What do you see in this image?
2. What are the natural resources you see in the picture (grass, trees)?
3. What do you NOT see in the pictures? (What do you have in your house that you don't see here?)
4. Why do you think that is there? Why do you think it is right there? (It's good to ask questions that may be harder to answer.)

Don't move to the next question until students can point out many

details in the image. You might ask several students to stand around the image and actually touch the details they see. This will give your students the building blocks they need to answer the higher-level questions you will soon ask.

Examples of Details

Level 2: Ask students questions that require them to formulate ideas or make inferences based on existing evidence. Sample questions to ask students:

1. Think about a mom trying to cook dinner. What would that be like in this home?
2. What if a kid were playing? What or how would he or she play?
3. What if a dad were responsible for getting food for the family?
4. What would it be like in this house if it started to rain? During the summer when it is really hot? During the winter when it is really cold? (Sample answer using tepee as image: Because of cone shape, water would run down the side, but the floor or ground of the tepee would get muddy; during the summer the tepee would offer shade and could let the breeze in via doorway; during winter the breeze getting in would make it very cold.)
5. How did the people who lived in these homes use the natural resources in their environment to survive? (Sample answer using

sod house as image: They used the easiest and most plentiful thing they could find—dirt/sod—to build their home. This made building cheap. Although they might have had to repair their home often, they could get more sod quickly and easily and without heavy lifting.)

Level 3: Ask students questions that require them to consider the scene as a whole and make hypotheses about what life was like in this home and why. Sample questions to ask students:

1. What do you think mattered to the people who (lived in tepees, traveled to Texas in covered wagons, lived in sod houses, lived in plank houses)? (Sample answer using tepee as image: They used what they found—animal skins, wood for poles. They were proud and artistic, making paint and decorating their homes with symbols and scenes so that others would know who lived in their tepee. Their home was easily moved so they could travel with their home from place to place. They lived close to one another. The tepee has only one space so people must have slept, sat, and cooked next to one another.)

Source: Question levels and some directions come from *Social Studies Alive! Engaging Diverse Learners in the Elementary Classroom* by Bert Bower & Jim Lobdell (Cordova, CA: Teachers' Curriculum Institute, 2003).

Tepee	Covered Wagon	Sod House	Lincecum House	Lindheimer House
Painted designs	Big wheels	Box/square house; flat roof	Box/square house with slanted roof in front	Slanted roof with flat porch roof
Crossed poles coming out through top	Articles (canteen, rope) hanging off outside	Fire pit outside house (not next to house)	Covered porch	Metal roof
Cone shape	Top curved, white material	Large (deep) black pot on the fire pit	Wood size and shape is consistent, planks/posts are being used	Furniture on porch
Small/short entrance	Steps leading to entrance	Texture of house looks rough	Trees, shrubs give shade to house	Window with shutters
Material of tepee wrinkled	Bottom inside has built-in bench	Open air window	Window with shutters	Lots of different plants, flowers, spiked leaves

Examples of Details Students Should Observe

of pioneer homes. Then, assign small groups (of three or four students) to each picture. Ask them to “act as a camera” and take a close look at the picture, then describe what they see (making notes in their journals as they observe). Once they have taken in the picture, ask them: What can you infer about life in this house? What conclusions can you make based on what you see? What kind of chores did pioneers do? What kind of foods did they eat? What were their houses like? How did they get heat and light? Remind students that they must answer these questions using information they can infer from the setting, not fiction!

Activity 5 (15 minutes): Have a “lead reporter” from each group share the group’s observations and inferences about their photo. Have the corresponding photo on the overhead projector for the whole class to see. Be sure to take notes on the class master chart as students share. Ask the other students to act as “fact checkers” by listening and looking to find all the observations being shared. Then, ask the other students if they believed all the inferences were logical conclusions that their observations supported. Be sure that students understand that inferences must be based on observable details and logical conclusions (facts only!).

Each student should take notes in his or her History Reporter journal during each presentation. In this way, each student will already have some observable details and inferences based on the photos. Tell students that when they go to the arboretum, they will collect additional ideas, information, and details to include in their notes. Plus, they may also need to make some changes. Sometimes a picture does not give a true story because it may make a room appear bigger than it actually is, or it may not include a detail that is important and could change the information you need to make a correct inference. So tell students to get ready to collect as much information as possible during their visit to these actual sites.

Day 3

Texas Pioneer Adventure Tour at the Dallas Arboretum

(Use In the Setting Cue Card)

Teacher note: It is very important for you to partner with the docent during your experience. The docent is the expert about the arboretum resources, but you are the expert about your students’ skill level and background knowledge.

Docent note: Please say something like: I know you'll be writing about what you learn later in your classroom. I'm really interested in reading what kids have to say because I give tours all the time. Your writing will tell me what kids notice, what I can point out to other kids, the best ideas kids have. When I get what you write, it will let me teach other docents and adults what kids notice, observe, and think about and what kids are really able to do.

