Observing Classrooms Through a Digital Lens:

Examining the Reliability and Feasibility of Video Observations in Pre-kindergarten Classrooms

Todd Grindal, Sarah Nixon Gerard,
Anne Partika, Nancy Perez, Gullnar Syed,
Morgan Solender, Anna Mark
SRI Education

February 2025

Authors

Todd Grindal
Sarah Nixon Gerard
Anne Partika
Nancy Perez
Gullnar Syed
Morgan Solender
Anna Mark

SRI Education

Suggested Citation

Grindal, T., Gerard, S. N., Partika, A., Perez, N., Syed, G., Solender, M., & Mark, A. (2025). *Observing classrooms through a digital lens: Examining the reliability and feasibility of video observations in pre-kindergarten classrooms*. SRI International.

Acknowledgments

The Gates Foundation supported this study (INV-034715) to advance the development of accurate, equitable, useful, and innovative tools to support coaching focused on the quality of center-based pre-kindergarten classrooms that serve children ages 3 to 5 years in the United States. Thank you to our program officers Snow Li, Jenny Mosley, Elizabeth Mokyr Horner, and Jacklyn Altuna Willard for your thoughtful guidance.

We would like to thank MC2 Education team members Jackie Mendez and Liah Watt, as well as observers Elizabeth Baldassarre, Tracee Billingsley, Ellen Frechette, Syreen Goulmamine, Saulondra (Santee) Jauhari, Cris Jimenez, Adriana Morell, Jenna Nguyen, Tasbita Reham, Israel Rodriguez, Jordan Smith, Katie Tripp, and Ho Yan (Bomly) Tsuen, for their meticulous data collection efforts. Additional thanks to Kate Borelli, Jeremy Fritts, Elisa Garcia, Erika Gaylor, Molly Goldberg, Ryan Grimm, Cat Hall, Charles Harding, Santiago Navia Jaramillo, Sharon Mattke, Jennifer Nakamura, Daisy Rutstein, and Krystal Thomas.

We would also like to thank Dr. Bridget Hamre and Dr. Sarah Caverly of Teachstone, Tracy Link of ERSI, and Dr. Stephanie Curenton of Boston University Wheelock College of Education & Human Development, as well as our advisory board members Maria-Isabel Ballivian, Dr. Courtney Bell, Dr. Nicole Gardner-Neblett, and Gonzalo Palza for their guidance throughout the study.

Study preregistration: Open Science Framework – osf.io/c2z38

Contents

Executive Summary

Accurate, reliable, and scalable measurement of classroom quality represent a critical tool for ensuring that young children benefit from early learning programs. Video-recorded classroom observations have the potential to increase the usefulness of classroom quality assessments and improve the capacity of those assessments to strengthen teaching practice (Curby et al., 2016; Kane et al., 2020). For example, video recordings may provide teachers with the opportunity to review video recordings with a coach or mentor (Pianta et al., 2014). Video recordings may also allow observers to see aspects of classroom practice that are difficult to capture during a live observation, such as how early educational experiences differ for racially, culturally, and linguistically minoritized learners (Meek et al., 2021). Finally, videos may reduce the costs of collecting classroom observation data, allowing more teachers to be observed more often (Clark et al., 2022; Kane et al., 2020).

The Early Childhood Classroom Observation (ECCO) study was designed to better understand how video recordings can support high-quality measurement of pre-kindergarten (pre-K) classrooms, using two common measures of early learning programs: Classroom Assessment Scoring System (CLASS) and Early Childhood Environmental Rating Scale, 3rd Edition (ECERS-3). For each measure, the study team compared the reliability of observations gathered through video and live (in-person) classroom observations. The team also interviewed teachers and program leaders to understand their perceptions of the challenges and benefits of video observations.

The study team conducted live and video scoring of 160 observations of pre-K classrooms, conducted surveys of nearly 60 teachers, and interviews with 17 teachers and program leaders. In general, **we find that pre-K classroom observations can reliably be conducted over video and that video scores are generally comparable to live scores.** There are some differences between live and video scores that may limit comparability for certain classrooms, different aspects of practice observed, and the purpose of the observation.

Finding 1: Video observations demonstrate reliability comparable to live observations.

- CLASS and ECERS-3 tools show good reliability when used with video observations.
- Most scores fall within acceptable ranges for inter-rater reliability and consistency.

Finding 2: Live and video observations produce mostly equivalent ratings, with some variations.

- Live and video score differences are more pronounced for dimensions requiring close observation of child behavior.
- The classroom quality level and the languages used in the classroom may influence the comparability of live and video observations.
- The quality of the video recordings varies across observations and may limit the accuracy of video scoring in some cases.
- Live and video scores do not appear to vary within an observation in ways that are meaningful for research use cases. However, CLASS scores vary in meaningful ways that impact their use for accountability purposes, and video ECERS-3 scores may not produce the same guidance for coaching purposes as would scores produced by live ECERS-3 observations.

Finding 3: Teachers have mixed experiences and perceptions about the use of video observations.

- Many teachers see potential benefits of video observations for their professional development.
- Teachers believe video recordings can provide a more comprehensive view of classroom dynamics, capturing subtle student behaviors and interactions that may be missed during live observations.
- Although most teachers do not express strong concerns about video observations, many noted potential challenges including accuracy, privacy, and potential classroom disruptions.

Finding 4: Video observations produce a substantial costs savings when conducting observations in support of coaching, but not when conducting observations for accountability purposes.

- When conducting observations for the purposes of coaching, video observations are less expensive than live observations.
- When conducting observations for the purposes of accountability, video observations are more expensive than live observations.

Introduction

The evidence is clear that high-quality early care and education (ECE) supports children's early learning and may have long-term impacts on their academic and life success (Phillips et al., 2016). In the ECE field, "quality" education is typically conceptualized as a combination of *process* and *structural* quality.

Process quality refers to the interactions between young children and their caregivers that are considered a cornerstone of children's development (Bronfenbrenner & Morris, 2007). This includes teacher practices such as providing an emotionally supportive and responsive classroom environment and individualized instruction that scaffolds children's development (Burchinal, 2018).

Structural quality refers to the features of a classroom or program that are expected to support high-quality interactions, such as class sizes and child–teacher ratios, classroom materials, or teacher training and experience (Burchinal, 2018).

Given the evidence that high-quality ECE is what drives positive impacts of ECE participation on children's development, much of ECE policy is focused on improving the quality of ECE programs (Office of Child Care, 2018; Office of Head Start, 2016). Accurate, reliable, and scalable measurement of ECE quality is therefore a foundational building block of efforts to improve ECE.

Although indicators of structural quality can be gathered through administrative data or surveys of teachers and program leaders, overall pre-kindergarten (pre-K) classroom quality is often measured through observations of children's classroom experiences and teachers' practice. The two most common measures across the United States are Classroom Assessment Scoring System (CLASS; Pianta & Hamre, 2022) and Early Childhood Environmental Rating Scale, 3rd Edition (ECERS-3; Harms et al., 2014). Both are intended as global measures of classroom quality; in other words, they capture quality for all children in the classroom across multiple areas of early learning. CLASS and ECERS-3 each have evidence of modest associations with children's learning outcomes (Burchinal, 2018). CLASS measures three domains of process quality, while ECERS measures both process and structural quality across six subscales.

