THINKING THROUGH ART

ISABELLA STEWART GARDNER MUSEUM

IMPACT STUDY TECHNICAL SUPPLEMENT

THINKING THROUGH ART:
A TRANSFORMATIVE MUSEUM-SCHOOL PARTNERSHIP

IMPACT STUDY MATERIALS AVAILABLE AT
GARDNERMUSEUM.ORG/ORGANIZATION/EDUCATION/RESEARCH

THINKING THROUGH ART: A TRANSFORMATIVE MUSEUM-SCHOOL PARTNERSHIP (REPORT)

EXECUTIVE SUMMARY

TECHNICAL SUPPLEMENT

DATA COLLECTION & MANAGEMENT PROTOCOLS

RESEARCH INSTRUMENTS

DATA CODING & RATING MANUALS

INFORMATIONAL VIDEOS

Contact us: education@isgm.org

Sara Egan, Claire Tratnyek, and Mary Ellen Munley
Isabella Stewart Gardner Museum, Boston
2024

Recommended citation:
Egan, S., Tratnyek, C., Munley, M. E. (2024). Thinking Through Art: A Transformative Museum-School Partnership. Boston: Isabella Stewart Gardner Museum.

Front cover: Isabella Stewart Gardner Museum, Boston. Photo by Carlie Febo

ISABELLA STEWART GARDNER MUSEUM

MEM & Associates
MUSEUMS · PEOPLE · PUBLIC VALUE

The lead sponsor of the Thinking Through Art program is the Vertex Foundation. Additional support is generously provided by the Rowland Foundation, Yawkey Foundation, The Beker Foundation, Liberty Mutual Foundation, The Hamilton Company Charitable Foundation, Thomas Anthony Pappas Charitable Foundation, and Janet Burke Mann Foundation. The Isabella Stewart Gardner Museum also receives operating support from the Massachusetts Cultural Council, which is supported by the state of Massachusetts and the National Endowment for the Arts.

THINKING THROUGH ART: A TRANSFORMATIVE MUSEUM-SCHOOL PARTNERSHIP

IMPACT STUDY TECHNICAL SUPPLEMENT

TABLE OF CONTENTS

1. INTRODUCTION TO THE TECHNICAL SUPPLEMENT 2

2. REVIEW OF LITERATURE 4

3. STUDY DESIGN 13

4. ETHICAL CONSIDERATIONS 22

5. PARTICIPANTS IN *THINKING THROUGH ART: A TRANSFORMATIVE MUSEUM-SCHOOL PARTNERSHIP* IMPACT STUDY 23

6. DATA COLLECTION METHODS 26

7. DATA CODING AND ANALYSIS METHODS 29

8. STUDY LIMITATIONS AND RECOMMENDATIONS 43

9. STUDY STRENGTHS AND LESSONS 45

10. REQUEST FOR DATA 46

11. SOURCES CONSULTED 47

1. INTRODUCTION TO THE TECHNICAL SUPPLEMENT

Purpose of *Thinking Through Art: A Transformative Museum-School Partnership* Impact Study

Thinking Through Art: A Transformative Museum School Partnership is part of a long tradition of educational programming evaluation and research at the Isabella Stewart Gardner Museum, which has sought to assess the efficacy of its school partnership programs through rigorous practitioner-research since 2003. In this study, our team investigated how the Gardner Museum's Thinking Through Art program with Boston Public Schools aligned with the school district's goals for teachers and students. We asked research questions across three broad domains:

- **Does participation in Thinking Through Art increase teachers' skills for Culturally Responsive Instruction?**
- **Does participation in Thinking Through Art increase students' skills for Social-Emotional Learning?**
- **Does participation in Thinking Through Art increase students' skills for Critical Thinking?**

As we designed the study, we added corollary questions, connected questions in one area to patterns we saw in another, and realized again and again that the program at the center of our research is a complex, deep, and generative one.

Project Management

Sara Egan, Associate Curator of Education at the Gardner Museum, served as the Principal Investigator and general manager for this project from 2020-2024, with support from the rest of the Gardner Museum Education Department. Mary Ellen Munley served as the Research Advisor for the study, and led the development of all study instruments, data coding manuals and processes, and coder training agendas. Claire Tratnyek, an alumnus of the Thinking Through Art program, former Boston Public Schools teacher, and doctoral candidate, oversaw all data collection and data analysis tasks as the Lead Researcher and Project Manager for the study.

After the data had been collected, coded, and prepared for analysis, Alicia Holden joined the team as Statistical Analyst. The design, construction, implementation, and analysis work was led by Ms. Egan, Ms. Munley, and Ms. Tratnyek, with support from Gardner Museum staff and Research Assistants. The final report was prepared by Ms. Egan and Ms. Tratnyek, reviewed by Ms. Munley, designed by Gardner Museum staff, and reviewed by all members of the research team.

Practitioner-Researcher Approach

Our research team was advised by a professional researcher, but all of the study design, data collection, coding, and analysis work was done collaboratively with a group of experienced Gardner Museum staff educators, who served as practitioner-researchers through each phase of the project. Museum staff forged close relationships with teachers and students during data collection sessions in the classrooms, and museum visits where classes learned together in the galleries. Some members of the team were former classroom teachers from Boston Public Schools, and understood how the district operates. Beyond the insider knowledge of the district, our team of practitioner-researchers brought expansive knowledge and rich experience with the Thinking Through Art program and its theoretical and practical underpinnings into each meeting, each process of revision, and each training session. A strength of our study is that the people who know the program played a key part in developing rigorous evaluation techniques to assess it. The team's interest in understanding the Thinking Through Art program's strengths – and commitment to improving its weaknesses – makes it ideally suited to undertaking a multi-year, multi-modal study of it.

A 360° Look at Thinking Through Art

Thinking Through Art at the Isabella Stewart Gardner Museum uses the Visual Thinking Strategies (VTS) structure of discourse, which is a whole-class conversation carefully facilitated by the teacher. Teachers may also include nonverbal signaling, pair-share or small group discussions, individual writing or drawing, or other strategies to ensure that all students are engaged. Teachers practice creating supportive learning environments that are open and accepting, encourage risk-taking, and elicit multiple perspectives. Teachers increase their ability to hear, value, and assess student thought and expression. The artworks included in the TTA curriculum represent diverse cultures, eras, and makers, offering windows, mirrors, and sliding glass doors (Bishop, R. S., 1990). Teachers in Thinking Through Art are also trained on methods to disrupt biases and stereotypes that may arise in group discussions.

2. REVIEW OF LITERATURE

Museum Education Practices and Student Outcomes

Schools and museums like the Gardner share a vision of learning that is inquiry-based and student-centered, valuing multiple perspectives and active discourse (Terrassa et al, 2016). Research from art education, as well as from the Gardner Museum specifically, demonstrates the pedagogical alignment between Thinking Through Art and Boston Public School's Culturally and Linguistically Sustaining Practices (Boston Public Schools, 2021).

Thinking Through Art was developed by the Gardner using Visual Thinking Strategies (VTS), the pedagogy developed by Harvard-trained cognitive psychologist Abigail Housen and former Museum of Modern Art director of education Philip Yenawine. Housen and Yenawine began to collaborate on work that linked Housen's theory of aesthetic development (Housen, 2002) with Vygotsky's (1978) theory on learning through social interaction. Based on these theories Housen and Yenawine argued that "structured discussion among peers of art that intrigues them will produce observations, insights and exchanges that spur not only thorough, rigorous examinations of works of art but also significant skill development in individuals" (Yenawine, 1999). In 1993-1998, Housen designed and implemented a longitudinal study. The results of the five-year study supported their hypothesis that the VTS strategies for teaching and learning accelerates aesthetic growth. More pertinent to the purpose of this study, they also found evidence that VTS contributes to gains in students' critical thinking. Dozens of research studies have now shown that VTS supports the development of creative and critical thinking skills and that the gains transfer to subject areas other than art (DeSantis & Housen, 2007). A summary of study descriptions and findings is available on the Visual Thinking Strategies website, vtshome.org/research/.

Schools across the country, often in partnership with their local museums, have adopted VTS because of the ways it contributes to building teachers' capacity for facilitating student-centered inquiry instruction that result in gains in critical thinking and visual literacy. The Isabella Stewart Gardner Museum was among the early adopters and has been a leader among the nation's museums as it embedded VTS into its partnership with Boston Public Schools. In 2003, The Gardner Museum contracted with the Institute for Learning Innovation and with funds awarded by the U.S. Department of Education undertook a three-year study that focused on assessing the effects of participation in a program similar to Thinking Through Art on the development of elementary students' critical thinking skills. They used a control group research design and concluded that students who experienced the Museum's partnership program did have greater increases in critical thinking skills than did the students in the control groups (Adams, et al. 2006).

Another Gardner Museum study, conducted with funding from the Institute of Museum and Library Services, found that 8th grade Boston Latin School students participating in a one-year partnership with the Gardner tripled their critical thinking skills in oral language and doubled them in writing about an artwork

(DeSantis, 2009). A follow up of this study showed that these gains were retained as the students completed their senior year of high school, as shown in Figure 1 (Egan & Grohe, 2013). Another set of Thinking Through Art case studies was conducted during the 2017-18 and 2018-19 school years. These studies included students in middle and high school and found that across all grades, students' writing samples showed an 84% increase from pre-program to post-program in critical thinking skills (Isabella Stewart Gardner Museum, 2019). Furthermore, teachers and students felt more connected to the Gardner at the end of the Thinking Through Art school year and teachers saw the museum as a relevant cultural resource.

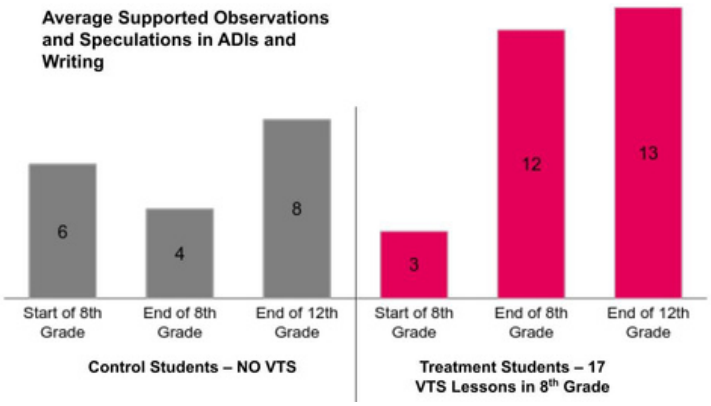

Figure 1: Student critical thinking skill growth in 2007-13 study of Boston Latin School (Egan & Grohe, 2013)

Aligning Outcomes for Teachers and Students with Boston Public Schools' Priorities

The outcomes of Thinking Through Art participation for this study have been defined and measured based on Boston Public Schools definitions and standards for culturally responsive instruction, social-emotional learning, critical thinking, and visual literacy.

Culturally Responsive Instruction

The Culturally Responsive Instruction Observation Protocol (Powell, et. al., 2017) adopted by Boston Public Schools operationalizes the district's commitment to Culturally and Linguistically Sustaining Practices. It calls for teachers to practice the six pillars of culturally responsive instruction. The Thinking Through Art professional development and curriculum explicitly supports CRIOP indicators, as outlined below in Table 1.