Activity 6 (60 minutes): Hand out students’ History Reporter journals and ensure that all students have a pen or pencil prior to beginning the tour. Then, as students visit the tepee, covered wagon, sod house, and cabins, have them add to and edit the notes they began in the classroom. At each stop on the tour, point out something that could NOT be observed in the photos you all reviewed in class. Ask students if this new detail gives them more or different information. At each stop or setting encourage students to do the following:

1. Take 30 seconds of quiet reflective time to take notes or make observations in your journals. Think of single, descriptive words to describe how it feels in the space. You will then have a few minutes to share these words with the class.
2. Take 30 seconds to tell me what questions you have before the docent begins talking.

Use In the Setting cue card to encourage and involve students throughout the rest of the tour.

In the Setting

Cue Card for Touring the Homes at the Arboretum

Big Idea: Ask questions that lead students to discover NEW information not visible in the photos of the settings.

At each stop on the tour, ask the three levels of questions. Remind students that they have an “expert” on the tour with them. They should look to the tour guide to help them. (*Teacher note:* Encourage the students to answer first and then ask the tour guide to elaborate, rather than have the tour guide answer every question. Remember, we want to engage the students.)

Level 1: Ask students to describe the details that they can touch or see in each setting or home. Sample questions to lead students:

1. What information does the setting give you?
2. What details jump out or catch your eye when you enter the home?
3. If you had to describe this setting to a blind person, what words would you use to describe how it looks and feels? Choose one good, descriptive word and share it with your classmates.

Level 2: Ask students to offer ideas or make inferences based on existing evidence in each setting or space. Tell them to use the details they observe to make informed or thoughtful conclusions. Sample questions or phrase to lead students:

1. What would it be like to live in this house if it started to rain? During the summer when it is really hot? During the winter when it is really cold?
2. Using the details you have observed, explain how people who lived in these homes used the natural resources in their environment to survive.
3. Using the details you have observed, explain how a feature or detail of this house would have made life better for a pioneer.

Level 3: Ask students to consider the scene as a whole and make hypotheses about what life was like in this home and why. Ask them to think about an 8- to 10-year-old pioneer growing up in each house. Then have them role-play using details and inferences to generate ideas about people who lived and worked in these homes. Sample questions to lead students:

1. What can you infer about the people who lived in this home or setting?
2. What problems or challenges might a pioneer living in this house have faced?
3. How might he or she have solved the problem or faced the challenge?

Day 4

Review Experience

(Back in the Classroom)

Activity 7 (20 minutes): Discuss as a class what the students learned about the people who settled Texas based on the types of early homes and gardens they saw on the Dallas Arboretum's Texas Pioneer Adventure tour. Update the master chart (created on day 2) with the new information the students captured in their journals—their observations of and inferences about the tepee, the covered wagon, the sod house, and the Lincecum and Lindheimer houses.

As students share from their journals, ask others to act as “fact checkers” by listening and looking through their notes to find all the observations being shared. Ask students if they believe all the inferences are logical conclusions that can be supported by observed details. Be sure that students understand that inferences must be based on observable details and logical conclusions.

Discuss how life was different for pioneers in each home using informed and thoughtful conclusions drawn from observed details and inferences.

Day 5

Theater Improvisation Workshop with Sarah Weeks

Teacher note: It is very important for you to partner with the artist during your experience. Ms. Weeks is the expert about theatrical and improvisational techniques, but you are the expert about your students' skill level and background knowledge.

Activity 8 (5-10 minutes): Ms. Weeks will work with students to translate their experiences into effective writing. She will begin by having students recall what the pioneer settings were like using their bodies and the information captured on the master chart and in their History Reporter journals.

Activity 9 (35 minutes): Ms. Weeks will then have students work in small groups to act out vignettes of the pioneer experience. As the students act out life within each home, the artist will prompt the students to have the following:

- **Main idea or thesis.** What do you want your audience to most understand about life as a pioneer?
- **Viewpoint.** What is your reason for this main idea, and what details in your scene support your reason?
- **Rich vocabulary.** What word(s) will best describe your idea? [She may have the other students close their eyes and then ask the student in the scene to use words that help the other students see the scene in their minds.]

Day 6

Informational Writing

Activity 10 (5-10 minutes): Discuss what it means to have a “big idea.” Explain to students that a big idea is something new that they have discovered or learned about pioneer life that they want to share with your readers. Because they want to share this idea, it controls what details or supporting information they write about. So, when they look through their journal, they should carefully choose those details, observations, and inferences that support their big idea. Also, a student's big idea will control the words that student chooses to use. Remind them that descriptive words that help the reader “see” and “feel” what they are sharing make their writing more powerful.