In the past decade, other researchers have developed new observational tools that build on frameworks established by CLASS or ECERS-3. However, these newer instruments focus on specific aspects of classroom quality that global measures of quality may underrepresent, or on the experiences of individual children in the classroom. Example tools include:

- Assessing Classroom Sociocultural Equity Scale (ACSES), which captures equitable interactions between teachers and racially minoritized learners (Curenton et al., 2020). Higher scores on this measure have been associated with improvements in children's math, executive functioning, and social skills scores (Curenton et al., 2022).
- Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA), which captures language and literacy supports for bilingual students (Freedson et al., 2011). Higher scores on this

measure have been associated with English, Spanish, math, and science skills for dual language learners (Figueras-Daniel & Li, 2021; Partika et al., 2021; White et al., 2020).

- Child Observation in Preschool/Teacher Observation in Preschool (COP/TOP; Farran & Anthony, 2014), in which observers conduct classroom sweeps where they alternate focus on each child and teacher in the classroom, recording what they are doing in that moment.

Classroom Observations Serve Multiple Purposes in ECE

Classroom observation tools can serve multiple purposes related to quality improvement in ECE. In a coaching context, expert teacher coaches typically conduct observations to gather information of a teacher's practice, often using observation scores as guides for focusing their coaching. In an accountability context, scores from classroom observations are used for monitoring or rating. For example, Head Start uses CLASS for monitoring as part of its Designation Renewal System. Specifically, Head Start grantees with average scores below a preestablished threshold on any of the CLASS domains must compete for funding in the next grant cycle (Office of Head Start, 2024a). Additionally, CLASS and ECERS-3 are commonly used as part of the formula to assign ratings for ECE programs in state Quality Rating and Improvement Systems (QRIS), with 75% of QRIS using either CLASS or ECERS-3 (Build Initiative & Child Trends, 2024). Finally, in research, classroom observation scores are often used to assess the effect of a teacher-focused intervention on teaching practice, and to examine associations between high-quality practices and children's learning outcomes.

Despite their widespread use, there are limitations to the effectiveness of classroom quality observations to support high-quality teaching. Associations between these measures and children's outcomes are often modest, raising questions about their ability to capture the key ingredients of high-quality ECE (Burchinal, 2018). Relatedly, classroom observations only reflect a snapshot in time; they are often conducted as infrequently as once a year and interpreted to reflect children's typical experience in the classroom (Weisberg et al., 2009). One reason for this infrequency is that observations are resource-intensive—requiring highly trained observers and taking several hours. The costs can be even greater for programs that need observers with specialized skills, such as a multilingual observer, or for programs in rural areas where observers must travel long distances between programs. Further, teachers sometimes report that having their classroom observed can be anxiety-inducing (Lasagabaster & Sierra, 2011), which can negatively impact their performance and the overall classroom dynamic (Bottiani et al., 2019; Hamre, 2014; Roberts et al., 2016). These constraints further limit the utility of observations; results are often not shared back with teachers, let alone in a timely manner or in a way that informs and incorporates coaching (Lasagabaster & Sierra, 2011).

Video Scoring of Observations May Better Support More Accurate and Scalable Measurement

Conducting video observations may address some of these challenges. Video recordings may allow observers to see aspects of classroom practice that are difficult to capture during a live observation, such as how early educational experiences differ for racially, culturally, and linguistically minoritized learners (Meek et al., 2021). Within the context of instructional coaching, providing teachers with the opportunity to view recordings of their classrooms may help teachers better understand the strengths and weaknesses of their practice (Pianta et al., 2014). Finally, videos may reduce costs, allow for more frequent observations, and enable coaches to share specific video examples with teachers (Clark et al., 2022; Kane et al., 2020).

Although CLASS was designed for live observations, several studies have used video coding (e.g., Farley et al., 2017; Justice et al., 2017; Pianta, Mashburn et al., 2008), and CLASS training is conducted over video. One study, to our knowledge, compared live and video observations of ECE classrooms, finding that CLASS video observations functioned well and that differences between live and video scores were minimal (Curby et al., 2016). ECERS-3 observations have not been systematically conducted on video (with a few exceptions during the COVID-19 pandemic), while observations for other tools such as ACSES were developed to be conducted over video (Curenton et al., 2022).

The Early Childhood Classroom Observation (ECCO) study is designed to better understand how video observations can support measurement of early learning classroom quality. The study team gathered evidence of the reliability and feasibility of video observations across a range of ECE settings and across different tools by conducting classroom observations of pre-K classrooms live and on video using CLASS and ECERS-3. We also conducted interviews and surveys with pre-K teachers and program leaders to determine the materials and conditions needed for conducting reliable video recordings of pre-K classrooms. In this study, we address four overarching research questions, with related sub-questions:

(1) Are video observations of pre-K classroom quality reliable?

a. Do observations of pre-K classroom quality demonstrate reliability, including inter-rater reliability, test-retest reliability, and internal consistency, when conducted over video?

b. Do observers assign similar ratings to the same observation scored on video and live (i.e., inter-rater reliability across live and video)?

(2) Do live and video observations of pre-K classroom quality produce equivalent ratings?

a. Do live and video scores vary systematically within an observation?

b. Do live and video scores vary differentially by program and observation characteristics?

c. Do live and video scores vary in meaningful ways for coaching, accountability, and research use cases?

(3) What are teachers' perceptions of the use of video recordings for observations in pre-K?

a. How do video observations change teachers' experiences with classroom observations?

b. In what ways do teachers envision using video observations?

(4) What are the approximate costs of conducting video observations relative to live observations?

Study Methods

Sample

The study team recruited a diverse set of 27 center-based preschool programs to participate in the study. We identified potential programs through lists provided by state and local education agencies, as well as through study team members' professional contacts and communities. Up to six classrooms participated at each program. Within these 27 programs, 142 teachers (82 lead teachers and 60 assistants) in 59 classrooms participated in the study. We conducted a total of 160 live observations in these classrooms.

Programs were in five states or districts: California (3 programs), Maryland (1 program), Massachusetts (8 programs), Virginia (1 program), and Washington, D.C. (14 programs). Program auspices included public pre-K, public charter schools, community-based childcare centers, faith-based programs, and cooperative nursery schools. Programs implemented district-mandated curricula (e.g., Creative Curriculum, HighScope) and instructional models (e.g., Montessori, Reggio Emilia, dual immersion). All programs were in urban or suburban areas. The sample included classrooms that were racially, culturally, and linguistically diverse; that included children with disabilities; and that were multilingual settings where both English and Spanish were spoken.

We surveyed lead teachers whose classrooms were observed as part of this study about their experiences with live and video observations along with their demographic characteristics. Fifty-eight lead teachers (71% of participating lead teachers) completed the survey. Most who completed the survey (87%) identified as women; 6% identified as Asian, 27% as Black, 18% as Hispanic/Latine, 36% as white, and 5% as multiracial. Among teachers who completed the survey, the median years of experience teaching in ECE settings was 5 years, ranging from 1 year to 45 years. More than half of teachers (71%) held at least a bachelor's degree. We also conducted interviews with 11 lead teachers and six program leaders regarding their experiences with live and video observations. Full sample descriptives are in Appendix A (Exhibit A1).