Table 1. Alignment between Thinking Through Art and the Culturally Responsive Instruction Observation Protocol (CRIOP)

CRIOP Pillar and CRI Indicator	**Thinking Through Art Components**
I. CLASSROOM RELATIONSHIPS	
The teacher demonstrates an ethic of care	Teacher explicitly begins each Thinking Through Art session by promoting an attitude that is safe and anxiety-free for all students, including culturally and linguistically diverse students; students seem comfortable participating in the Thinking Through Art lessons using the Visual Thinking Strategies (VTS) open-ended facilitated discussion method to structure lessons. Teacher refers to students by name throughout the Thinking Through Art lesson.
The teacher communicates high expectations for all students	Teacher expects all students to actively participate, responding to the same artwork (a visual text, that is both complex and decodable) using the same VTS prompts: What's going on in this picture? What do you see that makes you say that? What more can we find? Teacher introduces each Thinking Through Art lesson with the idea that we are all learners. Teacher uses sentence stems, translations, highlighting of cognates, and visual cues to ensure that emerging bilinguals understand directions and discussion content.
The teacher creates a learning atmosphere that engenders respect for one another and toward diverse populations	Teacher introduces each Thinking Through Art lesson with a reminder of respectful discussion norms. Teacher verbally links ideas that align and diverge. Students build on each other's ideas and consider each other's perspectives about the artwork. Teacher paraphrases all comments to model collaboration and respect, and to integrate students' lived experience and cultural knowledge into the classroom conversation. Thinking Through Art curriculum includes 8 image sets for each grade that align with the Enabling Text criteria and include positive and affirming images representing students' racial and ethnic identities. Teacher encourages students to share their stories and have pride in their history and linguistic and cultural identities.
Students work together productively	Students view each other as resources to interpret artwork. Students use each other as resources to interpret shared languages in cases when Teacher does not speak a shared language. Students have monthly Thinking Through Art discussions with peers.
II. FAMILY COLLABORATION	
The teacher reaches out to meet parents in positive, non-traditional ways	Teacher shares Isabella Stewart Gardner Museum resources with parents, including special family events and free family admission passes that allow 4 adults free admission throughout the school year and summer. All Thinking Through Art students receive letters for their families, translated into all 9 Boston Public Schools languages, about the program, the Museum, and how to redeem their free admission. Thinking Through Art staff from the Gardner Museum plan family activities at locations within the home community, as well as at the Gardner Museum

CRIOP Pillar and CRI Indicator	Thinking Through Art Components
II. FAMILY COLLABORATION (continued)	
The teacher encourages parent/family involvement	Parents are invited to participate in Thinking Through Art as chaperones for visits to the Museum. Thinking Through Art program materials include documents for chaperones about the visit and how to support learning.
III. ASSESSMENT PRACTICES	
Formative assessment practices are used that provide information throughout the lesson on individual student understanding	Thinking Through Art lessons act as informal formative assessment because students are asked to share their knowledge, understanding, and inferences on a range of subjects. Teacher immediately adjusts instruction based on student performance, for example integrating higher-tier vocabulary into paraphrasing or asking for more complex evidence. Students are able to voice their learning throughout the lesson by sharing how they revised ideas, as well as during reflective conversation or assignment after the lesson. The proposed study will use recordings of these classroom discussions as data to analyze the social and collaborative communication skills that are developed in Thinking Through Art.
Students are able to demonstrate their learning in a variety of ways	By design, Visual Thinking Strategies (VTS) encourages divergent responses and prompts students to share the processes and evidence they use to arrive at responses. Inherent to VTS is that there is no one correct answer. Students with limited English proficiency and/or limited literacy can show their conceptual understanding and learning orally or through visual representation. The proposed study aims to develop assessment methods that will allow all learners, including students with disabilities and English Learners, to demonstrate the depth and breadth of their critical thinking and understanding. This will include oral interviews conducted by research associates who share students' cultural backgrounds, as well as analyzing classroom discussions.
Authentic assessments are used frequently to determine students' competence in both language and content.	As part of Thinking Through Art, students complete a written analysis of an artwork at the beginning and end of the school year to assess their proficiency, while all other Thinking Through Art assessment is authentic, task-embedded Teacher observation of student performance. Teacher assesses both academic language and content. The proposed study will include a variety of assessments at the start and end of the school year to determine students' visual literacy, critical thinking, social-emotional development, and communication skills.
Students have opportunities for self-assessment	Teacher actively encourages students to evaluate their own learning through reflective activities before and after each Thinking Through Art lesson. Students generate criteria and goals, in conversation with Teacher.

CRIOP Pillar and CRI Indicator	**Thinking Through Art Components**
IV. INSTRUCTIONAL PRACTICES	
Instruction is contextualized in students' lives, experiences, and individual abilities	Thinking Through Art curriculum represents students' lives (acting as a window into students' worlds outside of school) and invites students to make connections to prior learning and their own lived experience. This curriculum of visual art is captivating and meaningful to students and promotes student engagement. All students are successful in the cognitively-demanding tasks of a Thinking Through Art lesson, including English Learners and Students with Disabilities.
Students engage in active, hands-on, meaningful learning tasks, including inquiry-based learning	Thinking Through Art lessons explicitly encourage exploratory learning where the teacher engages students in an inquiry process (i.e. asking "What's going on here?") and learns from students' investigations. Teacher probes for evidence by asking "What do you see that makes you say that?" and encourages further exploration by asking "What more can we find?" Students drive learning, with their questions becoming opportunities for close looking and/or further research into the artwork.
The teacher focuses on developing students' academic language	Teacher models academic vocabulary in context by paraphrasing and pointing to the details referred to by new terms. Teachers are also encouraged to scaffold language development by pre-teaching key academic vocabulary and sentence stems prior to the Thinking Through Art lesson. The repetitive structure of Visual Thinking Strategies (VTS) allows for students to hear key academic vocabulary and useful sentence structures many times during a VTS lesson.
The teacher uses instructional techniques that scaffold student learning	Teachers use visual art, which is a visual text, to engage all learners. Teacher models academic language through paraphrasing. The structure of a Thinking Through Art lesson, with the predictable sequence of a VTS discussion, allows students to enter from any point, and students gradually internalize the structure and the questions to support their thinking and participation. Teacher uses "comprehensible input" such as slowing down, using familiar language, and gesturing, to facilitate understanding. All contributions to the conversations are paraphrased orally and visually through pointing at the image.
Students have choices based upon their experiences, interests, and strengths	Students in high school grades select which 2 visual art images to discuss for each lesson, from a bank of 3-4. Students of all grades participating in Thinking Through Art lessons have ownership over their participation and learning because they decide what aspects of the artwork to discuss and probe.

CRIOP Pillar and CRI Indicator	**Thinking Through Art Components**
V. DISCOURSE	
The teacher promotes active student engagement through discourse practices	Thinking Through Art uses the Visual Thinking Strategies (VTS) structure of discourse, which is a whole-class conversation carefully facilitated by the Teacher. Teacher may also include pair-share or small group discussions, individual writing or drawing, or other strategies to ensure that all students are engaged.
The teacher promotes equitable and culturally sustaining discourse practices	VTS includes wait time before students begin to discuss the artwork and moments of wait time during the discussion. This gives all students an opportunity to organize their thinking and to have a personal interpretation of the artwork, while creating opportunity for those who process information at different speeds to all slow down. Teachers enforce turn-taking within the VTS structure and give equitable supportive feedback to each student through paraphrasing. Students build on each others' ideas, with Teacher explicitly naming moments of collaborative knowledge building. Students speak in their home language when appropriate, with students often translating for each other and developing multilingualism. Teachers are trained how to gradually scaffold participation to target language when appropriate.
The teacher provides structures that promote academic conversation	The structure of VTS includes: silent looking, repetition of key questions, paraphrasing and pointing, open-endedness, and appreciation. Students engage in genuine conversation without a predetermined endpoint. Teacher explicitly teaches and evaluates skills required for conducting effective academic conversations. Teacher provides prompts that elicit extended conversation and dialogue, using the VTS questions "what do you see that makes you say that?" and "what more can we find?" Teacher's open-ended facilitation encourages different points of view to emerge.
The teacher provides opportunities for students to develop linguistic competence	Thinking Through Art asks Teacher to lead Thinking Through Art lessons monthly, giving students ongoing opportunities to use academic language in this authentic context. In addition to classroom discussions facilitated by the Teacher, students visit the Isabella Stewart Gardner Museum at least twice during the academic year to transfer these academic conversations to a new setting. Students demonstrate the ability to extend their skills and behaviors beyond the classroom walls and with adults other than their Teacher, engaging in discourse in the Museum's galleries with trained Museum Teachers.

CRIOP Pillar and CRI Indicator	**Thinking Through Art Components**
VI. CRITICAL CONSCIOUSNESS	
The curriculum and planned learning experiences provide opportunities for the inclusion of issues important to the classroom, school, and community	Thinking Through Art curriculum includes visual art related to community-based issues such as civil rights, migration, poverty, racism, and humans' role in climate change. The curriculum also relates to individual concerns such as identity, interpersonal relationships, definitions of family, and gender roles. In all cases, these issues are presented through visual art that is developmentally- and age-appropriate. Students are engaged in deep conversations about these artworks, exploring real-world issues important to them and their communities. All artworks work as Enabling Texts, providing positive, interesting, complex, provocative depictions of characters that connect to issues that students find essential today, recognize and nurture multiple identities, demonstrate resilience and reflect the human condition.
The curriculum and planned learning experiences incorporate opportunities to confront negative stereotypes and biases	Thinking Through Art explicitly trains teachers how to confront negative stereotypes and biases during a Thinking Through Art lesson, disrupting these harmful narratives and examining their roots. Teachers are trained to further students' understanding of stereotypes and biases, and to examine how they are replicated in the media and texts. The Thinking Through Art curriculum intentionally includes visual art from diverse cultures to facilitate conversations about human differences and commonalities.
The curriculum and planned learning experiences integrate and provide opportunities for the expression of diverse perspectives	Thinking Through Art trains teachers to respond to all comments without judgment, allowing for divergent ways of thinking and diverse perspectives to be considered. (The only exception to this is comments containing bias or stereotypes, which are disrupted.) Students are encouraged to respectfully disagree with each others' arguments, providing evidence to support their views. The protagonists of the artwork in the Thinking Through Art curriculum are from diverse backgrounds.

Social-Emotional Learning

Boston Public Schools places social-emotional learning among its top strategic priorities. District documents define this aspect of instruction and learning as "a process whereby young people, and adults build strong, respectful, and lasting, relationships that facilitate co-learning to critically examine root causes of inequity, and to develop collaborative solutions that lead to personal, communal and societal well-being" (Boston Public Schools SEL Competencies & Skills, 2019).

Social-emotional learning is an integral part of the school district's Multi-Tiered Systems of Support, a model for integrating academic and behavior instruction (MA DOE, 2020). Relationship building through co-learning, empathy and conflict management skills are all components of a Tier 1 system of support that helps students to manage emotions, maintain positive relationships and make responsible decisions. Co-learning through conversation is also a keystone of the VTS method (Yenawine, 1999), used in Thinking Through Art instruction and curriculum and is a strong example of the Tier 1 instruction that Thinking Through Art provides by creating a learning environment that because embeds academic instruction and learning in a social contest.

Critical Thinking

Housen first developed an instrument to measure critical thinking, defined as supported observations and speculations, as an outcome of Visual Thinking Strategies (Housen, 2002). The Adams et. al. 2006 study at the Gardner Museum focused on the critical thinking outcomes of the museum's school programs, drawing on the foundations of Costa & Kallick's (2009) ground-breaking work about teaching habits of mind - a term that pre-dates the current use of "critical thinking" as a learning outcome (Adams et al., 2006). Most recently, the Gardner Museum has used an instrument informed by both of these studies to assess critical thinking in Thinking Through Art students' responses to works of visual art, see figure 2 (Egan, 2020). These outcomes correspond seamlessly with current Boston Public Schools' frameworks for considering critical thinking. Webb's Depth of Knowledge levels place skills such as synthesis and proving claims in the highest stage of critical thinking (Webb, 2005). The Think CERCA Framework identifies four key critical thinking skills: develop a claim in response to an essential question, look for credible evidence to support the claim, connect evidence to claims with reasoning, and consider other points of view (ThinkCERCA, 2022). Furthermore, the CERCA Framework aims to help students use, not only written texts but also "content-rich texts and multimedia. Thinking Through Art is grounded in examination of content-rich works of art (visual texts). Thus, Thinking Through Art students are instructed in the evaluation of visuals and their sources.

CRITICAL THINKING SKILL	Definition with Example	Tally of Uses by Student
Supported Observation	Comment about the image/object that is grounded in evidence. *"I think they're in the same room because they have the squares on the back."*	
Speculation	Considering something is possible; often uses conditional language such as "maybe, possibly, could be, might be." *"And they might seem like that even though they're different colors, they could be together."*	
Revision	Changing opinion about previously made statement *"The woman has, well, it looks like it has on a strange, or like, a strange outfit. Well, not strange, but uncommon for, like, for this day and age."*	
Multiple Possibilities	Acknowledges that more than one idea is plausible or recognizes different perspectives; often uses "or" to delineate the various options. *"So the back seems little, either discolored or it's reflecting from some kind of light."*	
	TOTAL CRITICAL THINKING SKILLS USED	

Figure 2: Isabella Stewart Gardner Museum Critical Thinking Skills Rubric (Egan, 2020)

Access to Cultural Resources

Programs like Thinking Through Art close opportunity gaps, giving urban students access to cultural organizations. In SY 2018-19, the last year not impacted by COVID-19, the Gardner welcomed over 3100 students on 156 class visits, and an additional 330 students and their families used the free admission passes to explore the Gardner on their own. Demonstrating the impact of these experiences, surveys from 2017-19 show a 65% increase in how many students feel connected to the Isabella Stewart Gardner Museum by the end of one year of Thinking Through Art (see Figure 3).