For example, explain that one big idea is that pioneer life was hard. What sort of details, observations, and inferences could they use to support this idea? Once students have come up with some answers, ask them to think of their own big idea (they can't use “life was hard”). Have them choose one of the homes and share their big idea. They should describe what life was like for a pioneer living in that home and remember to use their observations and inferences from the arboretum and from their work with Ms. Weeks. Their essays should have the following:

1. A short introduction that shares with the reader something new they learned about pioneer life.
2. Several (two or three) support paragraphs in which they use their observations and inferences to explain what they want their readers to understand about pioneer life. They should use their History Reporter journals and the master chart posted in the classroom.

If students are having trouble coming up with a big idea, ask these leading questions:

1. What can you share that will teach someone something that they may not understand about pioneer life?
2. What details did you see inside the house that helped you make these conclusions?

Activity 11 (20 minutes): Have students write a three- or four-paragraph informative essay about pioneers in Texas (see writing prompt on the next page). At a midpoint, pick out two or three strong examples of inferences in students' writing and ask the students to share them with the class.

Activity 12 (20 minutes): Have students continue writing their essays, making use of the examples they've just heard.

STUDENT WRITING PROMPT HANDOUT

STUDENT NAME: _____

SCHOOL: _____

DATE: _____

TEACHER: _____

Write a three- or four-paragraph informational essay about pioneers in Texas. Choose one of the homes you've been learning about (tepee, sod house, covered wagon, Lincecum house, or Lindheimer house) and describe what life was like for a pioneer living in that home. Remember to use your observations and inferences from the arboretum and from your work with Sarah Weeks. Your essay should have the following:

1. A short introduction that shares with your reader something new you learned about pioneer life.
2. Several (two or three) more paragraphs using your observations and inferences to explain what you want people to understand about pioneer life. Remember to use your History Reporter journal and the master chart posted in the classroom.

Note: Your informational essay must be nonfiction. Use only facts that you can observe to make inferences (because conclusions are based on facts).

APPENDIX E: FOR RESEARCHERS

Construction of Comparison Groups

To facilitate reliable and valid statistical analyses, four comparison groups were created by reconstructing previously gathered data files into one longitudinally designed database. These four groups were designated as (a) Focus Students, (b) Focus Class, (c) Focus Grade, and (d) Control Grade. All students who were enrolled in one of the eight study schools (four treatment and four control) at any time during the four years of the study (2001-2002, 2002-2003, 2003-2004 and 2004-2005) were coded by year into one of these groups. Therefore, group membership changed slightly each year. An Other category was comprised of students who repeated a grade or took their standardized tests at a school other than one of the study schools in which they were initially enrolled; they were not included in test score analyses. It was assumed that control schools may have used ArtsPartners materials, but did not receive special programming, as did the Focus students in the Focus classes in the treatment schools.

Focus students. Selected students from classes that had received specific ArtsPartners programming were identified as Focus students. During the four study years, the only added interventions these students received were interviews and literacy sample collections. By design, Focus students at each school remained together in the same class for all four years of the study.

Focus Class students. In the Dallas ISD, elementary teachers are given advisor numbers that uniquely identify the grade and section that they teach or that is their homeroom. The advisor numbers of the grade 1 through grade 6 teachers from the treatment schools (24 total) were taken from Dallas ISD databases. Any student that was not a Focus student, but had an advisor number that matched the six teachers at each school was designated as a Focus Class student for that year. There was an underlying assumption that the Focus Class students received the same ArtsPartners programming as the Focus students. Most probably did; however, there was no way to take into account absences or student mobility.

Focus Grade students. Any student enrolled in one of the four treatment schools at the correct grade level in any class other than the Focus Class for the corresponding year of the study was designated as a Focus Grade student.

Control Grade students. Control Grade students were enrolled in the control schools at the comparable grade as the Focus Grade students for that year.

Other database designations. Because one database contained all students who had one of the above designations at any time during the four years of the study, and because the population changed from year to year, during some years, students were coded as "Not in the study at this time," "Retained," "Tested at a different school" (meaning that they moved during the school year), or "Left the study." These students' test scores were not included in statistical analyses for that year.

Methodology

Observation Data Collection

Focus students were observed twice yearly, once in a writing activity that was part of an ArtsPartners lesson cycle (CL+AP) and the other in a writing activity following a regular classroom experience not containing ArtsPartners content (CL). Thus, there was a pair of observations for each of the four years of the study (2001-2002, 2002-2003, 2003-2004, and 2004-2005).

For the first time in 2004-2005, there was another set of observations. Students were observed not only in the writing exercise, but also in the pre-writing activities that would introduce and "inspire" their writing activities. Similar to the other observations, one prewriting experience occurred during ArtsPartners programming, while the other was a regular classroom lesson geared to content from the grade 4 *Open Court* basal reader. Trained researchers observed students and kept a running record of students' activities and statements during the observations. Records were coded into categories of individual learner behaviors.