Data Collection

Data used in this study are drawn from multiple sources, including classroom observations, coding of environmental and technical constraints present in video recordings, teacher surveys, and teacher and program leader interviews. The data sources, organized by research question, are presented in Exhibit 1.

Exhibit 1. Research questions and data sources

Research Questions	Data Sources
1. Are video observations of pre-K classroom quality reliable?	▪ Observations (CLASS and ECERS-3)
2. Do live and video observations of pre-K classroom quality produce equivalent ratings?	▪ Observations (CLASS and ECERS-3) ▪ Coding of environmental and technical constraints
3. What are teachers' perceptions of the use of video recordings for observations in pre-K?	▪ Teacher survey ▪ Teacher and program director interviews
4. What are the approximate costs of conducting video observations relative to live observations?	▪ Cost analysis

Classroom observations

The study team conducted classroom observations using the CLASS 2nd Edition (Pianta & Hamre, 2022) and ECERS-3 (Harms et al., 2014).

CLASS measures three domains of process quality through 20-minute observation cycles across which observers rate on a scale of 1 (*low quality*) to 7 (*high quality*). Each domain includes 3–4 dimensions:

- **Emotional Support:** Four dimensions (Positive Climate, Negative Climate, Educatory Sensitivity, Regard for Child Perspectives) assess support of children's relationships, with higher scores indicating that teachers provide a secure base for learning.
- **Classroom Organization:** Three dimensions (Behavior Management, Productivity, Instructional Learning Formats) assess how teachers manage child behavior and instructional time, with higher scores indicating a productive and engaged classroom with developmentally appropriate expectations.
- **Instructional Support:** Three dimensions (Concept Development, Language Modeling, Quality of Feedback) assess teachers' ability to support children's cognitive and language skills, with higher scores indicating opportunities for brainstorming and creativity with appropriate scaffolding and language feedback.

Multiple validation studies have established inter-rater reliability (IRR) and internal consistency for both CLASS dimensions and domains. CLASS dimensions demonstrate high internal consistency in preschool classroom observations: Emotional Support (α = .91), Classroom Organization (α =.87), and Instructional Support (α =.86; Pianta, La Paro et al., 2008). CLASS dimensions are listed in Appendix B (Exhibit B1).

ECERS-3 is designed to assess the quality of early childhood classrooms serving children 3–5 years of age. It measures both process and structural quality through a 3-hour observation during which the observer follows children wherever they go in a program. It is designed to measure physical space, groupings, materials, instruction, health, and safety. ECERS-3 consists of 35 items organized into six subscales: Space and Furnishings, Personal Care Routines, Language and Literacy, Learning Activities, Interaction, and Program Structure. Observers respond to a series of yes/no indicators for each item that are anchored to a 7-point scale. Scores are labeled as inadequate (1), minimal (3), good (5), or excellent (7). One study found ECERS-3 items had high internal consistency when used as a total score, with a Cronbach's alpha of .93, and statistically significant correlations with CLASS pre-K (Early et al., 2018). The ECERS-3 scales are listed in Appendix B (Exhibit B2). Key differences between the measures are outlined in Exhibit 2.

Exhibit 2. Key differences between CLASS and ECERS-3

	Classroom Assessment Scoring System 2nd Edition (CLASS)	Early Childhood Environmental Rating Scale (ECERS-3)
Quality Type	Process	Process + structural
Administration	Four 20-minute cycles	3 hours + additional time to explore materials
Scoring System	10 dimensions, scored 1–7 based on observer evaluation of the quality of behavioral markers	35 items, scored 1–7 based on the indicators checked off during the observation
Precedent for Video Use	Published video protocols; training conducted on video	Limited video use during COVID-19 public health emergency

Certified CLASS and ECERS-3 observers conducted observations between May 2023 and August 2024. Fourteen observers conducted CLASS observations, and four observers conducted ECERS-3 observations. Four CLASS observers and two ECERS-3 observers were bilingual in English and Spanish—these observers conducted the observations in classrooms where program staff indicated that teachers used Spanish.

The observation team conducted live observations according to instrument guidelines as described in Harms et al. (2014) and Pianta & Hamre (2022). During the live observation, the observer set up a Swivl recording device, a tripod device that rotates to track a teacher's movements. An iPad with a fisheye lens was attached to the device to record as much of the classroom as possible, with a focus on the lead teacher. We used three Swivl microphones (worn by the lead teacher, an assistant teacher, and the observer) to record audio. After completing the observation, the observer stored the videos on a secure cloud server for later scoring by a different observer.

Live and video observations were conducted by the same group of observers. All primary observers conducted at least one observation live and at least one on video. Each live observation was coded by one observer. Later, another observer independently coded the video recording of that observation. The video observer was blind to the scores given by the live observer. Observers were assigned observations via quasi-random assignment, considering availability, workload considerations, and where relevant, Spanish language skills. To assess IRR on video, a subset of observers double-coded video observations.

To mirror live observations as closely as possible, both CLASS and ECERS-3 observers were instructed to conduct observations without rewinding or fast-forwarding the videos, although they were permitted to pause occasionally for breaks. This approach is consistent with CLASS guidance for video coding as described in Pianta & Hamre (2022). Because ECERS-3 does not have published guidance for video coding, we collaborated with a certified ECERS-3 trainer to develop a coding protocol. This protocol mirrored the CLASS guidance but included a provision that allowed observers to mark any indicator as *Cannot Score* if the video format prevented them from scoring the item. Video observers never scored certain indicators or full items that would require leaving the classroom (e.g., space for gross motor, gross motor equipment, supervision of gross motor). For details on how we incorporated indicators marked as *Cannot Score* into the ECERS-3 scoring procedure, see Appendix C.

Coding of environmental and technical constraints

The study team developed a rating scale to identify the environmental and technical constraints that created challenges in completing the recordings. Challenges with the video component included classroom layout (e.g., columns or walls in view; cubbies, shelves, hidden nooks) and video technology (e.g., poor camera placement, camera focus, functionality of lead teacher camera tracking). Audio challenges include teacher microphone technology and difficulty hearing other adults or children in the classroom who were not wearing microphones. For more details on the coding protocol, see Appendix D.

After coding a consensus video, three coders watched the first 30 seconds of every 10-minute block, toggling between multi-teacher audio streams. They rated the recording on six video and five audio items on a 5-point scale ranging from 1 for *uncodable* to 5 for *no issues*.

Teacher survey

The study team created a brief survey for participating lead teachers to collect their perspectives on the accuracy, usefulness, and burden of video and live observations, covering their participation in the current study as well as any prior experiences. The survey included 23 questions organized into six sections: (1) background information, (2) experience with video observations during this study, (3) past experience with live observations, (4) past experiences with video observations, (5) comparative reflections on observation methods, and (6) demographic information. Participating teachers received a personalized survey link via email within 1 week of their observation. The survey took approximately 10–15 minutes to complete.

Teacher and program leader interviews

The study team conducted 45-minute virtual interviews with 11 teachers and six program leaders to provide deeper context for the survey findings. All 11 teachers had been observed during the study, and all six program leaders were from participating programs. Using a semi-structured protocol, interviewers asked participants about the potential disruption caused by both observation methods and about teachers' likelihood to trust results from live versus video observations. Each interview was recorded and transcribed.