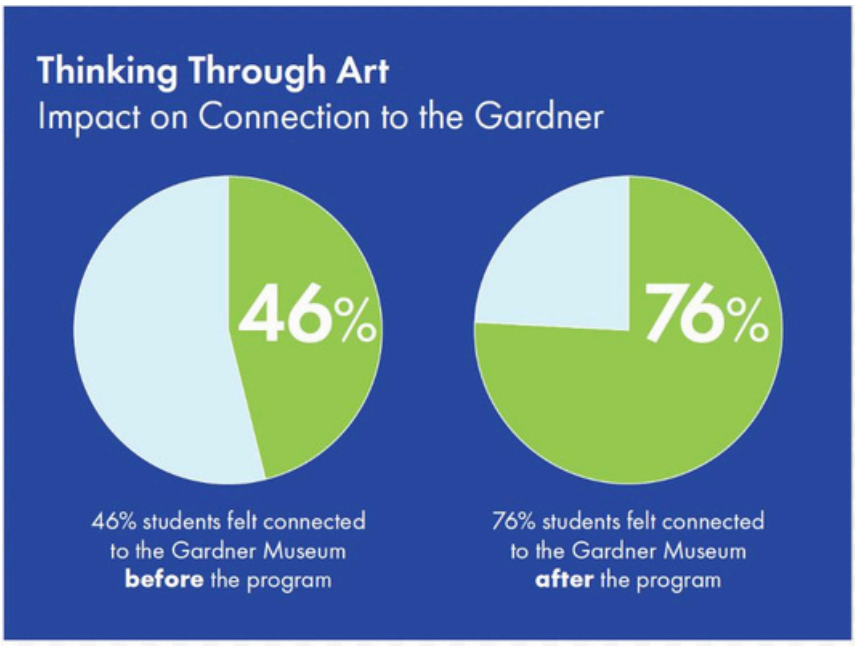

Figure 3: Thinking Through Art student data, school years 2017-19

3. STUDY DESIGN

Working with Boston Public Schools

Our research team underwent Boston Public Schools' process to apply for permission to do research with its teachers and students in Summer 2021. We had initially drafted a research proposal that primarily focused on the Thinking Through Art program's impact on the Gardner Museum's education department goals. When this first proposal was rejected by the Boston Public Schools Office of Data and Accountability, we met with members of the Data and Accountability team to discuss ways to make our research better aligned with the goals of Boston Public Schools, as well as the Gardner Museum. These meetings helped our team begin to clarify our research questions, and to think both more broadly and more deeply about how the Thinking Through Art program functions as part of multiple overlapping education ecosystems. We met with members of the Boston Public Schools English Language Arts team to discuss the district's Equitable Literacy initiatives, and to look for crossover between Thinking Through Art's program goals and these initiatives. We incorporated a number of tools used widely in Boston Public Schools into our instrument and analysis manual design processes, and sought regular feedback and counsel from current Boston Public Schools teachers.

We incorporated feedback and tools from our partners at the district and restructured our application to conduct research to center the district's goals for both teacher practice and student outcomes. By reframing our thinking about the bigger picture goals of museum and school partnerships, and putting our research queries into a framework that demonstrated how the project would support high quality teaching and learning in the district, we were able to get our second application to conduct research in Boston Public Schools approved in October 2021. We also successfully applied for an IRB exemption at this time.

With these approvals secured, we recruited a group of five Boston Public Schools teachers who had participated in Thinking Through Art in the past to provide counsel and to help develop and pilot test study instruments and frameworks. This group of teachers, our "Research Advisory Group" collectively worked with students from grades K-12, taught a diverse array of students including multilingual learners, students with disabilities, and general education students across all Boston Public Schools demographic groups. Between March and June 2022, the research team met with the Research Advisory Group three times to discuss the development of research questions, instruments, and frameworks. Research Advisory Group teachers were given specific pilot testing tasks to implement in their classrooms and due dates for their feedback (via Google Form) on the processes and instruments they pilot tested. The research team also conducted instrument pilot testing in these teachers' classrooms in the Spring of 2022.

Research Questions & Variables

In short, our research team sought to holistically evaluate the impact of the Thinking Through Art program on both the teacher and student participants. By investing in year-long, high-quality professional development for teachers, the Gardner Museum shares a goal with Boston Public Schools to support and improve teachers' culturally responsive instruction; we aimed to determine whether participation in Thinking Through Art improves teachers' skills in this area more than control group teachers' skills improve over the course of a year. We also asked what the impact of participating in Thinking Through Art was for students, in both social-emotional and cognitive domains. Seeking to replicate past research findings on students' increased critical thinking, we asked whether participation in Thinking Through Art improves students' general and specific critical thinking skills more than control group students' skills improve in this domain. Finally, we wanted to determine whether anecdotal evidence from past teacher participants about the program's positive impact on students' social-emotional learning would be visible in qualitative and quantitative data. We asked whether participation in Thinking Through Art improves students' social-emotional learning more than control group students' social-emotional learning. In Table 2, see a detailed description of our research questions, as well as the experimental, condition, and dependent variables in this project.

Table 2. Research Questions and Variables in *Thinking Through Art: A Transformative Museum-School Partnership* Study

RESEARCH QUESTION 1: Does participation in Thinking Through Art increase teachers' skills in Culturally Responsive Instruction?	
Experimental Variables: • Thinking Through Art training / no Thinking Through Art training • For the Thinking Through Art (treatment) group, dosage of training and of practice in the classroom	**Dependent Variables:** • Teachers' total score at the end of the school year for culturally responsive instruction • Degree of change from the beginning to the end of the year for culturally responsive instruction • Teachers' score for each category of culturally responsive instruction at the end of the school year (categories include: classroom relationships, instructional practices, student-centered teaching, and culturally responsive discourse) • Degree of change from the beginning to the end of the year for each category of culturally responsive instruction

RESEARCH QUESTION 2: Does participation in Thinking Through Art increase students' skills in Social-Emotional Learning?

Experimental Variables:

- Thinking Through Art lessons led by a trained teacher (treatment) / no Thinking Through Art lessons, training or curriculum (control)
- Focus image for study (art image or narrative/photojournalistic image)
- For the Thinking Through Art (treatment) group, dosage of Thinking Through Art lessons in the classroom

Condition Variables:

- Student race/ethnicity
- Students' multilingual learner status (English Language Development levels 1-2, 3-4, 5-6-FLEP)
- Students' disability status (no disability / disability categories: autism, communication impairment, developmental delay, emotional impairment, health impairment, multiple disabilities, neurological impairment, physical impairment, sensory impairment, specific learning disability)
- Student grade level (PreK-12) and grade band (PreK-2, 3-5, 6-12)
- Teacher scores at the end of the school year for culturally responsive instruction

Dependent Variables:

- Students' total score at the end of the school year for social-emotional learning
- Degree of change from the beginning to the end of the year for total social-emotional learning
- Students' score for each category of social-emotional learning at the end of the school year (categories include: possesses self confidence and belief in capacity to learn, demonstrates relationship skills, expresses empathy, respects multiple perspectives, and contributes to creating a community of learners)
- Degree of change from the beginning to the end of the year for each category of social-emotional learning

RESEARCH QUESTION 3: Does participation in Thinking Through Art increase students' skills in Critical Thinking?

Experimental Variables:

- Thinking Through Art lessons led by a trained teacher (treatment) / no Thinking Through Art lessons, training or curriculum (control)
- Focus image for study (art image or narrative/photojournalistic image)
- For the Thinking Through Art (treatment) group, dosage of Thinking Through Art lessons in the classroom

Condition Variables:

- Student race/ethnicity
- Students' multilingual learner status (English Language Development levels 1-2, 3-4, 5-6-FLEP)
- Students' disability status (no disability / disability categories: autism, communication impairment, developmental delay, emotional impairment, health impairment, multiple disabilities, neurological impairment, physical impairment, sensory impairment, specific learning disability)
- Student grade level (PreK-12) and grade band (PreK-2, 3-5, 6-12)
- Teacher scores at the end of the school year for culturally responsive instruction

Dependent Variables:

- Students' total score at the end of the school year for critical thinking
- Degree of change from the beginning to the end of the year for total critical thinking
- Students' score for each category of critical thinking at the end of the school year (categories include: identifies information, makes associations, constructs meaning, considers multiple perspectives, and exhibits reasoning)
- Degree of change from the beginning to the end of the year for each category of critical thinking

RESEARCH QUESTION 4: Does participation in Thinking Through Art increase students' skills in Social-Emotional Learning for both art images and narrative (photojournalistic) images?

Experimental Variables:	Dependent Variables:
• Focus image for study (art image or narrative/photojournalistic image)	• Students' total score at the end of the school year for social-emotional learning • Degree of change from the beginning to the end of the year for total social-emotional learning • Students' score for each category of social-emotional learning at the end of the school year (categories include: possesses self confidence and belief in capacity to learn, demonstrates relationship skills, expresses empathy, respects multiple perspectives, and contributes to creating a community of learners) • Degree of change from the beginning to the end of the year for each category of social-emotional learning

RESEARCH QUESTION 5: Does participation in Thinking Through Art increase students' skills in Critical Thinking for both art images and narrative (photojournalistic) images?

Experimental Variables:	Dependent Variables:
• Focus image for study (art image or narrative/photojournalistic image)	• Students' total score at the end of the school year for critical thinking • Degree of change from the beginning to the end of the year for total critical thinking • Students' score for each category of critical thinking at the end of the school year (categories include: identifies information, makes associations, constructs meaning, considers multiple perspectives, and exhibits reasoning) • Degree of change from the beginning to the end of the year for each category of critical thinking

Study Images

The research team chose two images to use in this study: one work of figurative art, and one narrative photojournalistic image (Figure 4 and 5, below).

Figure 4: Art Image

*F*ernand Leger (1881-1955), *Les Loisires–Hommage à Louis David,* 1948-1949
© 2024 Artists Rights Society (ARS), New York / ADAGP, Paris
Digital Image © CNAC/MNAM, Dist. RMN-Grand Palais / Art Resource, NY

Figure 5: Narrative Image

D. Gordon, *Untitled,* 1978
Used with permission of the artist.

Using the guidelines for Image Selection for Beginning Viewers published by the Visual Thinking Strategies (VTS) organization VUE (Visual Understanding in Education), we selected the two images for the study to use with both the Thinking Through Art (treatment) and control groups, across all grades from PreK-12 (VUE, 1998). The VUE criteria included the following considerations:

- Images should be accessible, so that beginning viewers can recognize elements in the image and make sense of them
- Images should contain expressive content, which allows them to be open for interpretation, with several possible valid readings or meanings
- Images should be narrative, so that viewers can partake of developmentally appropriate instincts to find or create stories about what they see
- Images should be stylistically realistic, which helps viewers access their elements smoothly

Approximately two-thirds of the Thinking Through Art (treatment) group (25 classes out of 38 total) and the control group (10 classes out of 17 total) looked at, discussed as a whole class, and responded independently to the art image. The other third of the Thinking Through Art (treatment) group (13 classes out of 38 total) and the control group (7 classes out of 17 total) used the narrative image for the same observation, discussion, and response tasks.

Sequence of Pilot Testing, Data Collection, and Data Analysis

The planning and data collection phases of this study were scheduled across two school years. Phase 1: Pilot Testing (January - June 2022) consisted of the work of the research team and a Research Advisory Group made up of experienced Thinking Through Art teachers to pilot test all of the research procedures, data collection instruments, and initial versions of rubrics for data analysis coding. During Phase 2: Implementation and Data Collection (September 2022 – June 2023) the research study was implemented and data was collected from fifty-five Thinking Through Art and control group teachers and their students. Phase 3: Data Analysis (June 2023 – April 2024) consisted of coding teacher and student data, statistical analyses, and writing this report.

PHASE 1: Pilot Testing Instruments and Procedures (January - June 2022)

During the January - June 2022 Pilot Testing Phase, five experienced Thinking Through Art teachers in the Research Advisory Group and their students occasionally used all, or a part, of the study procedures and data collection instruments with the purpose of providing feedback about their clarity, ease of implementation, adaptations for students with special needs and capacity to produce data that are valid and therefore, tested for their veracity in addressing the research questions. The research team used the Pilot Testing Phase for refining procedures and methods so that they met high standards for validity (they measure what they are intended to measure) and reliability (the consistency with which instruments are administered and inter-rater consistency when applying observation protocols and coding data).

The research team members worked closely with Research Advisory Group teachers during this phase, meeting three times over zoom as a group to discuss the utility of instruments and ease and effectiveness of procedures. Members of the Research Advisory Group made their own decisions about how and when they would seek feedback from students in their classes without interfering with classroom procedures or instruction; the goal was for feedback about instruments and procedures to be integrated into already planned instruction and classroom activities. Our team coordinated the schedule and teacher assignments for pilot testing tasks and made revisions that were pilot tested iteratively. For some pilot testing tasks, members of the research team visited the Research Advisory Group teachers' classrooms to pilot test video and audio recording procedures. For others, the teacher administered versions of instruments to their students without members of the research team present.