Some students' observations occurred in the fall semester, while others occurred in the spring. A previous analysis of ArtsPartners literacy data found differences in performance based on the content of the specific ArtsPartners lesson. Because of this, the semester of the observation was used as an independent variable for most analyses. These analyses also can be used to discern whether the content of the lesson impacted observed behaviors.

Learner Behavior Categories

Four overarching learner behavior categories (distinguished by capital letters) were created from 21 observed and coded individual learner behaviors: Classroom Participation Behaviors, Self-Initiated Learning, Drawing on Resources, and Using Alternate Learning (see Appendix B). The individual learner behaviors that contributed to each learner behavior category are listed in Appendix A and Appendix B.

Analyses assessed differences in learner behavior constructs for Focus students when they were in (a) ArtsPartners lesson cycle observations (CL+AP) and classroom observations (CL) and (b) CL+AP and CL prewriting observations. Multivariate analyses of variance (MANOVA) for each year that observation data were collected were computed with the four aggregated learner behavior constructs as the dependent variables and the observation setting and semester as the independent variables.

Literacy Sample Collection

Literacy samples were collected from Focus students that received special ArtsPartners programming each semester. As part of the Focus students' ArtsPartners integrated lesson cycle (CL+AP), they were asked by their evaluator to write about a specific topic. Most students had two experiences per year, one in the fall and the other in the spring semester. Within two weeks, Focus students' teachers were asked to submit a writing sample created during a typical classroom assignment. The scores for these samples were coded CL samples. Sample sizes remained stable, but some students changed for the six semesters because the Focus population changed due to mobility. Lesson descriptions can be found in Appendix D.

Literacy samples were scored by trained evaluators using the Northwest Regional Educational Laboratory's (NWREL) 6+1 Trait writing assessment (see Appendix C). Scores were given to the evaluator for statistical analyses.

Differences between Focus students' ArtsPartners lesson cycle samples (CL+AP) and classroom samples (CL) were computed using multi-variate analyses of variance (MANOVA) for each semester that samples were collected. Scores on each of the six traits were the dependent variables, and the type of sample was the independent variable.

A MANOVA procedure was used to measure changes in six trait ratings

of Focus students on ArtsPartners lesson cycle writing samples. Procedures were conducted by trait. The dependent variables were the six traits, and the independent variable was time. Mean ratings on traits for these analyses were frequently different from those for other sections due to the changing nature of the population. As long as students were Focus students during that semester, their ratings were included in the analysis. Content of lessons was compared to trait score patterns.

Texas Assessment of Knowledge and Skills (TAKS) Preparation of Test Scores

TAKS reading and writing scores were matched by Dallas ISD student identification number for all students in the longitudinal database, regardless of their database coding for any one specific year. For both the reading and writing subtests, the state uses the raw score to set a "cut score" designating a Pass-Fail determination.

Writing Composition Scores

The writing portion of the state criterion-referenced test contains both multiple-choice items and a written composition. This composition receives a score ranging from 0 to 4, with scores of 2, 3, or 4 considered as "Pass." A 0 is given when the composition has a nonscorable response. For this report, the written composition scores were described for the grades 4 (grade 1 cohort) and 7 (grade 4 cohort) students in 2005. A designation of "Pass" or "Fail" on the entire writing test is determined from the aggregate of the multiple-choice items and the written composition.

Percentage Correct of Total Possible Items

The percentage correct of total possible items on the Texas Assessment of Academic Skills (TAAS), given in 2001 and 2002, or TAKS reading test, used beginning 2003 to the present, was computed for each testing period. This score was used to assess growth across the four years for the grade 4 cohort only and to compare the cohorts for the differing time periods. Only students who remained in the same comparison group for grades 4-6 and were enrolled in a Dallas ISD middle school in grades 7 and 8 were included in analyses of growth.

Statistical Comparisons

Chi-square (χ^2) tests were used to assess differences in (a) percentages of students in each of the comparison groups that passed the TAKS reading or writing tests for the respective year and (b) TAKS writing composition scores in each of the comparison groups. Scores of 0 or 1 were considered "Fail," while scores of 2, 3, or 4 were considered "Pass."

A repeated-measures design was used to assess growth across the six years, 2001-2002, 2002-2003, 2003-2004, 2004-2005, 2005-2006 and 2006-2007 for the grade 4 cohort only. Because tests had a different number of possible items each year, percentages of correct responses were used as the dependent variable. The independent variable was the student comparison groups.

Results

Differences in Learner Behavior Categories and Literacy Sample Ratings for ArtsPartners and Classroom Writing

Although all grade levels were analyzed, only the grade 4 observations were presented in this report. The lessons are described in Appendix D.