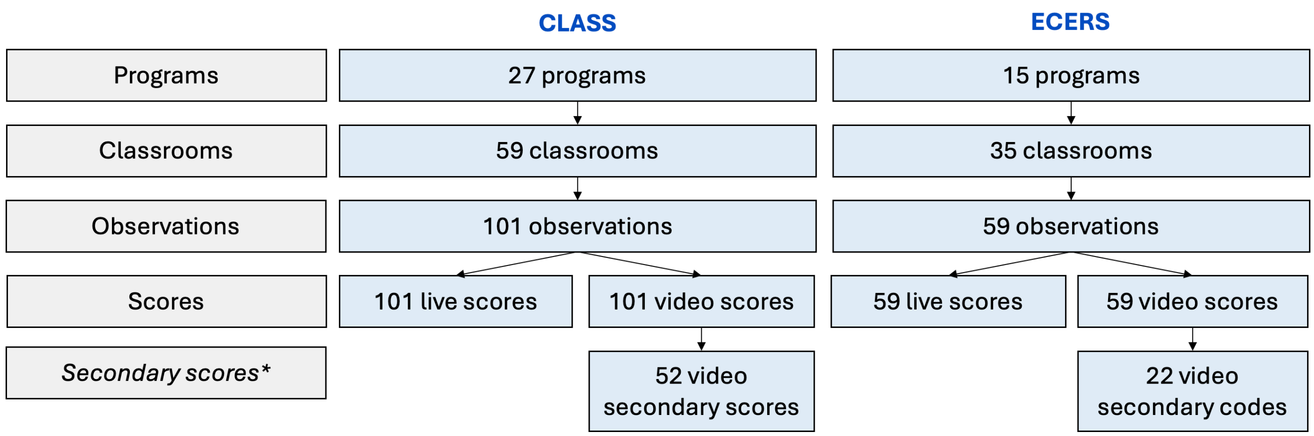
Cost analysis

The study team collected information about costs to conduct live and video observations using CLASS and ECERS-3 tools throughout the study. We used these data to calculate and analyze net costs for the actual observations conducted in this study.

Analytic Methods

Observation data were structured with multiple levels of nesting (Exhibit 3). Below, we describe the analytic methods for each research question and note the level at which each analysis was conducted.

Exhibit 3. Nesting structure of classroom observation data



*** Secondary scores are used only for the video IRR analysis.*

RQ 1a: Do observations of pre-K classroom quality demonstrate reliability when conducted over video?

The study team calculated reliability for video scoring for CLASS and ECERS-3 in the following ways:

- **Inter-rater reliability** for all double-coded video scores using percent-within-one IRR. Percent-within-one IRR represents the percentage of dimensions (CLASS) or items (ECERS-3) in which two observers provided scores within 1 point of each other. This approach is used when determining whether novice CLASS or ECERS-3 observers can be certified to conduct live observations. Consistent with certification standards used to certify observers on both measures, the level of analysis for CLASS was an observation cycle score and for ECERS-3 was a full observation score. We used the cutoff of 80% or higher as adequate IRR (Harms et al., 2014; Pianta & Hamre, 2022).
- **Test-retest reliability** for each classroom that observed at multiple timepoints, using percent-within-one calculations reflecting the percentage of dimensions (CLASS) or items (ECERS-3) in which scores on two observations in the same classroom are within 1 point of each other. We used 80% or higher as the threshold for adequate reliability.
- **Internal consistency statistics** (Cronbach's alphas) for each video observation score for each subscale of CLASS and ECERS-3. An alpha of .70 or higher is considered desirable, although lower can also be acceptable, especially for scales with fewer than 10 items (Taber, 2018).

RQ 1b: Do observers assign similar ratings to the same observation scored on video and live?

The study team calculated percent-within-one IRR comparing scores produced by a video observer to those produced by a live observer for the same observation. Specifically, we calculated IRR for each observation cycle (CLASS) or observation (ECERS-3) by considering the live observer as Rater 1 and the video observer as Rater 2, using the 80% threshold used to certify observers.

RQ 2a: Do live and video scores vary systematically within an observation?

To address this question, the study team conducted multivariate regression analyses to predict *scores* from observation modality (live or video; see Appendix E for details of the analytic models). The models incorporated fixed effects at the *observation* level, accounting for all factors that remained constant within an observation. These include measurable factors that theory and prior research suggest may be related to classroom quality, such as teacher education and experience, as well as unobservable characteristics of the classroom, including features of high-quality teaching not captured by the tool.

The models also incorporated controls for variables that vary across live and video observations: (1) observer identity and (2) number of days since the observer was last certified in the measure. Even though CLASS and ECERS-3 observers achieved reliability standards for their respective observations, we included a control for observer identity because observers were not randomly distributed across live and video observations and may have individual differences in their scoring. We also controlled for the number of days since last certified to control for observer drift (Casabianca et al., 2015). The analysis used classroom-level cluster-robust standard errors to address the multilevel data structure.

RQ 2b: Do live and video scores vary differentially by program and observation characteristics?

The study team tested whether live and video scores varied differentially across several program and observation characteristics (see Appendix A for descriptive statistics):

- **Program type:** Program type data was drawn from study administrative records. We created a variable that collapsed program types into two categories: (1) public programs, including public pre-K and public charter school pre-K; and (2) non-public programs, including community-based childcare centers, faith-based programs, and cooperative nursery schools. Fifty-one percent of CLASS observations and 42% of ECERS-3 observations were conducted in public programs.
- **Language:** Observation language data was drawn from study administrative records in which observers reported the language used in the classroom following live observation. We created a variable that indicated whether an observation was (1) monolingual English, meaning that no language besides English was used more than superficially (e.g., to count to 10 or sing songs); or (2) multilingual

English-Spanish, meaning that substantial Spanish was used in the classroom. Fifteen percent of CLASS observations and 24% of ECERS-3 observations were multilingual in English and Spanish.

- **Observed quality:** We examined moderation in observation scores by quality level, using both data- and policy-driven approaches to operationalize quality that were based on the live observation scores to reflect business-as-usual conditions. For the data-driven approach, we estimated models with a continuous interaction term, using live scores. Additionally, we estimated models where we split the data into sample-specific quintiles and collapsed the middle three quintiles into one group, creating categories of low-quality (bottom quintile), average-quality (middle three quintiles), and high-quality (top quintile) classrooms. For the policy-driven approach, we categorized observations by whether they met the Head Start competitive and quality thresholds for CLASS and the typical QRIS thresholds for ECERS-3. For the CLASS Emotional Support, 4% of observations did not meet the Head Start competitive threshold and 60% met the quality threshold. For Classroom Organization, 12% did not meet the competitive threshold and 36% met the quality threshold. For Instructional Support, 45% did not meet the competitive threshold and just 27% met the quality threshold. For the ECERS, 42% of observations scored high enough to meet the typical QRIS 3-star level. See Appendix E (Exhibits E1 and E2) for additional details.
- **Video and audio technology challenges:** Data were drawn from the coding of environmental and technical constraints. We focused moderation analyses on the issues that most reflected malleable components of conducting video observations in classrooms, including those related to video technology (camera focus, Swivl placement, primary marker tracking) and microphone technology (primary microphone, secondary microphone). Analyses for the other more environmental constraints are provided in Appendices F and G. We collapsed the 5-point Likert-type scale originally assigned into three categories: (5) no issues, (4) minimal issues that likely did not affect coding, and (3) issues that affected coding (collapsed ratings of 2–3). No observations were rated (1) uncodable across any CLASS or ECERS-3 observations.