Student privacy and rights were honored during the pilot testing phase; student names were not requested, and no list of student participants was created. The purpose of the pilot testing data was to examine the strength of the instruments themselves, not to analyze the content of the students' responses. Thus, the questions that were addressed were different from the research questions for the 2022-2023 Data Collection year (Phase 2). The questions to guide pilot testing information gathering and analysis were:

- Was the choice of words, length and format of student survey and instructions for reflecting on a work of art or narrative image appropriate for different grade levels, multilingual students (English language learners) and students with disabilities?
- What augmented and alternative communication methods and devices needed to be created, tested and made available in order for all students to comfortably participate in the study?
- Did the proposed study procedures for gathering data and videotaping classroom discussion minimize burden on teachers, disruption of the classroom, and/or interruption in the school day?

The research team and Research Advisory Group pilot tested the following in preparation for their use during the 2022-2023 Data Collection (Phase 2):

- Methods of distributing forms and receiving parent/guardian written consent
- Prompt for the individual student response activity
- Selection of images for group discussion and individual student response activity
- Methods for audio recording a student discussing an image
- Methods for video recording classroom discussions
- Methods for administering student survey
- Student survey design and content (including grade level and language versions)

PHASE 2: Data Collection (September 2022 - June 2023)

Implementation of the Thinking Through Art program and treatment and control group data collection took place during the 2022-2023 school year. Teachers enrolled in the 2022 Thinking Through Art Summer Institute and their students participated in the study as members of the treatment group. A control group of teachers was recruited by the Thinking Through Art teachers and members of the Thinking Through Art program team (see more on study participants in Section 5, below).

Beginning of year, or pre, data was collected from September through December 2022, first in the Thinking Through Art classrooms, then in the control group classrooms. During the first months of 2023, the research team began inputting data (video and audio recordings, scanned student responses and surveys) from the beginning of year data collection sessions into a secured drive, transcribing written and audio recorded student responses, and tabulating student survey responses. We also began sending responses in languages other than English to professional translators for transcription into English. Our team began scheduling data collection sessions for the end of year, or post, data in February 2023. The post data

collection sessions began in March 2023 and continued through the end of the school year in June 2023. Data from the post sessions was input, transcribed, and translated into English throughout this period, enabling the team to begin the analysis process immediately upon data collection completion. During Phase 2, the research team was also pilot testing and revising early versions of the data analysis manuals in preparation to train the research assistants in coding, rating, and analysis beginning in June 2023.

PHASE 3: Data Coding and Analysis (June 2023 - April 2024)

Throughout Phase 3, the core research team members revised and tested the various data coding manuals, then created agendas for multi-day coding training "marathons" where we taught other members of the research team how to use the manuals to code either the student responses (for critical thinking, using the Students' Critical Thinking Coding Manual), student behaviors (for social-emotional learning, using the Students' Social-Emotional Learning Coding Manual), or teacher behaviors (for culturally responsive instruction and facilitation skills, using the Teachers' Culturally Responsive Instruction and Facilitation Skills Ratings Manual). Different groups of members from the research team worked independently and with partners or in small groups to code or rate different types of data for the study. The table below lays out the schedule of Data Coding and Analysis in Phase 3, and includes which members of the team worked on each type of data coding and subsequent analysis. The Core Research Team refers to Ms. Egan, Ms. Munley, and Ms. Tratnyek. The Research Assistants included Lead Researcher Ms. Tratnyek and a team of part-time researchers, some of whom were hired for this study and some of whom are also Museum Teachers at the Gardner Museum.

Table 3. Data Analysis Timeline and Personnel

Type of Data Coding / Analysis	Timeline	Task Description and Personnel
Data Coding: Critical Thinking	June 2023 - October 2023	**Core Research Team:** led manual training, administered inter-rater reliability checks, led weekly coding meetings **Research Assistants:** attended manual training, participated in regular inter-rater reliability checks, coded student responses for critical thinking skills
Data Coding: Teacher Facilitation Skills	June 2023 – October 2023	**Research Assistants:** rated teachers' facilitation skills using rubric developed by the core research team for the study and in use by members of this team since 2020.
Data Coding: Student Survey	June 2023 – August 2023	**Research Assistants:** tabulated student survey response data

Type of Data Coding / Analysis	Timeline	Task Description and Personnel
Data Coding: Student Social- Emotional Learning Skills	October 2023 – December 2023	**Core Research Team:** led manual training, administered inter-rater reliability checks, led bi-weekly coding meetings **Research Assistants:** attended manual training, participated in regular inter-rater reliability checks, coded video-recorded student behaviors for social-emotional learning skills
Data Coding: Teacher Culturally Responsive Instruction	January 2024 – March 2024	**Core Research Team:** led manual training, administered inter-rater reliability checks, led bi-weekly coding meetings **Research Assistants:** attended manual training, participated in regular inter-rater reliability checks, coded video-recorded teacher behaviors for culturally responsive instruction
Data Analysis: All Collected Data Types	January 2024 – April 2024	**Core Research Team:** conferred with the Statistical Analyst to plan a framework for a variety of types of quasi-experimental data analysis, on multiple types of student and teacher data, and putting these different types of data into conversation with one another **Statistical Analyst:** Cleaned, scaled, and statistically analyzed teacher practice, student social-emotional learning, and student critical thinking data using OLAP Cubes and Cross tabulations; determined statistical significance or lack thereof using Chi-Squares, Mann-Whitney U and Pearson R evaluations.

4. ETHICAL CONSIDERATIONS

The research team underwent an IRB exemption review process, and applied to conduct research with teachers and students in a school setting with Boston Public Schools, both in 2021. The IRB exemption and permission to conduct research were both granted by the end of 2021, before the pilot testing phase was scheduled to begin. All members of the research team underwent a CORI background check and were therefore granted permission from Boston Public Schools to enter classrooms for data collection sessions throughout the 2021-2022 and 2022-2023 school years.

As this study evaluated students who are almost all children under the age of 18 (pre-Kindergarten through grade 12), their names, likenesses, and other identifiable information all must be considered "private" and protected by the research team and the students' teachers. Students and their parents or guardians received a consent form at the beginning of the school year, which included information on the impact study, the instruments and data to be collected. The form was available in multiple languages to meet the Boston Public Schools student/parent population's needs. Parents (and high school students aged 18 and older) were asked to provide written consent separately for participation in the impact study broadly (including audio-recorded responses and survey responses), and participation in video-recorded discussions. Researchers also asked for student assent at the time of data collection.

Study data was delinked-anonymized (identifiers collected, link obliterated); the research associate assigned a unique alpha-numerical identifier to each student participant. Thinking Through Art and control group teachers were responsible for labeling student-generated data (audio-recordings, surveys, etc.) with the alpha-numeric identification codes, and removing student names before the research team accessed the data. Thinking Through Art and control group teachers were briefed on how to do this, and the research team could assist teachers with work labeling and anonymizing during data collection sessions. Thinking Through Art teachers signed a data security agreement agreeing to these protocols as a requirement for participation in the Thinking Through Art Program (for Thinking Through Art teachers) and the Impact Study (for the control group teachers).

All data were transported to the Gardner museum directly by members of the research team or through pre-paid USPS mail by Thinking Through Art and control group teachers. All student and teacher records were saved in secure, password protected spreadsheets by the Gardner Museum education department. All written, survey, audio-recorded, video-recorded, and other data for this study were stored in a secure drive in the Gardner Museum education department. Content stored in this drive is restricted to be viewed only by members of the research team. All video and audio recordings were transferred immediately from the device that took them to the project computers for storage and deleted from the original recording devices. All of the files and logs will be destroyed after the three-year required period for keeping study records expires in May 2027.

5. PARTICIPANTS IN *THINKING THROUGH ART: A TRANSFORMATIVE MUSEUM-SCHOOL PARTNERSHIP* IMPACT STUDY

School District and Study Universe Overview

Boston Public Schools is a moderately large urban public school district, the 93rd largest in the United States as of 2019. In the 2022-2023 school year, when the our team conducted the Thinking Through Art Impact Study, the district educated 46,269 students, most of whom qualify as both "high needs" and "low income" according to the Commonwealth of Massachusetts (Department of Elementary and Secondary Education, 2022 District Report Card). Teachers in Boston Public Schools are almost universally licensed, and more than eighty percent of teachers in the district have three or more years of teaching experience.

55 teachers from 28 schools in the Boston Public Schools district participated in the Thinking Through Art Impact Study during the 2022-2023 school year. The treatment group consisted of 38 teachers who had enrolled in the Thinking Through Art program at the Gardner Museum, and then opted into the Impact Study treatment group (five teachers who were also enrolled in Thinking Through Art either opted out or were ineligible to participate in the study). Thinking Through Art teachers are generally motivated and experienced educators with at least a few years of classroom teaching, and all come from Boston Public Schools. Seventeen teachers participated in the study as members of the control group. To assemble the control group, the research team first reached out to teachers who had expressed interest in Thinking Through Art in the past, but who had never attended the professional development sessions. We also reached out to other teachers at the schools where prior and current Thinking Through Art teachers worked to recruit control teachers with similar student populations, by virtue of being at the same schools. Some Thinking Through Art teachers recruited interested coworkers, as well. In the end, the 55 teachers in the treatment and control groups included 21 early childhood teachers (working with students in preschool, kindergarten, first, and second grade), 23 upper elementary teachers (working with third, fourth, and fifth grade students), and 11 secondary teachers (working with sixth–eighth or ninth–twelfth graders).

Both Thinking Through Art and control group teachers provided the research team with anonymized student demographic data for all students in their class at the start of the school year (note that secondary teachers and specialists who worked with multiple classes selected one group to participate in the study); the total universe of student participants in this study was 942 across the fifty-five teachers' classrooms. Of these students, 714 provided consent to participate in the study as respondents. For critical thinking analysis, the research team sampled responses from 388 students from this group of students with consent.

Table 4 below shows that the total universe of student participants is closely aligned demographically to the district as a whole. About 76% of all the students in the universe of the study (including 79% of treatment group students and 70% of control group students) gave consent to participate in the study, and to have their responses collected. Tables 5 and 6 show the grades in the study universe.

Table 4. Comparing Student Demographics in Boston Public Schools and Treatment and Control Groups

Percentage of students in Race/Ethnicity and Selected Population Category	School District	Student Universe for the Thinking Through Art Impact Study	
	Boston Public Schools Students	Treatment Group (Thinking Through Art Students)	Control Group
African American and/or Black	28%	28%	26%
Asian American	9%	9%	4%
Hispanic	44%	43%	45%
Multiple Races (Non- Hispanic)	4%	2%	7%
Native American	1%	0.6%	0%
White	15%	14%	11%
Other / Blank	N/A	3%	7%
Multilingual Learners *	32%	38%	31%
Students with Disabilities	22%	19%	15%

* More than 74 different languages are spoken by Boston Public Schools students, as of 2024. We counted students as Multilingual Learners (or "English Language Learners") if they had been tested and given an English Language Development Level (ELD), or if they had once been classified as having an ELD level in an earlier year of school. During the 2022-2023 school year when our team collected data, we had all our instruments translated into the top nine languages spoken by students in the district (Arabic, Mandarin Chinese, English, French, Haitian Creole, Portuguese, Spanish, Somali, and Vietnamese), as well as Cape Verdean Creole, which multiple participating Thinking Through Art teachers requested for their school community. We had student responses professionally transcribed and translated from American Sign Language, Spanish, Portuguese, and Vietnamese into English for this study.

Table 5. Student Universe by Grade Band and Special Populations

		Pre-Kindergarten - Grade 2	Grades 3 - 5	Grades 6 - 12
Total Number of Students	Treatment	233	295	92
	Control	135	107	80
Multilingual Students	Treatment	65	115	55
	Control	41	42	16
Students with Disabilities	Treatment	26	73	17
	Control	24	23	0

Table 6. Teacher Universe by Grade Band and Special Populations

		Pre-Kindergarten - Grade 2	Grades 3 - 5	Grades 6 - 12
Total Number of Teachers	Treatment	13	17	8
	Control	8	16	3

6. DATA COLLECTION METHODS

The Gardner Museum's partnership with the Thinking Through Art teachers and the control group teachers was an essential component of this study. We had to convey clearly and concisely what we needed them to do ahead of the research team visits to their classrooms to collect data, and to make these preparatory tasks as simple, unobtrusive, and easy as possible. It was a boon to have former classroom teachers on the research team, and to have teachers in the Research Advisory Group available to weigh in on expectations for teachers and students before the data collection phase began.