ArtsPartners and Classroom Writing Lessons: Fall 2002

The multivariate test of differences between trait scores for the grade 2 fall writing samples was not significant [Wilks lambda (6, 57) = .532, p = .782]. Similarly, none of the between-subject effects were statistically significant. The greatest difference between the two samples was Voice, where the ratings were higher on the CL+AP than the CL samples, and Conventions, where the CL+AP ratings were higher than the CL ratings (see Figure 16).

ArtsPartners and Classroom Writing Lessons: Fall 2004

For the grade 4 fall 2004 writing samples, there were two between-subject effects that were statistically significant: Ideas/Content [$F(1, 62) = 4.394$, p = .04, η^2 = .066] and Voice [$F(1, 62) = 4.092$, p = .047, η^2 = .062]. Ratings were higher in the Arts-Partners setting for all traits except Conventions, where ratings were the same (mean = 2.9) (see Figure 17).

Differences in Observed Behaviors in ArtsPartners and Classroom Prewriting Sessions

When the observed behavior in prewriting sessions for grade 1 cohort students (now in grade 4) were assessed, there was a significant multivariate difference among the four learner behaviors for both the setting [Wilks lambda (4, 56) = .527, p < .001, η^2 = .473] and the semester [Wilks lambda (4, 56) = .776, p = .006, η^2 = .224]. The setting explained almost half (47 percent) of the variance in learner behavior construct scores. There were

There were significant differences in between-subjects effects for learner behaviors for setting, semester and the interaction between the two. Students were observed in Classroom Participation Behaviors more often in the CL+AP setting than in the CL setting [$F(1, 59) = 12.728$, p = .001, η^2 = .177], with setting explaining 18 percent of the variance (see Figure 9). Significant differences by semester were noted for Drawing on Resources [$F(1, 59) = 6.558$, p = .013, η^2 = .098]. Behaviors defining Self-Initiated Learning also were seen more often in the fall AP setting (mean = 5.44, standard deviation = 4.9) than the spring AP setting (mean = 2.1, standard deviation = 3.5). In Figure 9, Classroom Participation Behaviors were graphed separately from the other constructs because its scale was so disparate from the other constructs' scales.

Longitudinal Changes in the Six Traits of Writing

A final comparison assessed the changes over the four years in the ratings on the six traits. The multivariate test for differences in the six traits across time was statistically significant [Wilks lambda (30, 706) = .681, p < .001, η^2 = .074], with time of measurement or sample explaining 7 percent of the variance in trait scores. In addition, between-subjects effects were significant for all six traits, with effect sizes explaining from 9 percent (Conventions) to 22 percent (Ideas/Content) of the variance in ratings.

There was considerable growth in ratings of the traits from the first grade 2 sample in fall 2002 to the second rating in spring 2003, then to the fifth rating in fall 2004 (see Figure 15). Differences in these ratings were responsible for the majority of the statistical differences.

State Criterion-Referenced Reading and Writing Assessments

Spring 2005 TAKS Reading and Writing

There was a significant difference in writing composition scores of grade 4 students among the comparison groups [$\chi^2(12) = 31.522$, p = .002, V = .149]. A greater percentage of Focus students scored a 2 (45 percent) or a 3 (48 percent), while the other comparison groups had greater percentages scoring a 2 (65 percent to 67 percent) (see Figure 12).

Several analyses indicated that there was a significant positive residual effect from participation in ArtsPartners programming for grades 7 (2005) and 8 (2006) students that were part of the original grade 4 cohort.

Although there was no significant differences in grade 7 writing composition scores [$\chi^2(12) = 10.653$, p = .559], 100 percent of the Focus students received passing scores on their compositions (see Figure 13). Focus Class and Control Grade students had the smallest percentage of students passing (88 percent).

There was a significant difference in the percentage of grade 7 students passing the 2005 TAKS Reading test among the four student comparison groups [$\chi^2(3) = 9.676$, p = .022, V = .142]. Even though Focus students had received no further ArtsPartners intervention during the 2004-2005 school year, a significantly higher percentage passed than the other student comparison groups (see Figure 9).

Differences in Growth in Reading

When the grade 1 cohort students were assessed in 2004 (now in grade 3), Focus students scored 10 percentage points higher than their Focus Grade peers. For the 2005 administration, Focus students (78 percent correct answers) performed better than the other two comparison groups, which clustered together (71 percent correct) (see Figure 10). Although teachers may have used ArtsPartners lessons

in 2006, the longitudinal study concluded in 2005. For the grade 5 TAKS administration, students from the Focus schools (Focus, 78 percent; Focus Grade, 77 percent) had a greater mean percentage of answers correct than the control (72 percent). There was a statistically significant difference between the Focus and Focus Grade students and the Control Grade students [$F(2, 362) = 4.495$, p = .012, χ^2 = .024], with group accounting for 2 percent of the variance in scores.