For each interaction variable, we estimated separate models between the interaction term and video indicator. Details on the models used are in Appendix E.

RQ 2c: Do live and video scores vary in meaningful ways for coaching, accountability, and research?

The study team also aimed to identify whether scores on live and video observations have meaningful differences for three use cases: **coaching, accountability,** and **research**.

Coaching use cases: We used the predicted values of the regression models estimated in RQ 2a to identify the lowest scoring dimension (CLASS) or item (ECERS-3) within each domain/subscale for each observation. Although coaches take many factors into account to decide where to support teachers, understanding whether the data point they are most likely to use to decide (i.e., the lowest scoring dimension or item) varies across live and video observations can provide insights into how coaches might reach different conclusions about where to focus improvement efforts depending on the observation method.

Accountability use cases: We predicted whether a CLASS observation would score below the Head Start competitive threshold (which requires a grantee to compete for renewed funding), at the competitive but below the quality threshold, or above the quality threshold (Office of Head Start, 2024b). For ECERS-3, we predicted whether an observation would meet the average ECERS-3 score required to reach the 3-star level in most QRIS that use ECERS-3. Of the 12 state QRIS with ECERS-3 cutoffs to reach the 2-star level, the cutoff is typically 3.5 for the average scores across all ECERS-3 items (Build Initiative & Child Trends, 2024).

Research use cases: To understand the implications of using video observations for research, we explored the extent to which within-observation variance in scores was explained by video compared to a common source of variance in live observations: observers. To do so, we compared adjusted R-squared values for models with a control for observer to those that control for both the observer and the video indicator. As a robustness check, we also explored the residual variance in random effects multilevel models controlling for observer compared to models controlling for modality, with observations nested in classrooms.

RQ 3: What are teachers' perceptions of the use of video recordings for observations in pre-K?

For teacher surveys, the study team produced simple descriptive statistics related to teachers' demographic information, their experience participating in our study, and their experience with both live and video observations. We analyzed teacher and program leader interviews using Braun and Clarke's (2006) six-step approach to thematic analysis. The interview team developed a preliminary list of codes based on the semi-structured interview protocol, and inductive sub-codes were added during the coding process. For coding reliability, two study team members used Dedoose to double-code 24% of transcripts to establish reliability before independently coding the remaining transcripts. Once each transcript was coded, two study team members engaged in an iterative process to identify emergent themes.

We use the following guidance to characterize survey and interview responses in this report: "few" refers to fewer than 25% of participants, "some" refers to 26% to 50%, "many" refers to 51% to 75%, "most" refers to 76% to 90%, and "nearly all" refers to 91% to 100%.

RQ 4: What are the approximate costs of conducting video observations relative to live observations?

The study team conducted a cost analysis using the ingredients method (Levin et al., 2018). The ingredients method is a detailed process for identifying and calculating the cost of all resources needed to implement a program, including personnel, materials and equipment, facilities, and other inputs. For this study, we focused on personnel and materials and equipment but did not include the construction costs of the classrooms. No ingredients were categorized as "other."

Cost components included training fees, labor hours, travel, and equipment costs for SRI staff to conduct the live and video observations. For districts and programs that plan to conduct these observations, some of these costs would represent net costs. Net costs are expenditures unique to conducting the observation and would be beyond an ECE program's standard operating budget, such as fees to train staff to use the CLASS and ECERS-3 tools. Training and labor hours and travel would represent opportunity costs, or the value of a resource that is used for one purpose so that it cannot be used for another, such as the time needed to train an observer so their time could not be spent on completing other duties.

We used the data collected from the study team's observations and translated these into cost estimates for staff in schools or ECE programs to conduct both live and video observations with the CLASS and ECERS-3 tools. We used the same labor hours, travel, and equipment, but assumed the observer would be an instructional coordinator who trains or coaches teachers. According to data gathered from the Bureau of Labor Statistics, the national mean hourly wage for this position is $37.12 (Bureau of Labor Statistics, 2024). We assumed the recorder would be an audio/video technician sent by the program or district to set up the recording equipment; the national average hourly wage for this position is $29.48 (Bureau of Labor Statistics, 2024).

In terms of labor, a live CLASS observation requires 3.5 hours on average, and a live ECERS-3 observation requires 4.5 hours; these estimates include 1 hour of travel to arrive at the school site. A video CLASS observation requires 2 hours of recording and 3 hours of coding, while a video ECERS-3 observation requires 3 hours of recording and 5 hours of coding.

Observations could be conducted in a variety of locales requiring varying amounts of travel regardless of whether the observation method was live or video. We created two travel categories: local and distant. Local travel was defined as travel that is less than 20 miles, does not require a per diem for meals, and does not require a hotel for overnight stay; mileage costs were calculated using a fully loaded rate of $0.655 per mile. Distant travel was defined as travel that is at least 50 miles using the same mileage rate (total cost of $32.75), requires a per diem for meals ($16), and requires an overnight stay at a hotel (estimated at $202.94).

Findings

Reliability of Video Observations (RQ 1)

We found that all CLASS domain scores and the ECERS-3 total score meet reliability standards typically used to certify observers when conducted on video (RQ 1a; see Exhibit 4 for reliability statistics and Appendix H for details on analyses). When two different observers coded the same video, at least 80% of scores they assigned were within 1 point of each other. Additionally, when the same classroom was coded on video at two different timepoints, the classrooms received scores within 1 point at least 80% of the time.

With respect to internal consistency—that is, the extent to which CLASS dimension scores are consistent within a domain or ECERS-3 item scores are consistent within a subscale—all CLASS domains and the CLASS total score had good internal consistency (α = .72–.86), as did the ECERS-3 total score and the Learning Activities and Interaction subscales (α = .77–.89). The Program Structure subscale had acceptable reliability (α = .63), but the ECERS-3 Space and Furnishings, Personal Care Routines, and Language and Literacy subscales had unacceptable consistency (α = .33–.58) and therefore could not be used as subscale scores in later analysis.

Further, observers assigned similar ratings to the same observation scored on video and live. Specifically, when one observer coded on video and the other coded live, they assigned scores within 1 point of each other at least 80% of the time (RQ 1b). The one exception is the ECERS-3 Personal Care Routines subscale, for which the IRR across live and video was 77%.

Exhibit 4. Reliability statistics for CLASS and ECERS-3

	CLASS	ECERS-3
Inter-rater reliability on video	88%	88%
Test-retest reliability	88%	87%
Inter-rater reliability across live and video	88%	86%

Note. *Details on reliability analyses are provided in Appendix H.*

Summary: Video observations using CLASS and ECERS-3 tools demonstrate reliability comparable to live observations.

All CLASS domains and the ECERS-3 total score meet typical percent-within-one reliability standards when conducted on video. IRR and consistency across timepoints are generally strong, with most scores falling within acceptable ranges. Internal consistency is good for CLASS domains and most ECERS-3 subscales, although some ECERS-3 subscales show unacceptable consistency. Comparisons between video and live observations reveal similar ratings, with scores typically within 1 point of each other. However, some ECERS-3 subscales demonstrate low IRR across live and video formats. These findings suggest that video observations can provide reliable measures of classroom quality, although some specific areas may require additional attention when using video formats.