We created a "Teacher's Guide to the Impact Study" with a timeline for the study (including due dates for specific tasks), an overview of the study design and of each instrument, a guide to the data collection sessions (including scripts for teachers to use to set up the lesson and administration of instruments), and detailed instructions on creating an anonymized list of their students so that they could share student demographic data with the museum without sharing any student names with members of the research team.

Teachers' first task was to create a list of students with demographic data, information on students' status as Multilingual Learners, and/or their disability categories, with the students' names redacted on the version of this list the teachers shared with the research team. Teachers next submitted their own consent to participate in the study, and collected consent to participate forms from their students. Teachers recorded information about consent on the student lists, then submitted them (without students' names) to the research team. The research team provided teachers with unique 4-digit student identification numbers for each student in their class. Teachers were the only individuals able to connect students by name to their assigned identification number. This procedure ensured that students were anonymous to the research team, yet linked to their data through their identification number.

Throughout the first data collection period in fall 2022, members of the research team reached out to each participating teacher to schedule a data collection session. For Thinking Through Art teachers, it was imperative that these sessions be scheduled as soon as possible, before they began teaching the Thinking Through Art lessons in October, so that it would truly represent their beginning of year baseline. As the teachers in the control group were receiving no intervention from the Thinking Through Art program, we were able to schedule their data collection sessions later in the fall of 2022.

A data collection session required at least two Research Assistants, and ideally three or more Research Assistants for classes in Pre-K through second grade. A larger team of researchers allowed for a high ratio of adults to students in the lower grades, so there was plenty of support for students in the room during the data collection session. Research Assistants traveled to the school, bringing with them all the materials needed for the session. Teachers were sent emails the week before the data collection session with a link to the Study Images slide deck and reminders to have their copy of their class list ready for researchers to use to connect student ID#s to individual students correctly.

Upon arrival at the classroom, Research Assistants introduced themselves to the class and then set up two small video recorders on tripods, one angled to capture the teacher and the study image (either projected with the standard-issue projector all Boston Public Schools teachers have access to, or broadcast on a large TV screen, also provided by the district to all teachers) and the other angled to capture the students. We also used audio recorders throughout the class discussions to pick up supplemental audio, as students in the classroom often spoke very quietly, or were masked and had muffled voices. The participating teacher (and, if applicable, their support staff, which could include other teachers and paraprofessionals) helped get the students set up for a whole class discussion, and provided students who had not provided consent to participate in the study with brightly colored stickers so that the research team could easily identify them and not code their comments and behaviors during the data analysis phase. All students were able to participate in the data collection session, including the facilitated class discussion, the student survey, and the written or recorded student response, regardless of whether or not they had provided consent to participate; the research team simply destroyed data from students without consent to participate after the collection session, and did not code their words or behaviors throughout the video recorded discussion. The goal for data collection sessions was for the lesson to run as it would without a team of researchers in the room, with the members of the research team behaving as unobtrusively as possible, and all students participating in classroom activities.

The teacher decided whether to administer the student survey before or after the facilitated discussion about the art or narrative image, but all data collection sessions began with one of these two activities. The student survey took 5-15 minutes to administer, depending on the age of the students and how much support they needed to answer the nine questions. The classroom discussion was facilitated by the teacher, who was instructed either to "facilitate a student-centered whole group discussion about an image for 12-15 minutes" (directions for control group teachers) or to lead a Thinking Through Art lesson using Visual Thinking Strategies (directions for treatment group teachers). One Research Assistant filmed the teacher and one filmed the students; at least one Research Assistant also recorded the lesson on an audio recorder. Because we were interested in capturing the entire classroom ecosystem during these lessons, we left the cameras in one place and generally did not zoom in or out or change the camera angles once filming began. Once teachers reached the 12-minute mark, a Research Assistant gave them a signal to wrap up in the next three minutes. After video and audio recording the classroom lesson, the classroom teacher transitioned students either to the survey if they hadn't done it yet or to the student response activity. In classes where most or all of the students were most comfortable writing their responses to the image they had just discussed, the research team was able to put away the recording equipment and help the teacher label and anonymize student work (using printed labels with students' ID#s) while students wrote their responses. In classes where some or all students were not yet able to express, or less comfortable expressing themselves in writing, the Research Assistants and the teacher(s) in the room worked with students individually to either audio record or scribe their responses.

The research team always deferred to the teachers about how much or little support their students needed for each of the data collection activities. Research team members worked with students across all grades who could benefit from the ability to speak instead of write responses (the classroom teachers made

these recommendations) by taking them aside, speaking their ID# into the recorder so their responses could be identified, and saying the prompt "What's going on in this picture?". Students were able to speak into the recorder for up to five minutes. Students who wrote responses were given up to fifteen minutes of writing time. All students were encouraged to write or speak in the language that was most comfortable for them, and our team had their data transcribed and translated professionally.

Researchers used the following templates to collect and organize class lists and student identification numbers:

- Class List
  docs.google.com/spreadsheets/d/1A0OBmvLfxGAYJB9hHq6DUlTS1E6MeBo1bq3vfgqdh0/copy
- Control Class List ID#s + Tracking
  docs.google.com/spreadsheets/d/1snQRGcmC-UKFbnXTl6YsfGpo6J1mGMytVz_3nRy4wy4/copy
- Treatment Class List ID#s + Tracking
  docs.google.com/spreadsheets/d/1nOyIeISEUQq3Ixb-yZCHFN5-1lHfy3vYfYyiHud2SUk/copy
- Data Collection Tracking Master Spreadsheet
  docs.google.com/spreadsheets/d/124H5TRAnpE7AE3OWI6076at_B5ZaiSD-CinKXfki-S8/copy

Student Survey

The student survey is divided into three sections: "What I Think About Myself", "How I Learn", and "My Classroom". Each section contains three questions, all of which were designed to be answered using a labeled five-point Likert scale (1= Never, 2 = A few times, 3 = Sometimes, 4 = A lot of the time, 5 = Almost always). Students were directed to choose their answer and circle it on the form. Students who needed additional support were able to complete the survey with the help of a Research Assistant.

Student Response

The student response form simply has the prompt, "What's going on in this picture?" and lines on the front and back of the sheet for drawing or writing. Researchers can easily mark whether the student is responding using an audio recorder, or whether an adult has scribed their response.

Classroom Lesson Video Recording

Classroom lesson video recordings were saved as AVI files, and labeled with the help of some analog technology: video "clapperboards" held in front of the camera by members of the research team to identify the teacher, grade band, school, and date. Audio recordings were saved as mp3 files.

7. DATA CODING AND ANALYSIS METHODS

Our team sought to answer research questions about both teachers' and students' end-of-year outcomes, and their growth from the beginning to the end of the year, across a number of different categories of analysis. To evaluate teacher practice, we developed a survey and two coding manuals (one for teachers' culturally responsive instruction and one for teachers' facilitation skills). To evaluate students' social-emotional learning, we developed a survey (to evaluate students individually) and a coding manual (to evaluate the social-emotional learning present across all members of each classroom's student community). To evaluate students' critical thinking, we developed an instrument to capture students' individual responses to a study image. Students could write or type their responses, or have their responses be taken by a researcher through one-on-one audio recording / scribing, or use another type of adaptive technology to record their responses.

Description of Statistical Analyses with Examples

Chi-square

A chi-square test is a statistical test that is used to compare observed and expected results. The goal of this test is to identify whether a disparity between actual and predicted data is due to chance or to a link between the variables under consideration. As a result, the chi-square test is an ideal choice for aiding in our understanding and interpretation of the connection between our two categorical variables.

The Mann-Whitney U test

The Mann-Whitney U test is the non-parametric counterpart to the t-test for independent samples. It tests if there is a difference between two groups. The Mann-Whitney U test is used when the requirement of normal distribution for the t-test is not met.

Pearson r

The Pearson product-moment correlation coefficient (or Pearson correlation coefficient, for short) is a measure of the strength of a linear association between two variables and is denoted by r. The r indicates how far away all data points are to this line of best fit. (i.e., how well the data points fit this new model/line of best fit).

The Pearson correlation coefficient, r, can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. A value greater than 0 indicates a positive association; that is, as the value of one variable increases, so does the value of the other variable. A value less than 0 indicates a negative association; that is, as the value of one variable increases, the value of the other variable decreases.

Coding and Analyzing Teacher Data

The research team created one coding manual and 0-5 scale rubric to evaluate teachers' culturally responsive instruction, and another manual with the same 0-5 scale rubric to evaluate their facilitation skills. Research Assistants were trained across multiple days to code teacher data using these two manuals. Once a standard of at least 80% inter-rater reliability had been established during these training sessions, the Lead Researcher assigned each member of the coding team a list of teachers to evaluate using the manuals.

Research Assistants independently watched teacher-facing videos one time each for evidence of teachers' facilitation skills, using a note-taking worksheet to help organize their thoughts, and entered scores of 0-5 for fifteen indicators across three categories. First, each Research Assistant watched the teacher's pre video from Fall 2022 and code the teacher's facilitation skills, then they watched and coded the same teacher's post video from Spring 2023. Each Research Assistant had an assigned partner with whom they could check in about any questions they had, and the Lead Researcher followed up with this team of Research Assistants periodically to check for at least 80% inter-rater reliability. While our team did evaluate both Thinking Through Art and control group teachers' facilitation skills, the Thinking Through Art staff primarily used the facilitation scores to give targeted feedback to Thinking Through Art teachers during professional development sessions. We chose not to include our facilitation skills findings in this report because, as expected, the teachers participating in the Thinking Through Art program were adept at facilitating using the Visual Thinking Strategies protocol we taught them, and we did not expect the control teachers (who were not trained by our team in any facilitation techniques) to be able to do the same. The other type of teacher practice analysis (see below), culturally responsive instruction, is a more broadly applicable measure of strong teaching, and more relevant to our research questions, as well as the school district's goals.

To assess the twelve indicators across four categories of teachers' culturally responsive instruction, Research Assistants watched each teacher-facing video twice. The first round of watching gave Research Assistants an overall impression of the lesson, and the second watch was for careful note taking. As with the coding of teachers' facilitation skills, each Research Assistant watched first the pre video and coded it, then watched the post video and coded it. Every two weeks, this group of Research Assistants met with the Lead Researcher to discuss one teacher's set of pre and post videos as an inter-rater reliability check. Research Assistants entered their scores for teachers' culturally responsive instruction and facilitation skills in a spreadsheet, which was then shared with the statistical analyst for the project. Using SPSS version 29.0.1.0 (171), the statistical analyst scaled each type of teacher practice score using the scales in Table 7. Scaled teacher practice data was statistically analyzed using OLAP Cubes and Cross tabulations; statistical significance was determined using Mann-Whitney U and Pearson R evaluations.

Table 7. Teacher Practice Score Scaling Guidelines

	Indicator Score Range / Number of Indicators	Raw Score Range	Score Scale (Raw Score Range = Scaled Score)	Scaled Score Range
Teachers' Facilitation Skills TOTAL SCORE	Indicator Score Range = 0 - 5 Number of Indicators = 15	0 - 75	0 - 15 = 1 16 - 30 = 2 31 - 45 = 3 46 - 60 = 4 61 - 75 = 5	1 - 5
Teachers' Culturally Responsive Instruction TOTAL SCORE	Indicator Score Range = 0 - 5 Number of Indicators = 12	0 - 60	0 - 12 = 1 13 - 24 = 2 25 - 36 = 3 37 - 48 = 4 49 - 60 = 5	1 - 5
Teachers' Culturally Responsive Instruction INDIVIDUAL CATEGORY SCORES	Indicator Score Range = 0 - 5 Number of Indicators = 12 (3 per Category) Number of Categories = 4	0 - 15	0 - 3 = 1 4 - 6 = 2 7 - 9 = 3 10 - 12 = 4 13 - 15 = 5	1 - 5

Table 8. Indicator Score Range and Descriptions for Teacher Practice

Low Scores	Score = 0	Rarely descriptive of teacher practice • Happens almost never (less than 10% of the time) • Includes teacher practice that are counterproductive or foster disruptions
	Score = 1	Occasionally descriptive of teacher practice • Happens about 10-25% of the time • Teacher frequently misses opportunities, or opportunities do not emerge
	Score = 2	Sometimes descriptive of teacher practice • Happens about 30% of the time • Teacher behavior is inconsistent or ineffective
High Scores	Score = 3	Often descriptive of teacher practice • Happens about 50% of the time • Meeting expectations • Teacher consistently attempts to do this, but with moderate effectiveness
	Score = 4	Frequently descriptive of teacher practice • Happens about 75% of the time • Teacher is accomplished at exhibiting these skills • Teacher choices elevate participation in the discussion
	Score = 5	Highly descriptive of teacher practice • Happens nearly 90% or more of the time • Teacher is highly successful • Discussion is characterized by deep, rigorous investigation