There are six years of data for the grade 4 cohort, beginning with a "pretest" score in 2001. At that time, all groups were in grade 3 and scored similarly (between 75 percent and 79 percent correct). For the next five years, Focus students had a mean percentage of correct answers higher than all other comparison groups, even though none of the students received ArtsPartners programming of any kind after the 2004 administration (see Figure 11). By the 2006 administration, when these students were in grade 8, Focus and Focus Grade students (91 percent and 88 percent correct answers, respectively) scored well above total district eighth-graders (77 percent). Control Grade students began a steady increase in percentage of answers correct in 2004 (65 percent), which continued through 2006 (87 percent), when their scores were almost as high as those of Focus Grade students. Although there were no statistically significant differences among the groups in 2006 [$F(2, 277) = 2.122$, p = .122], 13 Focus and Focus Grade (9 percent) and 6 Control Grade students (4 percent) students answered 100% of their grade 8 TAKS reading questions correctly.

APPENDIX F: ARTSPARTNERS LONGITUDINAL STUDY LEAD PARTNERS

Lead Partners

Big Thought

Big Thought is a learning partnership inspiring, empowering, and uniting children and communities through education, arts, and culture. The big thought is that a community, working together, can lift children up and better their lives using arts and culture as tools and catalysts.

Big Thought supports community partnerships, cultural integration for academic achievement, youth development, and family learning.

Annenberg Institute for School Reform

The Annenberg Institute for School Reform is a national policy-research and reform-support organization based at Brown University that focuses on improving conditions and outcomes in urban schools, especially those serving disadvantaged children. The institute works through partnerships with school districts and school reform networks and in collaboration with national and local organizations skilled in educational research, policy, and effective practices to offer an array of tools and strategies to help districts strengthen their local capacity to provide and sustain high-quality education for all students.

Dallas Independent School District

The district's mission is to prepare all students to graduate with the knowledge and skills to become productive and responsible citizens.

Goals

- Improve student achievement
- Nurture and develop teachers and other employees
- Earn the community's trust through good financial management
- Improve the district's facilities
- Maintain a safe and secure environment

City of Dallas, Office of Cultural Affairs

The mission and purpose of the City of Dallas, Office of Cultural Affairs (OCA) is to enhance the vitality of the city and the quality of life of all Dallas citizens by creating an environment wherein artists and cultural organizations can thrive. One of the major roles of the OCA is to provide management assistance to promote stability and collaboration and resource development to increase the productivity of the organization. The OCA is responsible for implementing the City's Cultural Policy, which advocates access to the arts, public-private cooperation, creative expression, cultural diversity, and citizen involvement.

School Partners

Casa View Elementary

Principals during the study

- Mike Paschall (2001-2005)

Teachers during the study

- 2001-2002
 - Sally Colden: first-grade teacher
 - Janet Burgess: fourth-grade teacher
- 2002-2003
 - Tammy West: second-grade teacher
 - Nadine Corbo: fifth-grade language arts teacher
- 2003-2004
 - Dawn Jiles: third-grade teacher
 - Bryan Robinson: sixth-grade teacher
- 2004-2005
 - Sally Ragle: fourth-grade teacher

Hogg Elementary

Principals during the study

- Manuel Rojas (2001-2003)
- Mrs. Watanebe (2003-2005)

Teachers during the study

- 2001-2002
 - Raul Alvarez: first-grade teacher
 - Amanda Terrice: fourth-grade teacher
- 2002-2003
 - Jesse Thomas: second-grade teacher
 - Sherletrice Johnson-Berry: fifth-grade language arts teacher
- 2003-2004
 - Mary Price: third-grade teacher
 - Keisha Wright: sixth-grade language arts teacher
- 2004-2005
 - Sanya Dozier: fourth-grade social studies teacher

Marsalis Elementary

Principals during the study

- Gloria Lett (2001)
- Patricia Weaver (2001-2002)
- Cheryl Malone (2002-2005)

- 2001-2002
 - Brenda Cooper: first-grade teacher
 - Sheila Lyons: fourth-grade teacher
- 2002-2003
 - Jeanine Hill: second-grade teacher
 - Rhonda Hays: fifth-grade language arts teacher

- 2003-2004
 - Sonya Giles: third-grade teacher
 - Imogene Binion: sixth-grade language arts teacher
- 2004-2005
 - Deborah Polk: fourth-grade language arts teacher

Walnut Hill Elementary

Principals during the study

- JoAnne Hughes (2001-2004)
- Brian Lusk (2004-2005)

Teachers during the study

- 2001-2002
 - Michelle Wood: first-grade teacher
 - Linda Moody: fourth-grade teacher
- 2002-2003
 - Lashon Gales: second-grade teacher
 - Viveca Wilson: fifth-grade language arts teacher
- 2003-2004
 - Diane James: third-grade teacher
 - Angela Bell: sixth-grade language arts teacher
- 2004-2005
 - Cindy Slater: fourth-grade teacher