Comparison of Live and Video Observation Scores (RQ 2)

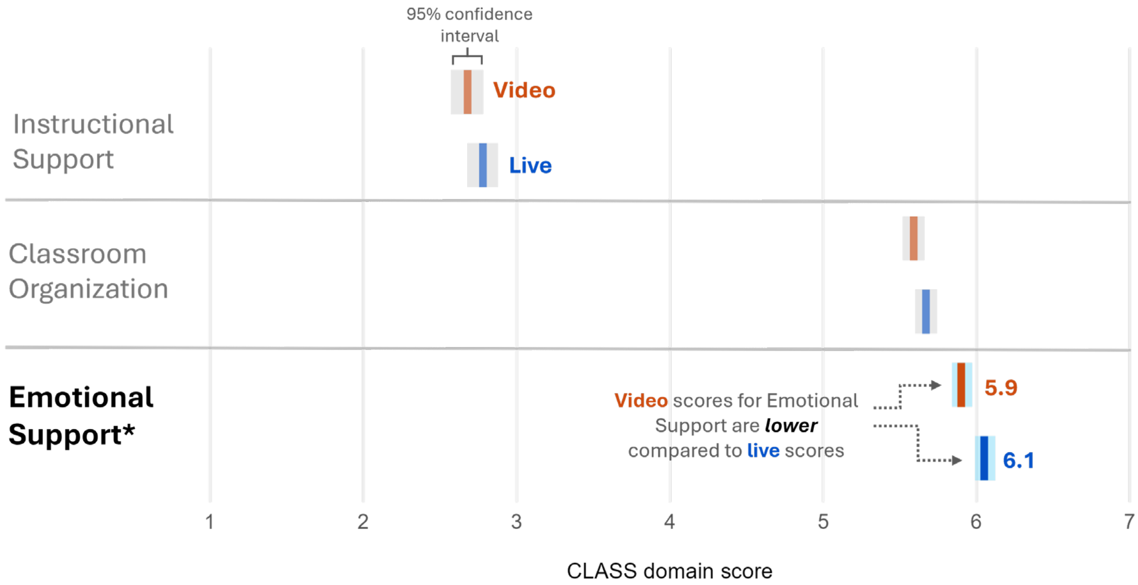
Within-observation comparison of live and video scores (RQ 2a)

Overall, we found few statistically significant differences between live and video CLASS and ECERS-3 scores of the same observation.

Video scores for CLASS Emotional Support domain are slightly higher than live scores, but there are no differences between live and video scores for the other CLASS domains.

CLASS Emotional Support scores produced via a live observation were, on average, higher than those produced via video scoring of the observation, but differences were small: Live scores were approximately 0.2 points higher than video scores (Exhibit 5). We did not observe statistically significant differences between live and video scores for the Classroom Organization or Instructional Support domains.

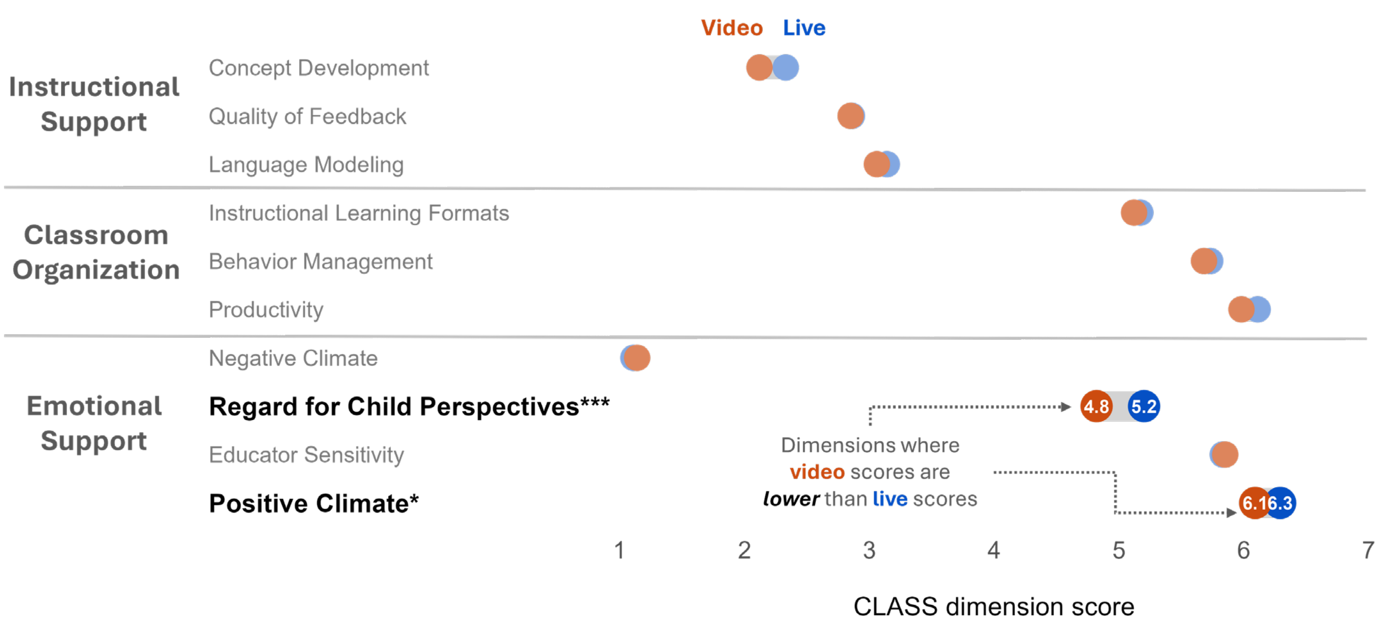
Exhibit 5. Comparison of live and video CLASS domain scores



Note. *Graph shows regression-adjusted means. See Appendix F (Exhibit F1) for detailed regression results.*
***** *p < .001, ** p < .01, * p < .05*

Within the Emotional Support domain, scores were statistically significantly higher on live for both the Positive Climate and Regard for Child Perspectives dimensions, but scores were not different for Negative Climate or Educator Sensitivity (Exhibit 6).