FINDINGS: Teachers' Culturally Responsive Instruction Scores

Table 9. FINDINGS: Teachers' Culturally Responsive Instruction Scores

- **Comparisons between Thinking Through Art teachers and control group teachers**
- **Comparisons between Culturally Responsive Instruction category Pre scores and Post scores**
- **Percent growth in scores from beginning to end of year in each Culturally Responsive Instruction category**

Culturally Responsive Instruction Category	Thinking Through Art Teachers (n=38)			Control Group Teachers (n=17)		
	Mean pre score (range 0 - 15)	Mean post score (range 0 - 15)	% Growth from pre to post	Mean pre score (range 0 - 15)	Mean post score (range 0 - 15)	% Growth from pre to post
Student-Centered Teaching (Mann-Whitney U: *P*=003) (x2= 20.856a, df=4, *P*<.001)	7.89	10.26	**30.0%**	4.82	5.12	**6.1%**
Culturally Responsive Discourse (Mann-Whitney U: *P*<.001) (x2= 23.498a, df=4, *P*<.001)]	8.34	10.82	**29.7%**	5.29	5.47	**3.3%**
Classroom Relationships (Mann-Whitney U: *P*=.039) (x2= 13.810a, df=3, *P*<.003)]	11.79	13.03	**10.5%**	10.00	9.94	**- 0.6%**
Instructional Practices (Mann-Whitney U: *P*=.054) (x2= 14.283a, df=4, *P*<.006)]	10.34	12.08	**16.8%**	8.47	9.35	**10.4%**

* Denotes statistically significant findings

The research team also created a survey for both Thinking Through Art (treatment) and control group teachers to complete at the beginning and the end of the school year. Survey questions asked teachers to answer using a seven-point Likert scale, or to provide short answers. Teachers in the Thinking Through Art group completed their pre survey in the spring of 2022 when they signed up to participate in the Thinking Through Art program. Teachers in the control group completed their pre survey in late summer or early fall of 2022 when they signed up to participate in the control group for the Impact Study. Teachers in both the Thinking Through Art and control groups all completed their end-of-year surveys after their class's final data collection session in the late spring of 2023.

Research Assistants compiled teachers' survey responses into a spreadsheet and compared average treatment group and average control group pre and post scores, as well as average score change for each question. Our team also compared individual teachers' pre and post scores, as well as each individual teacher's score change for each question. Some survey questions asked teachers to write an open response to a prompt; we analyzed teachers' reflections of their own learning and educational philosophies and practices at the end of the year for evidence in their own words of the impact they perceived the Thinking Through Art program had on them and their students over the course of the school year.

Researchers used the following documents to organize data collection and coding for teacher data:

- Culturally Responsive Instruction Notes Worksheet
  docs.google.com/document/d/1wqYRqAzNv6nVGqKi04MTAiQG23LiksYfbjIbyMLMl90/copy

- Teacher Practice: Culturally Responsive Instruction and Facilitations Skills Ratings Results
  docs.google.com/spreadsheets/d/19bhrQSauaYnAi3tMgKkJmnU2pB5sr_61-2kpR1b4at8/copy

Coding and Analyzing Student Data

The research team created one coding manual and 0-4 scale rubric to evaluate five categories of students' social-emotional learning as students discussed an image collaboratively with their peers, and another manual to evaluate each individual students' critical thinking in a response to the image.
The team also created a short student survey to evaluate individual students' sense of their own social-emotional learning using a five-point Likert scale.

Evaluating Social-Emotional Learning (Classroom Lesson Video Coding)

Research Assistants were trained across multiple days to code students' social-emotional data using the manual and 0-4 scale rubric. Once a standard of at least 80% inter-rater reliability had been established during these training sessions, the Lead Researcher assigned pairs of coders to watch the student-facing videos of classroom lessons, discuss them together, and come to consensus on how to score students' behaviors for social-emotional learning.

Each member of the coding pairs independently watched student-facing videos twice each for evidence of students' social-emotional learning from October to December 2023. Coders used a note-taking worksheet to help organize their thoughts, and entered their scores of 0-4 for thirteen indicators across five categories. The first round of viewing was designed to give coders an overall impression of the gestalt of the lesson, and the second was for careful notetaking of the conversation amongst the teacher and their students. As with the teacher-facing videos, each coder would watch the class pre video from Fall 2022 and code the students' social-emotional learning, then watch and code their post video from Spring 2023. Every 1-2 weeks, coders would meet with their coding partner to discuss pre and post videos for each of the classes they had been assigned, and to finalize scores for each indicator. Every other round of coding, all coding pairs would watch and coder the same teacher's set of pre and post videos as an inter-rater reliability check, and meet as a whole group to discuss and finalize scores.

Coders entered their scores for students' social-emotional learning in a spreadsheet, which was then shared with the statistical analyst for the project. Using SPSS version 29.0.1.0 (171), the statistical analyst scaled each type of social-emotional learning score using the scales in Table 10. Scaled student social-emotional learning data was statistically analyzed using OLAP Cubes and Cross tabulations; statistical significance was determined using Mann-Whitney U and Pearson R evaluations.

Table 10. Students' Social-Emotional Learning Score Scaling Guidelines

	Indicator Score Range / Number of Indicators	**Raw Score Range**	**Score Scale (Raw Score Range = Scaled Score)**	**Scaled Score Range**
Students' Social-Emotional Learning INDIVIDUAL CATEGORY SCORES (Categories 1 - 4)	Indicator Score Range = 0 - 4 Number of Indicators = 12 (3 Indicators per Category)	0 - 48	0 - 12 = 1 13 - 24 = 2 25 - 36 = 3 37 - 48 = 4	1 - 4
Students' Social-Emotional Learning INDIVIDUAL CATEGORY SCORES (Category 5)	Indicator Score Range = 0 - 4 Number of Indicators = 1	0 - 4	Raw score = scaled score	0 - 4
Students' Social-Emotional Learning TOTAL SCORE	Indicator Score Range = 0 - 4 Number of Indicators = 13	0 - 52	0 - 13 = 1 14 - 26 = 2 27 - 39 = 3 40 - 52 = 4	1 - 4

Table 11. Indicator Score Range and Descriptions for Students' Social-Emotional Learning

Grades Pre-K - 2	Grades 3-12
0 - No Evidence • 1 Seldom descriptive of classroom behavior – happens or isn't seen very often	**0 - No Evidence** • 1 Seldom descriptive of classroom behavior – happens or isn't seen very often
1 - Beginning • 1 Seldom descriptive of classroom behavior – happens or isn't seen very often Students are curious about many different things in the environment - each other, objects on display - they move their bodies to look more closely and to touch, they turn attention and gaze in many directions With consistent reminders, students focus on looking at the image Students share space, seating and objects with each other, with minimal disruptions With support, students begin to balance their own needs with those of others. Students make comments and point out observations that they see with little or no attention to the comments made by other students.	**1 - Beginning** • 1 Seldom descriptive of classroom behavior – happens or isn't seen very often Students are slow to provide what they see and often repeat what another student has already pointed out; very few questions or statements of "I wonder?" Students are easily distracted by something happening in another section of the room; the overall energy is low Students stay within themselves – very little eye contact, gazing or turning to see other students; students much more attentive to the teacher than to other students Only a few students actively participate in the discussion and comments move from one topic or section of the image with no connections There are dead/silent moments during the lesson that are longer than time needed to gather thoughts before speaking; teacher works hard to get student responses with little or no response from most of the students Comments focus on the obvious and provide little or no opportunities for finding connections across the comments
2 - Developing • 2 Sometimes descriptive of classroom behavior – happens or is seen occasionally Students increasingly turn their attention to the comments others are making about the image Students demonstrate readiness to participate by raising hand or approaching the image, and with support, point out their observations mostly to the teacher With support, students listen and demonstrate attention when others talk Students understand and practice taking turns Students repeat what others say or continue a list of observations made by others.	**2 - Developing** • 2 Sometimes descriptive of classroom behavior – happens or is seen occasionally A handful of students generate responses to the question: "What more can we find?"; some students are quiet, their body language isolates them from the group – head down, few facial expressions in response to comments from others Some students have attentive posture, eye contact and a few ways of showing connection with others Individual students offer observations and other comments, there is little extension of ideas from one student to another; communication is more frequently directed to the teacher than to other students A handful of students offer their ideas and observations without prompting from the teacher; others are called on by the teacher to participate; some students wait for the prompt from the teacher to provide evidence for their interpretations Comments remain related to the elements of the image with very little connection to big ideas or their lives outside of school

3 - Accomplished

- 3 Often descriptive of classroom behavior - happens or is seen more often than not

Students are increasingly able to maintain focus and attention on the image and make comments that relate to the image

Students are increasingly aware of the presence of others and respect their personal space

Students mostly listen when others are talking, without prompting from teacher or other students

Students are comfortable with observations and ideas that are different from their own

Students build on the comments of others, but not in explicit ways

3 - Accomplished

- 3 Often descriptive of classroom behavior - happens or is seen more often than not

Students generate a large number of responses to the question: "What more can you find?" Most are straightforward literal descriptions of an aspect of the image.

Most students exhibit behaviors of respect and empathy; they mostly listen attentively; they signal connection to each other with smiles and gestures

Conversation about a particular aspect of the image or idea are more frequent and include more than one student, many, but not all students exhibit interest in the conversation

Most students are comfortable expressing their ideas and feelings without the prompting of the teacher and follow the Thinking Through Art "rules of engagement" without prompting by the teacher – many provide evidence for their inferences and interpretation without prompting from the teacher

Occasional instances of one or more students making a connection between observations and ideas and social issues, big concepts and/or their lives outside of school

4 - Exemplary

- 4 Highly descriptive of classroom behavior – happens frequently

Students focus on image, with little or no prompting by teacher and are eager to contribute comments related to the image

Students understand and follow rules for participation in the discussion and sometimes remind other students to follow the rules

Students begin to notice and respond to facial and body cues that represent the feelings of others. Without prompting, students treat others with respect and kindness

Students recognize differences, similarities and positive qualities among their classmates

Students acknowledge other students' comments and build on ideas of others by linking their comments to them

4 - Exemplary

- 4 Highly descriptive of classroom behavior – happens frequently

Students engage in identifying patterns and underlying ideas or principles when describing what they see

Group has established a culture of respect and empathy; students listen attentively to a variety of perspectives; they pick up on verbal and nonverbal cues to understand what others are thinking and feeling

There are extended conversational exchanges involving several students while others pay attention and express interest nonverbally

Students are comfortable expressing their thoughts, don't hesitate to ask a question and are not reliant on teacher to remind them of appropriate behavior or the Thinking Through Art structure for group conversation

The conversation expands beyond literal descriptions of image and includes making connections between observations and concepts and issues they know about and/or their own ideas

FINDINGS: Students' Social-Emotional Learning Scores

Table 12. FINDINGS: Students' Social-Emotional Learning Scores

- **Comparisons between all Thinking Through Art students and all control group students**
- **Comparisons between Social-Emotional Learning category Pre scores and Post scores**
- **Percent Growth in scores from beginning to end of year in each Social-Emotional Learning category**

Social-Emotional Learning Category	Thinking Through Art (Treatment) Classes (n=38)			Control Group Classes (n=17)		
	Mean pre score	Mean post score	% Growth from pre to post	Mean pre score	Mean post score	% Growth from pre to post
	(score range 0 - 4)			(score range 0 - 4)		
Creating a Community of Learners (Mann-Whitney U: *P*=.002)	1.87	2.76	**47.9%***	1.94	2.12	**9.1%**
	(scaled score range 0 - 12)			(scaled score range 0 - 12)		
Respects Multiple Perspectives (Mann-Whitney U: *P*=.044)	4.37	6.11	**39.8%***	5.00	5.29	**5.5%**
Demonstrates Relationship Skills (Mann-Whitney U: *P*=.011)	5.47	7.45	**36.1%***	5.29	5.94	**12.2%**
Possesses Self Confidence and Belief in Capacity to Learn (Mann-Whitney U: *P*=.010)	6.71	9.00	**34.1%***	6.88	7.65	**11.1%**
Expresses Empathy (Mann-Whitney U: *P*=.300)	3.16	3.87	**22.5%**	2.53	3.06	**20.9%**

* Denotes statistically significant findings

Researchers used the following documents to organize data collection and coding for student social-emotional learning data:

- Social-Emotional Learning Notes Worksheet
  docs.google.com/document/d/1sPNN-vVD8nYe4EXiSz17e2D7WGu7o7lKQJ7ly-3jcrE/copy

- Social-Emotional Learning Coding Results Spreadsheet
  docs.google.com/spreadsheets/d/1qyPW6GrexwTI5_OBASYrBcXWYHBa46f85Zv2soZrwtc/copy

Evaluating Social-Emotional Learning (Student Surveys)

The research team created a nine-question survey for students that asked them to state how much they agreed or disagreed with statements about their confidence, the way they thought about their own learning, and what they thought about their classroom environments, on a five-point Likert scale. We expected students to be able to answer these questions with support from their teachers and the Research Assistants during the data collection sessions, and that we would see some clear patterns from students' pre responses at the beginning of the school year to their post responses at the end of the school year. What we found was that there was a large amount of survey data that was unusable because students had chosen more than one response per question, not to respond at all, or had written in an answer not on the Likert scale. We found that individual students' answers were so idiosyncratic that no clear patterns emerged across the treatment and control groups, or across pre to post data. We ultimately decided that surveys were an inappropriate tool to evaluate social-emotional learning in a classroom setting, because students' self-reported answers were not reliable and did not get at the complexity of what we saw going on in the dynamic social space of a classroom discussion. We also felt that our coding manual and procedures for evaluating students' social-emotional learning at the classroom level were robust enough to give us a clear picture of students' abilities and progress in these areas. Therefore, we opted not to factor students' responses to the survey into our overall social-emotional learning analysis.