Control Schools

Arcadia Park Elementary
Cabell Elementary
Stemmons Elementary
Thornton Elementary

Cultural Community Partners 2001-2006

A.R.T.S. for People
African American Museum
Anita N. Martinez Ballet Folklórico, Inc.
Artreach-Dallas, Inc.
Arts District Friends
Black Cinematheque Dallas
Black Dallas Remembered, Inc.
CAMP (Collaborating Artists Media Project)
Cara Mia Theatre Company
Circle Ten Council, Learning for Life
Crow Collection of Asian Art
Dallas Aquarium at Fair Park
Dallas Arboretum
Dallas Black Dance Theatre
Dallas Children's Museum at the Museum of Nature and Science
Dallas Children's Theater
Dallas Classic Guitar Society
Dallas Community Television
Dallas Dance Council
Dallas Heritage Village
Dallas Historical Society
Dallas Holocaust Museum

Dallas Museum of Art
Dallas Nature Center
Dallas Opera
Dallas Summer Musicals
Dallas Supporters of the Fort Worth/Dallas Ballet
Dallas Symphony Orchestra
Dallas Theater Center
Dallas Trees and Parks Foundation
Dallas Zoo
Daniel de Cordoba Bailes Espanoles
Fine Arts Chamber Players
Frontiers of Flight Museum
H.O.P.E. (Honoring of People Everywhere)
Imagine the Impossible Dance Workshop
International Museum of Cultures
Japan-America Society of Dallas/Fort Worth
Jewish Community Center of Dallas
Junior Players
Kennedy Center Imagination Celebration at Dallas
MADI Museum
Making Connections, Inc.
Meadows Museum
Multicultural Arts Academy
Museum of the American Railroad
Museum of Nature and Science
Nasher Sculpture Center
New Arts Six
Ollimpaxqui Ballet Company Inc.
Our Endeavors Theater Collective
Partnership for Arts, Culture & Education
Shakespeare Dallas
Soul Rep Theatre Company
Southwest Celtic Music Festival
SPCA of Texas
Teatro Dallas
TeCo Theatrical Productions
Texana Living History Association
Texas Ballet Theater
Texas Discovery Gardens in Fair Park
Texas Trees Foundation
Texas Winds Musical Outreach
The Dallas Center for Contemporary Art
The Sixth Floor Museum at Dealey Plaza
The Women's Museum - An Institute for the Future
The Writer's Garret
Theatre Three
Tipi Tellers
TITAS, Texas International Theatrical Arts Society
USA Film Festival
Voices of Change
Young Audiences of North Texas

FUNDING PARTNERS

Legacy Donors

Thank you to these extraordinary contributors that have made the most significant investments in Dallas ArtsPartners. Their generosity has touched thousands.

American Airlines
Bank of America
Citigroup Foundation
City of Dallas Office of Cultural Affairs
ExxonMobil Corporation
Ford Foundation
The Horchow Family Charitable Trust
IBM
W.K. Kellogg Foundation
Kraft Foods of North America, Inc.
Lucent Technologies Foundation
The Meadows Foundation, Inc.
National Endowment for the Arts
Charles Vincent Prothro Family Foundation
The Howard Rachofsky Foundation
Mr. and Mrs. Edward W. Rose III
Michael L. Rosenberg Foundation
U.S. Department of Education**
Young Audiences, Inc.

Sustaining Donors

Thank you to these special contributors that helped sustain Dallas ArtsPartners by providing support for four or more years.

American Airlines
Bank of America
Centex Corporation
Citigroup Foundation
IBM
Charles Vincent Prothro Family Foundation
Mr. and Mrs. Edward W. Rose III
Michael L. Rosenberg Foundation
Texas Instruments Foundation

Trailblazing Donors

Thank you to the following visionaries that saw our potential in our earliest years and helped plant the seeds of our success. Each first gave in 1999.

American Airlines
Bank of America
Communities Foundation of Texas
Constantin Foundation
ExxonMobil Corporation
The Leland Fikes Foundation
Edith and Gaston Hallam Foundation
IBM
Lucent Technologies Foundation
The Howard Rachofsky Foundation
The Michael L. Rosenberg Foundation
SBC Texas

CUMULATIVE GIFTS

TO DALLAS ARTSPARTNERS
SINCE 1999

\$1,000,000+

US Department of Education

\$200,000 - 300,000

Ford Foundation
Mr. and Mrs. Edward W. Rose III

\$150,000 - 199,999

Bank of America
The Meadows Foundation, Inc.
National Endowment for the Arts
Michael L. Rosenberg Foundation

\$100,000 - 149,999

Citigroup Foundation
City of Dallas Office of Cultural Affairs
ExxonMobil Corporation
Charles Vincent Prothro Family Foundation

\$50,000 - 99,999

The Horchow Family Charitable Trust
W.K. Kellogg Foundation
Kraft Foods of North America, Inc.
Lucent Technologies Foundation
The Howard Rachofsky Foundation
Young Audiences, Inc.