Exhibit 6. Comparison of live and video CLASS dimension scores

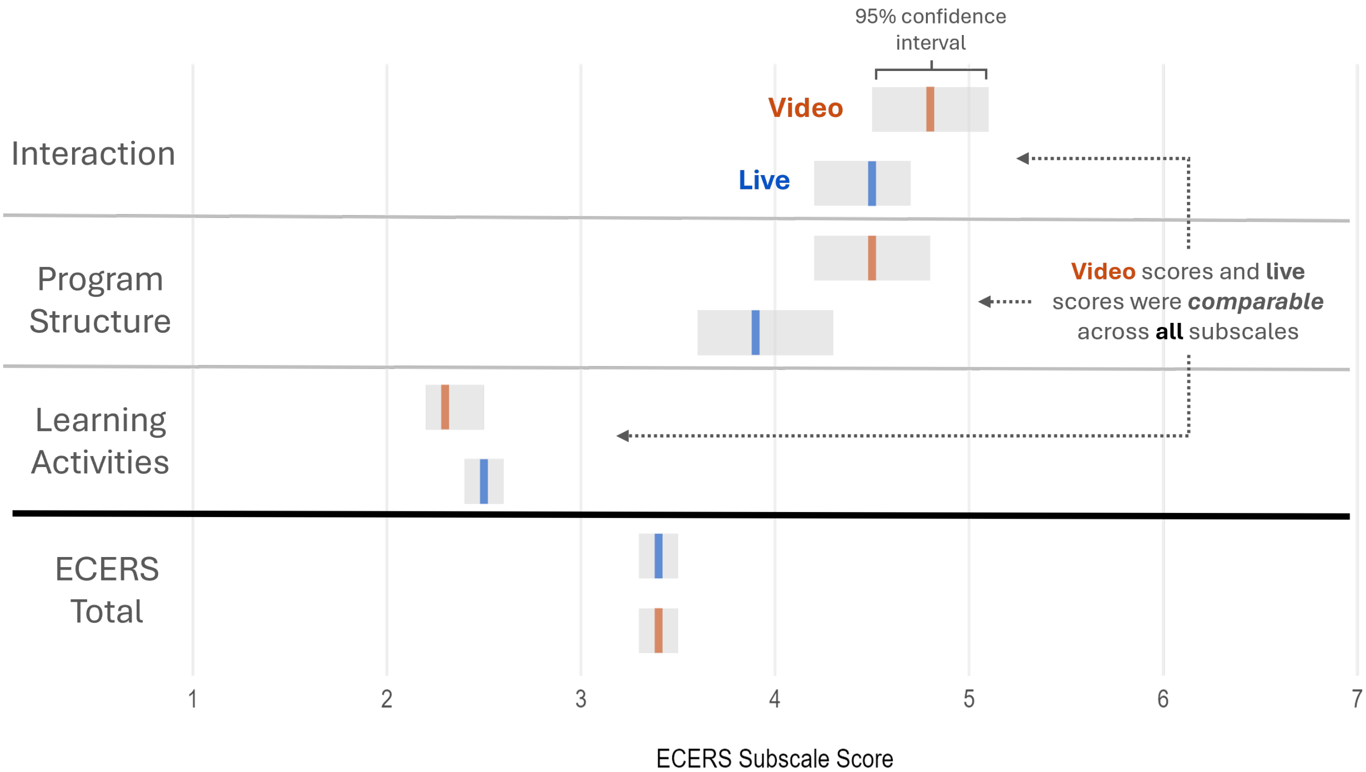


Note. *Graph shows regression-adjusted means. See Appendix F (Exhibit F2) for detailed regression results.*
***** *p < .001, ** p < .01, * p < .05*

Live and video scores did not vary across ECERS-3 subscales, but they did vary across several ECERS-3 items.

For the ECERS-3 subscales with acceptable internal consistency (see Appendix C for details on internal consistency), we did not observe statistically significant differences across live and video scores within an observation (Exhibit 7).

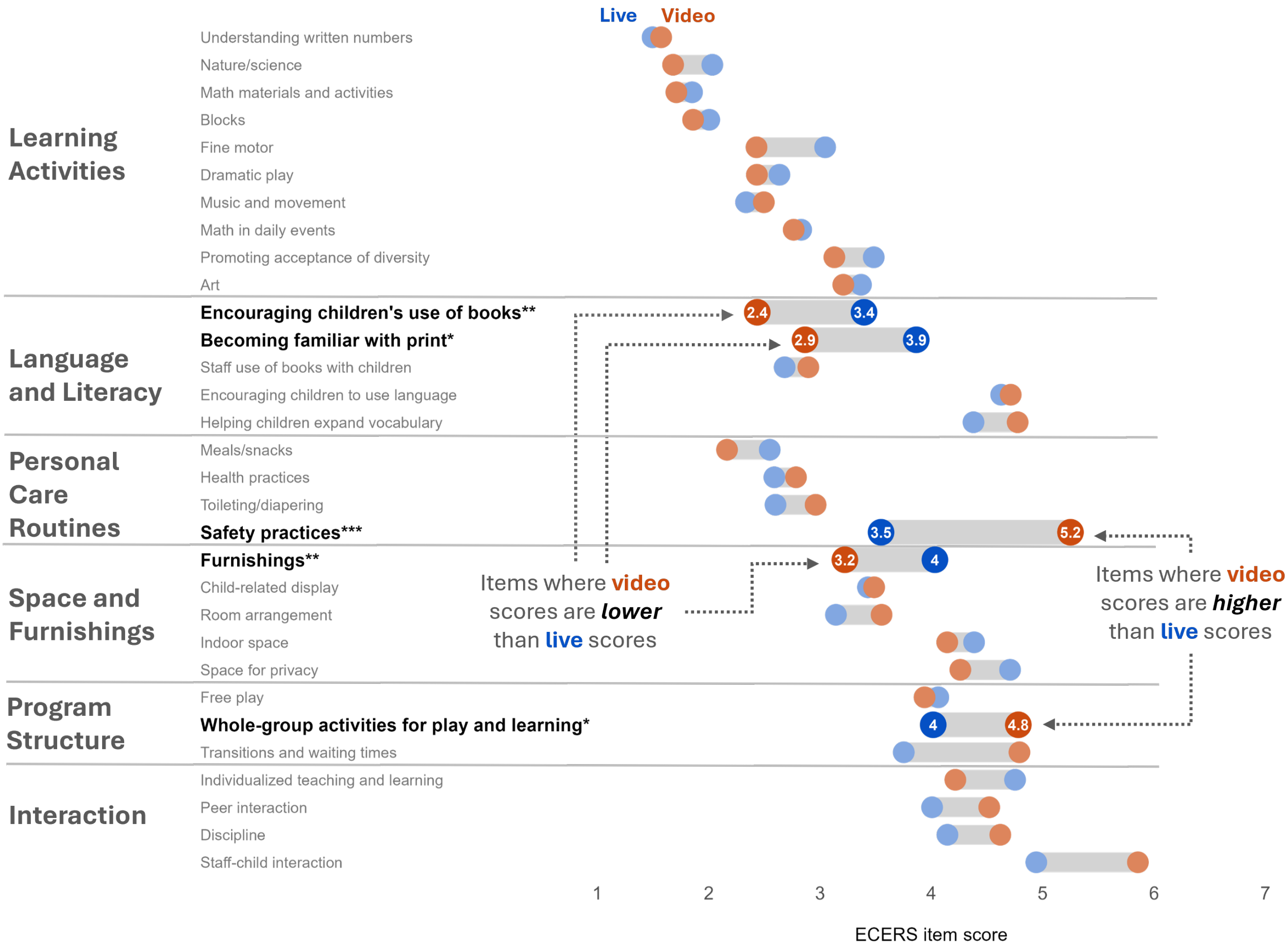
Exhibit 7. Comparison of live and video ECERS-3 subscale scores



Note. *Graph shows regression-adjusted means for ECERS-3 subscales with sufficient internal consistency. See Appendix G (Exhibit G1) for detailed regression results.*
***** *p < .001, ** p < .01, * p < .05*

However, these subscale averages mask variation between the scores produced by live and video observations for some specific ECERS-3 items: live and video scores were significantly different for 5 of the 31 included ECERS-3 items (16%; Exhibit 8). Unlike CLASS, in which live scores were consistently higher, the pattern of differences across live or video scores varied across ECERS-3 items. Live scores were higher than video scores for furnishings for care, play, and learning; encouraging children's use of books; and becoming familiar with print; but scores were higher on video compared to live for safety practices and whole-group activities for play and learning.