Evaluating Critical Thinking (Student Response Coding)

Student individual responses to the study images were captured in a variety of ways in accordance with this study's Universal Design principles and teachers' understandings of the least restrictive and most robust ways for each of their students to respond: handwritten, typed into a word processor, audio recorded by a member of the research team, scribed in real time by a member of the research team, or video recorded by the student using Microsoft Flipgrid. Other than five deaf or hard of hearing students using Flipgrid to record their responses in American Sign Language, all other students in the study chose whether to write or speak, or have their teacher choose the most appropriate mode of response for them.

All student responses written or recorded in English were transcribed exactly by the research team for coding. Research team members transcribed students' misspellings and only added clarifications in brackets if necessary. All student responses written or recorded in languages other than English were

transcribed and translated into English by multilingual members of the research team and/or by a professional translation service. While the majority of student responses were written and/or recorded in English, several dozen were created in a language other than English (primarily Spanish, Vietnamese, and Portuguese). Any drawings or other non-text elements students wrote on the response forms were copied into the transcription, and were evaluated by coders according to guidelines laid out in the Critical Thinking Manual.

To evaluate critical thinking in a holistic and inclusive way, the research team created a coding manual which includes guidelines for assessment of five distinct categories of critical thinking. Four of the categories of critical thinking could be evaluated by counting instances of specific indicators in students' responses to an image. Each of these four categories contains three distinct indicators, for a total of twelve critical thinking indicators. The fifth critical thinking category, "Exhibits Reasoning", was designed as a gestalt measure to evaluate each student's ability to synthesize the other four types of critical thinking into a cogently and critically reasoned response. Rather than looking for indicators, as in the first four critical thinking categories, coders scored students' reasoning using grade-band-specific 0-4 scale rubrics.

As with the coding manual training for students' social-emotional learning and teachers' culturally responsive instruction and facilitation skills, Research Assistants worked together collaboratively across a two-day "coding training marathon" to learn the critical thinking manual and to begin to establish at least 80% inter-rater reliability. Each Research Assistant was assigned several classes-worth of sampled student responses for each round of coding from June 2023 to October 2023. From the 714 students with consent to participate and share their responses across the treatment and control groups, we took a demographically representative sample of 388 students (237 from the treatment group and 151 from the control group) and coded their responses to the art or narrative images from the beginning and the end of the year. Coders worked independently, but met as a whole group with the Lead Researcher weekly to discuss any questions or unclear elements in student responses. These discussions served as regular inter-rater reliability checks; every few coding rounds the Lead Researcher also assigned a class set of student responses to all coders as an extra inter-rater reliability check. In the first rounds of critical thinking coding, coders scored students' pre responses from Fall 2022. In the last rounds of coding, they scored students' post responses from Spring 2023.

Coders entered their scores for students' critical thinking in a spreadsheet, which was then shared with the statistical analyst for the project. Using SPSS version 29.0.1.0 (171), the statistical analyst scaled each type of social-emotional learning score using the scales in Table 13. Scaled student social-emotional learning data was statistically analyzed using OLAP Cubes and Cross tabulations; statistical significance was determined using Chi Squares, and Pearson R evaluations.

Table 13. Students' Critical Thinking Score Scaling Guidelines

	Score Scale (Raw Score Range = Scaled Score)	Score Scale (Raw Score Range = Scaled Score)	Score Scale (Raw Score Range = Scaled Score)
Category 1 - OBSERVATIONS Total Category Score Range = 0 - 12 Total Category Score Scaling: 0 (No Evidence) = 0 1 - 3 (Rarely descriptive) = 1 4 - 6 (Sometimes descriptive) = 2 7 - 9 (Often descriptive) = 3 10 - 12 (Highly descriptive) = 4	**1.a IDENTIFICATION** 0 (No Evidence) = 0 1 - 4 (Rarely descriptive) = 1 5 - 9 (Sometimes descriptive) = 2 10 - 14 (Often descriptive) = 3 15+ (Highly descriptive) = 4	**1.b MOVEMENT** 0 (No Evidence) = 0 1 - 3 (Rarely descriptive) = 1 4 - 7 (Sometimes descriptive) = 2 8 - 11 (Often descriptive) = 3 12+ (Highly descriptive) = 4	**1.c DESCRIPTION** 0 (No Evidence) = 0 1 - 4 (Rarely descriptive) = 1 5 - 9 (Sometimes descriptive) = 2 10 - 14 (Often descriptive) = 3 15+ (Highly descriptive) = 4
Category 2 - ASSOCIATIONS Total Category Score Range = 0 - 12 Total Category Score Scaling: 0 (No Evidence) = 0 1 - 3 (Rarely descriptive) = 1 4 - 6 (Sometimes descriptive) = 2 7 - 9 (Often descriptive) = 3 10 - 12 (Highly descriptive) = 4	**2.a PRIOR KNOWLEDGE** 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4+ (Highly descriptive) = 4	**2.b COMPARISON** 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4+ (Highly descriptive) = 4	**2.c SIMILE/METAPHOR** 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4+ (Highly descriptive) = 4
Category 3 - INTERPRETATION Total Category Score Range = 0 - 12 Total Category Score Scaling: 0 (No Evidence) = 0 1 - 3 (Rarely descriptive) = 1 4 - 6 (Sometimes descriptive) = 2 7 - 9 (Often descriptive) = 3 10 - 12 (Highly descriptive) = 4	**3.a CLAIM, NO EVIDENCE** 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 - 4 (Sometimes descriptive) = 2 5 - 7 (Often descriptive) = 3 8+ (Highly descriptive) = 4	**3.b CLAIM WITH EVIDENCE** 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4+ (Highly descriptive) = 4	**3.c POSSIBLE CONDITIONS** 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4+ (Highly descriptive) = 4
Category 4 - MULTIPLE PERSPECTIVES Total Category Score Range = 0 - 12 Total Category Score Scaling: 0 (No Evidence) = 0 1 - 3 (Rarely descriptive) = 1 4 - 6 (Sometimes descriptive) = 2 7 - 9 (Often descriptive) = 3 10 - 12 (Highly descriptive) = 4	**4.a UNCERTAINTY** 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4+ (Highly descriptive) = 4	**4.b MULTIPLE POSSIBILITIES** 0 (No Evidence) = 0 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4+ (Highly descriptive) = 4	**4.c REVISION** 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4+ (Highly descriptive) = 4
Category 5 - CRITICAL REASONING Total Category Score Range = 0 - 4 RAW SCORE = SCALED SCORE	RAW SCORE = SCALED SCORE: 0 (No Evidence) = 0 1 (Rarely descriptive) = 1 2 (Sometimes descriptive) = 2 3 (Often descriptive) = 3 4 (Highly descriptive) = 4		
TOTAL (ALL CATEGORIES) Score Range = 0 - 52 TOTAL (ALL CATEGORIES) Score Scaling: 0 (No Evidence) = 0 1 - 13 (Rarely descriptive) = 1 14 - 26 (Sometimes descriptive) = 2 27 - 39 (Often descriptive) = 3 40 - 52 (Highly descriptive) = 4			

FINDINGS: Students' Critical Thinking Scores

Table 14. FINDINGS: Students' Critical Thinking Scores

- **Comparisons between Thinking Through Art students and control group students**
- **Comparisons between Critical Thinking category Pre scores and Post scores**
- **Percent growth in scores from beginning to end of year in each Critical Thinking category**

	Thinking Through Art (Treatment) Classes (n=38)			Control Group Classes (n=17)		
Critical Thinking Category	**Mean pre score**	**Mean post score**	**% Growth from pre to post**	**Mean pre score**	**Mean post score**	**% Growth from pre to post**
	(scaled score range 0 - 4)			**(scaled score range 0 - 4)**		
Exhibits Reasoning (Mann-Whitney U: *P*<.001)	1.61	2.23	**38.6%***	1.85	1.84	**- 0.4%**
	(raw score range 0 - 15+)			**(raw score range 0 - 15+)**		
Identifies Information (Mann-Whitney U: *P*=.050)	17.64	19.19	**8.8%**	17.66	18.84	**6.7%**
Constructs Meaning (Mann-Whitney U: *P*=.050)	4.70	4.85	**3.2%**	4.65	4.89	**5.1%**
Makes Associations (Mann-Whitney U: *P*=.050)	2.60	2.47	**- 4.7%**	2.79	2.49	**-10.9%**
Considers Multiple Perspectives (Mann-Whitney U: *P*=.050)	0.576	0.546	**- 5.1%**	0.695	0.808	**16.2%**

* Denotes statistically significant findings

Researchers used the following documents to organize data collection and coding for student critical thinking data:

- Student Response: Critical Thinking Coding Work
  docs.google.com/document/d/12l5cQ2DGFjMRCZPQigyY2mvomZgBtnbQ7T38lvSUwJQ/copy

- Critical Thinking Coding Results
  docs.google.com/spreadsheets/d/1sUTaflKpxkSmquTiFsoNLc0NGIYmDf2FlYdGOhLn90k/copy

8. STUDY LIMITATIONS AND RECOMMENDATIONS

As with any study involving work with teachers and students within a large school district, this one featured many unexpected complexities that required flexibility, creative thinking, and collaboration. We hope that the following examples can inform future research.

- **Sample limitations.** Based on previous cohorts of Thinking Through Art teachers and students, this study was designed with the expectation that the 2022-2023 participants would have a roughly even distribution across all grade levels. The study sample ended up having 80% of teachers and 82% of students in grades PreK-5. Without a larger sample size of middle and high school students, our findings across the secondary grade band (grades 6-12) are less robust than those in the early childhood and upper elementary grade bands. We recommend replicating this study with a cohort of teachers and students from grades 6-12.

- **Collecting data from students with disabilities.** This study's student sample under-represented students with disabilities, as they were more likely than students without disabilities to deny consent to participate in this research. Students with disabilities make up 22% of Boston Public Schools students, but 17% of this study's student 'universe', and only 16% of the students who provided consent to participate. We strongly recommend future researchers work to build trust with school communities, particularly the families of students with disabilities, before asking to conduct research with these students.

- **Avoid mixed modalities.** A limitation of our data collection methods was that we allowed teachers and students to decide what modality of response to use during both the beginning and end of year data collection, which resulted in inconsistencies in our data. Across many classes, including most classes in grades 1 and 2, teachers opted to have students respond orally in the fall and then practice their emerging writing skills in the spring. At the time, this seemed like an appropriate way to honor students' learning and the pride they take in their work. When we began to analyze the data, however, we found that many students' critical thinking scores decreased from the beginning of the school year, simply because they were able to say more into an audio recorded in the fall than they were able to write in the spring. In order to remove that variable, future researchers should consider allowing students and teachers to make the choice of mode of response for the pre assignment, but should keep modalities consistent for the post assignment.

- **Survey limitations:** The use of surveys as key methodological tools is a given in many social science and education studies, and initially our team sought to learn a lot from teachers and students through surveys. While the data from the teacher surveys revealed some trends and patterns regarding teachers' comfort with discussing art with their students, the student survey data didn't add anything to our understanding of students' social-emotional learning in a group setting. Many students in grades PreK-5 had trouble using the five-point Likert scale to answer questions about their learning style and

socialization within their classroom, even with assistance. Ultimately, we determined that student self-report on a survey is an insufficient tool to capture social-emotional learning, especially when compared with the richness of our methods of analysis of observed student behaviors during classroom discussion.