\$25,000 - 49,999

Centex Corporation
Chase
The Constantin Foundation
Entertainment Industry Foundation
Leland Fikes Foundation
Hillcrest Foundation
Hoglund Foundation
Eugene McDermott Foundation

\$10,000 - 24,999

Harry W. Bass Jr. Foundation
Crowley-Carter Foundation
Dallas Foundation
Darden Restaurants Foundation
Depth Magazine
EDS Foundation
M.R. and Evelyn Hudson Foundation
Olive Garden
SBC Texas
Texas Commission on the Arts
Texas Instruments Foundation
George and Fay Young Foundation

\$5,000 - 9,999

Communities Foundation of Texas
Embrey Family Foundation
Edith and Gaston Hallam Foundation
Ben E. Keith
The John F. Kennedy Center
for the Performing Arts
Lockheed Martin Vought Systems
Perot Foundation
Vinson & Elkins L.L.P.

IN-KIND DONORS

Thank you to these generous companies who kindly provided essential services, equipment and products that have helped Dallas ArtsPartners flourish.

American Airlines
BKM Total Office of Texas, Inc
Cisco Systems
Container Store
Crescent Court Hotel
Dallas Community Television
Dallas Morning News/WFAA
Darden Restaurants Foundation
Daryl Flood Warehouse and Moving
Dennard, Lacey, & Associates
Fiesta Mart
Gardere Wynne Sewell L.L.P.
Grisoft
IBM
Imaginuity Interactive, Inc.
IPS Advisors
Lockheed Martin Vought Systems
M/C/C
Microsoft Corporation
Olive Garden
Padgett Printing
Salmon Beach & Company
Scholastic, Inc.
Southwest Airlines
Stark Technical Group
Steelcase
Stutzman, Bromberg, Esserman & Plifka PC
United Systems Integrators

** The contents of this publication were developed under a grant from the Department of Education. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

**Printing generously
underwritten by the**

W. K. Kellogg Foundation

Throughout this report, opportunities for building the capacity of stakeholders are highlighted to illustrate how a rigorous evaluation can build the knowledge and skills a community needs to design and improve programs for children and youth. In their own words, here are a few examples of what stakeholders gained through participating in the ArtsPartners longitudinal study:

The experience became a part of who I am as a teacher. I definitely see the connections. I think about that and the activities that brought about those connections and those exciting moments of learning.

—**Michelle Wood**, FIRST-GRADE TEACHER

Having other researchers as colleagues—building the codes, arguing about what we meant, all of that taught me a lot about the process of examining program outcomes and impacts. It was a combination course/support group/booster club for my own work with inner-city youth.

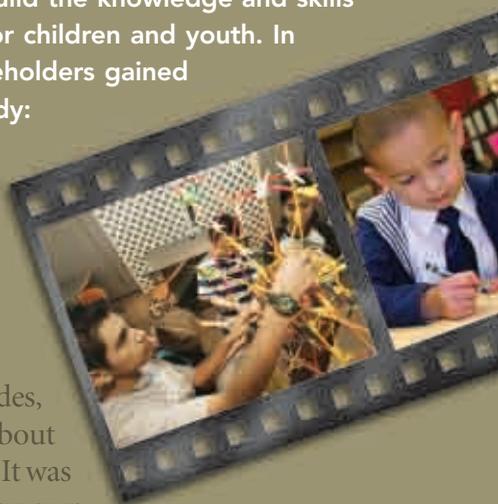
—**Janet Morrison**, FIELD RESEARCH AND COMMUNITY-BASED ORGANIZATION STAFF

I arranged for a member of my staff to participate as a researcher in this study for two reasons: to provide a valuable employee with a professional development opportunity and to stay informed about the progress of the evaluation. Very quickly I and my organization felt the positive impact of her participation. Through her work with the children and teachers, this staff member brought us insights into which lessons truly impacted students, how teachers were using support materials in the classroom, and the methodology of the evaluation process. This perspective guided internal discussions about future programming.

—**LeAnn Binford**, DALLAS SYMPHONY ORCHESTRA DIRECTOR OF EDUCATION, 1994-2005

It is the responsibility of educational institutions as well as the community to provide the outlets for discovery, for all students. As we increase the role of cultural learning, we want to ensure that what we are offering is the very best. Therefore, we all—the central office, principals, teachers, and our partners—must be able to evaluate and improve the learning we offer every day. This collaborative research shows that the Dallas community has the capacity to work in this way.

—**Michael Hinojosa**, GENERAL SUPERINTENDENT, DALLAS INDEPENDENT SCHOOL DISTRICT



ISBN: 0-9769704-1-4

ISBN: 0-0000000-0-0



0 000000 000000