- **Additional research on empathy.** The sole category of students' social-emotional learning where we did not have statistically significant findings was *Expresses Empathy*. Visual Thinking Strategies, our program's core pedagogical framework, was not explicitly designed for conversations about emotions – neither those felt by participating students, nor those that might be expressed by figures depicted in art images. Furthermore, we found through watching 110 classroom lesson videos that a whole-class discussion lasting just 12-15 minutes was unlikely to include opportunities for students to demonstrate empathy. One reason for this might have been our choice of images to use in the study (see study images' in Figures 4 and 5, above), neither of which depicts strongly emotionally expressive content. Future research to determine how programs like Thinking Through Art could better support building students' empathic understandings and behaviors would likely be useful for both museums and schools.

- **Refine Critical Thinking Manual.** Due to the prevalence of research on critical thinking in programs using Visual Thinking Strategies, the Critical Thinking Manual was designed to look for types of critical thinking that would be rare but exciting, such as students' use of simile and formulating hypotheses. Unsurprisingly, these forms of critical thinking did not appear frequently enough in either Thinking Through Art or control student responses to generate a statistically significant finding. Furthermore, we underestimated the complexity of evaluating twelve different indicators of critical thinking – including those of which may have inverse relationships – on the same scale. For future research, we would strongly recommend making significant revisions to the Critical Thinking Manual to better measure and categorize the types of critical thinking present in Thinking Through Art or other school partnership programs.

9. STUDY STRENGTHS AND LESSONS

Our team learned many lessons in the course of completing this research that will strengthen the Thinking Through Art program, guide future education research at the Gardner Museum, and strengthen the museum's relationship with Boston Public Schools, as well as the teachers who participate in Thinking Through Art.

- **The value of researcher-practitioner expertise.** Early on, we found that our model of researcher-practitioners would help us leverage the expertise of Thinking Through Art staff to collect and code data. We did not need to train additional research assistants in using and evaluating Visual Thinking Strategies, for example, because the Thinking Through Art staff members were experienced in evaluating and coaching teachers in these practices. The existing relationship between the teachers in the Thinking Through Art group and the Gardner Museum staff members turned data collection sessions into an opportunity for all parties to build stronger connections. This was particularly important for the museum staff members, who gained considerable concrete knowledge about how the teachers worked with their students in their classrooms This unexpected feature of the study led to the first programming change based on this research for Thinking Through Art: staff coaching visits to the classrooms are available to all participating teachers in the 2023-2024 school year.

- **Inclusive design is vital for education research.** For any researchers who are interested in research in schools, our team strongly recommends designing study instruments that adhere to the principles of Universal Design, particularly in that they offer myriad access points and support student engagement using flexible modalities.

- **Lower the barriers for participation for teachers.** At every choice point, we minimized burdens on teachers and simplified their tasks as much as possible. This approach covered every decision from instrument design to data collection, and we are grateful to the Research Advisory Group for being our thought partners. The result of taking on all possible tasks as a research team instead of pushing them onto the teachers is that teachers found participation in the study to be not only a light lift but a benefit to their classrooms. Both treatment and control teachers returned to the Gardner Museum's programs in subsequent years, eager to maintain the relationships developed during this study.

10. REQUEST FOR DATA

Please email **education@isgm.org** to request anonymized study data (available prior to June 2027).

ADDITIONAL STUDY MATERIALS

AVAILABLE AT GARDNERMUSEUM.ORG/ORGANIZATION/EDUCATION/RESEARCH

THINKING THROUGH ART: A TRANSFORMATIVE MUSEUM-SCHOOL PARTNERSHIP (REPORT)

EXECUTIVE SUMMARY

TECHNICAL SUPPLEMENT

DATA COLLECTION & MANAGEMENT PROTOCOLS

RESEARCH INSTRUMENTS

DATA CODING & RATING MANUALS

INFORMATIONAL VIDEOS

11. SOURCES CONSULTED

Museum Education & Visual Thinking Strategies

Adams, M., Foutz, S., Luke, J. & Stein, J. (2006) Thinking Through Art: School partnership program, year 3 report. Annapolis, MD: Institute for Learning Innovation (ili). https://www.gardnermuseum.org/sites/default/files/uploads/files/Year_3_Report.pdf

Burchenal, M. and Grohe, M. (2007). Thinking Through Art: Transforming museum curriculum. *Journal of Museum Education.* 32(2), 111-122.

DeSantis, K. & Housen, A. (2007). Highlights of Findings – San Antonio: Aesthetic Development and Creative and Critical Thinking Skills Study. Visual Understanding in Education. http://www.issuelab.org/permalink/resource/15725

DeSantis, K. (2009). Report to the education department of the Isabella Stewart Gardner Museum on the 8th grade school partnership program Visual Thinking Strategies adaptation 2008-2009. New York: VUE.

Egan, S. & Grohe, M. (2013). Follow-Up to 8th Grade School Partnership Study. Boston, MA: Isabella Stewart Gardner Museum. https://www.gardnermuseum.org/sites/default/files/uploads/files/IMLS%20Follow%20Up%20Report.pdf

Egan, S. & Grohe, M. (2014). Edward M. Kennedy Academy for Health Careers grade 9 (class of 2016) year 1 program critical thinking skills trends. Boston, MA: Isabella Stewart Gardner Museum.

Egan, S. (2020). Isabella Stewart Gardner Museum critical thinking skills coding manual. Unpublished.

Isabella Stewart Gardner Museum. (2019). Thinking Through Art writing sample findings. Unpublished.

Isabella Stewart Gardner Museum. (2016). Isabella Stewart Gardner Museum partnership with Edward M. Kennedy Academy for Health Careers: critical thinking skill growth." Unpublished.

Housen, A.C. (2001). Aesthetic thought, critical thinking and transfer. *Arts and Learning Research Journal, 18(1)*, 99-131.

Terrassa, J., Hubard, O., Holtrop, E. & Higgins-Linder, M. (2016). Impact of art museum programs on students: Literature Review. Alexandria, VA: National Art Education Association, p. 6-9.

Vygotsky, L. S. (1978). In M. Cole, V. John-Steiner, s. Scribner & E. Souberman (Eds.), *Mind in Society: The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press.

Yenawine, P. (1999). Theory into practice: The visual thinking strategies. Visual Thinking Strategies.

Yenawine, P. (2013). *Visual Thinking Strategies: Using art to deepen learning across school disciplines.* Cambridge, MA: Harvard Education Press.

Boston Public Schools

Boston Public Schools. (2020, April 29). *Imagine Boston Public Schools: Boston Public Schools 20/25 Strategic Plan. Boston Public Schools Strategic Plan*. Retrieved from https://www.bostonpublicschools.org/strategicplan

Boston Public Schools. (2021). *Equity & CLSP Toolkit - Equitable Literacy instruction.* Equity & CLSP Toolkit - Equitable Literacy Instruction. Retrieved from https://sites.google.com/bostonpublicschools.org/ogequityandclsptoolkit/equitable-literacy-instruction.

Boston Public Schools. (n.d.). Opportunity index. Retrieved from https://www.bostonpublicschools.org/domain/2301

Boston Public Schools. (n.d.). Social Emotional Learning Competencies & Skills. Retrieved from https://docs.google.com/document/d/1E7qJd3zchnHxx3O-TSXkKqJ6Vb8eh-YYmIF3lZhIwco/edit.

Massachusetts Department of Elementary and Secondary Education. (2022). 2022 district report card: Boston. Retrieved from https://reportcards.doe.mass.edu/2022/districtreportcard/00350000.

Critical Thinking

Costa, A. L., & Kallick, B. (Eds.). (2009). Habits of mind across the curriculum: Practical and creative strategies for teachers. ASCD.

Facione, P. A., Facione, N. C., & Giancarlo, C. A. F. (2001) California critical thinking disposition inventory: CCTDI, California Academic Press.

Hitchcock, David (Fall 2020) "Critical Thinking", The Stanford Encyclopedia of Philosophy Edward N. Zalta (ed.) https://plato.stanford.edu/archives/fall2020/entries/critical-thinking/>.

Insight Center. (2016). Social-Emotional Skills: Critical Thinking. Boston After School & Beyond. Retrieved from https://insight.bostonbeyond.org/social-emotional-skills/critical-thinking/.

National Council of Excellence in Critical Thinking. (1987) The foundation for critical thinking. https://www.criticalthinking.org/pages/defining-critical-thinking/766.

Tatum, Alfred. (2009). *Reading for their Life: (Re)Building the textual lineages of African American adolescent males.* Portsmouth, NH: Heineman. Enabling Text Rubric. (n.d.) Retrieved from https://drive.google.com/file/d/13zGWSsbzerzILbSDewr9AdBfNU2HZgux/view via https://wakelet.com/wake/w2HN6LRv5X-TNrz7Wd-lo

Think CERCA. (2022). CERCA Framework. Retrieved from https://www.thinkcerca.com/cerca

Webb, N. L. et. al. (2006). Web Alignment Tool. Wisconsin Center of Educational Research. University of Wisconsin-Madison.

Wolpert-Gawron, H. (2017) *Getting critical about critical thinking.* Los Angeles, CA: Lucas Foundation, Edutopia. Retrieved from https://www.edutopia.org/blog/getting-critical-about-critical-thinking-heather-wolpErt-gawron

Social Emotional Learning

Collaborative for Academic, Social Emotional Learning (CASEL). https://casel.org.

Massachusetts Department of Education. (2020). Multi-tiered system of support. Retrieved from https://www.doe.mass.edu/sfss/mtss/

Massachusetts Standards for Preschool and Kindergarten. (2015). Social emotional learning and approaches to play and learning. June 2015.

Pianta, R. C. (1999). *Enhancing relationships between children and teachers.* Washington, DC: American Psychological Association.

Pianta, R. C. (2006). Classroom management and relationships between children and teachers: Implications for research and practice. In C. M. Evertson, & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues.* Mahwah, NJ: Lawrence Erlbaum Associates, 685-710.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Education Research Journal,* 45(2), 365-397.

Weinstein, C., Curran, M., & Tomlinson-Clarke, S. (2003). Culturally Responsive Classroom Management: Awareness into Action. *Theory Into Practice*, 42(4), 269–276.

Culturally Responsive Instruction

Cantrell, S.C., et. al., (2014) Culturally responsive instruction observation protocol (CRIOP) professional development: Year 2. Lexington, KY: Collaborative Center for Literacy Development. Retrieved from https://nisnresourcehub.org/wp-content/uploads/2021/01/Culturally-Responsive-Instruction-Observation-Protocol.pdf

Geneva, G. (2010) *Culturally responsive teaching: Theory, research and practice.* New York: Teachers College Press.

Hammond, Z. (2015) *Culturally responsive teaching and the brain: Promoting authentic engagement and rigor among culturally and linguistically diverse students.* Thousand Oaks, CA: Corwin, a Sage Co.

Ladson-Billings, G. (2021) *Culturally responsive pedagogy: Asking a different question.* New York: Teachers College Press.

Muniz, J. (2019) *Culturally responsive teaching: A 50-state survey of teaching standards*. Washington, DC: New America

Muniz, J. (2020) *Culturally responsive teaching: A reflection guide.* Washington, DC: New America

National Academies of Science, Engineering and Medicine (2018) *How people learn II: Learners, contexts and cultures.* Washington, DC: National Academies Press.

Powell, R. et.al., (2017) Operationalizing culturally responsive instruction: Preliminary findings of CRIOP research. *Teachers College Record,* 118(1), 1-46.

Powell, R., Cantrell, S. C., Correll, P. K., & Malo-Juvera, V. (2017). Culturally Responsive Instruction Observation Protocol (4th ed.). Lexington, KY: University of Kentucky College of Education.

Ronfeldt, M. and Truwit, M. (2023). Considerations for use of teaching quality measures. EdInstruments Brief, Annenberg Institute for School Reform, Brown University.

Methodology

Baldwin, S. & Berkeljon, A. (2010). Quasiexperimental design. In *Encyclopedia of Research Design*. Neil J. Salkind (ed.). Newbury Park, CA: Sage Publishing.
http://sk.sagepub.com/reference/researchdesign

Campbell, D.T. and Stanley, J.C. (1963) *Experimental and quasi-experimental designs for research. Boston:Houghton Mifflin.*

CAST (2018). Universal Design for Learning Guidelines version 2.2. Retrieved from http://udlguidelines.cast.org

Chun, M. (2010, March). Taking teaching to performance task: Linking pedagogical and assessment practices." *Change: The Magazine of Higher Education.*

McTighe, J. (2015), April). What is a performance task?. Mc Tighe Consulting. https://jaymctighe.com/what-is-a-performance-task/

Massachusetts Department of Elementary and Secondary Education School Profiles: Definitions of disabilities. https://www.doe.mass.edu/sped/definitions.html

Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment Research and Evaluation,* 13(4).

Photo by Bearwalk Cinemas