Exhibit 8. Comparison of live and video ECERS-3 item scores



Note. *Graph shows regression-adjusted means. See Appendix G (Exhibits G2–G7) for detailed regression results.*
****** *p < .001,* *** *p < .01,* ** *p < .05*

Summary: Few CLASS and ECERS-3 live and video scores varied systematically within an observation. Where differences exist, live scores were typically higher than video scores for dimensions or items that require child-level evidence or visual cues that reflect *high quality*. However, video scores were higher on some ECERS-3 items that include indicators of *low quality*.

For CLASS, live scores were higher compared to video scores for the Emotional Support domain and the Positive Climate and Regard for Child Perspectives dimensions. One possible reason for this is that both the Positive Climate and the Regard for Child Perspectives dimensions rely on close observation of *children's* language and behavior, not only what the teacher says and does. For example, Positive Climate assesses whether children and teachers share affection and enjoy being in each other's company, and Regard for Child Perspectives assesses whether the teacher follows children's lead and creates a relaxed environment where children can move frequently while still engaged in learning. Our video technology included microphones for two teachers, and the camera tracked the lead *teacher* around the room but did not specifically follow *children*; therefore, video may insufficiently capture child-level evidence.

For ECERS-3, live and video scores did not vary within an observation for any of the subscales; however, the Space and Furnishings, Personal Care Routines, and Language and Literacy subscales were not included in analysis because of insufficient consistency within the subscales. Scores did vary across five ECERS-3 items: As with CLASS, live scores were higher than video scores for furnishings for care, play, and learning; encouraging children's use of books; and becoming familiar with print; but video scores were higher for safety practices and whole-group activities for play and learning.

This differing pattern of results may be due to whether indicators that are difficult to see on video for each of these items are indicators of high quality or low quality. The furnishings, encouraging children's use of books, and becoming familiar with print items all involve many visual indicators of *high quality*, such as the presence of specific furniture, the number of available children's books, and staff actively pointing out letters or words. While visual indicators are common in ECERS-3, these three items contain mostly visual rather than audio indicators, and audio indicators may be easier to notice given the microphones. The safety practices and whole-group activities items, however, include indicators such as the presence of safety hazards and children being distracted. If these indicators are difficult to observe on video, it may result in higher video scores because they indicate lower quality.

Variation in live and video scores by program and observation characteristics (RQ 2b)

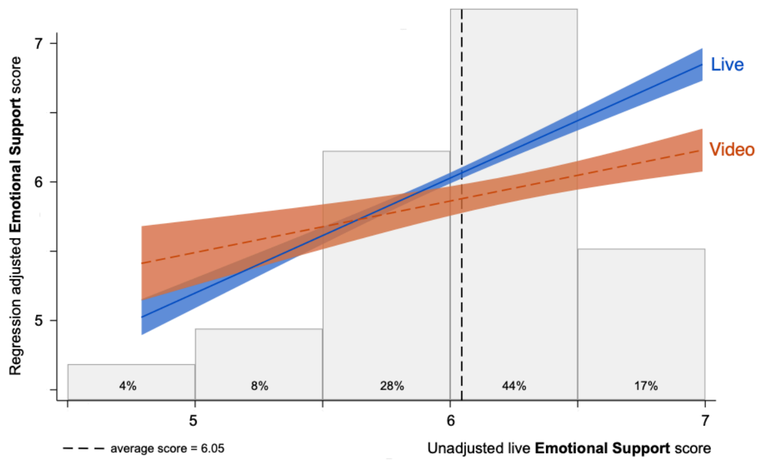
Differences between live and video scores are most pronounced for observations with the highest and lowest scores.

Live and video scores for most CLASS domains and ECERS-3 subscales varied differentially by the quality level of the program (as determined via unadjusted live observation scores). Although we found few differences, on average, in live and video scores across the full sample of observations, we did observe differences in those observations that earned comparatively low and high scores (Exhibits 9 and 10). Exhibits 9 and 10 display the interactions between quality level (unadjusted live score) and video. The dotted line reflects the average unadjusted live scores; the histogram overlayed reflects the distribution of unadjusted live scores.

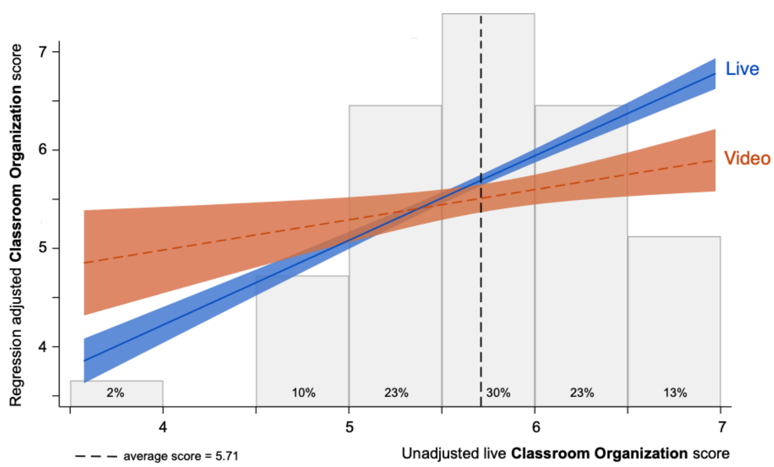
In observations that had the lowest unadjusted live scores, video scores were typically higher compared to live especially for the CLASS Classroom Organization and Instructional Support domains and the ECERS Interaction and Program Structure subscales. This pattern reversed in the highest scoring observations, in which live scores were higher than video scores, for all CLASS domains (Emotional Support, Classroom Organization, and Instructional Support), and for ECERS Learning Activities. Notably, this pattern emerged even for CLASS domains and ECERS-3 subscales for which we did not see differences between live and video scores in our main effect findings.

Exhibit 9. CLASS live and video domain scores vary differentially by quality level

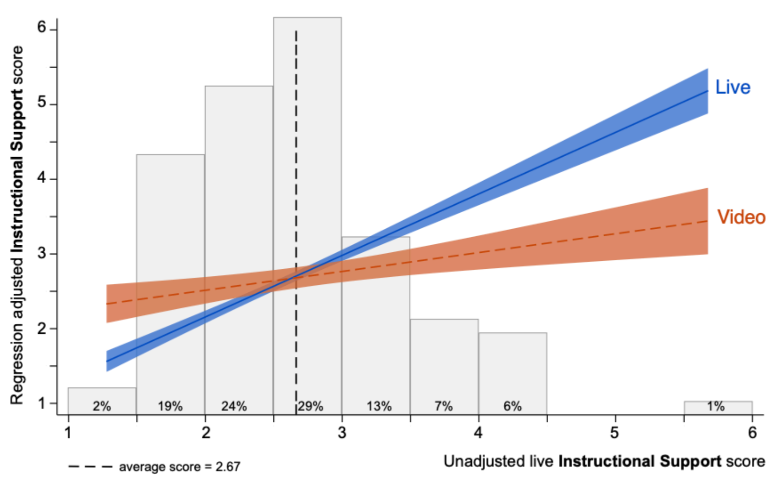
Emotional Support



Classroom Organization



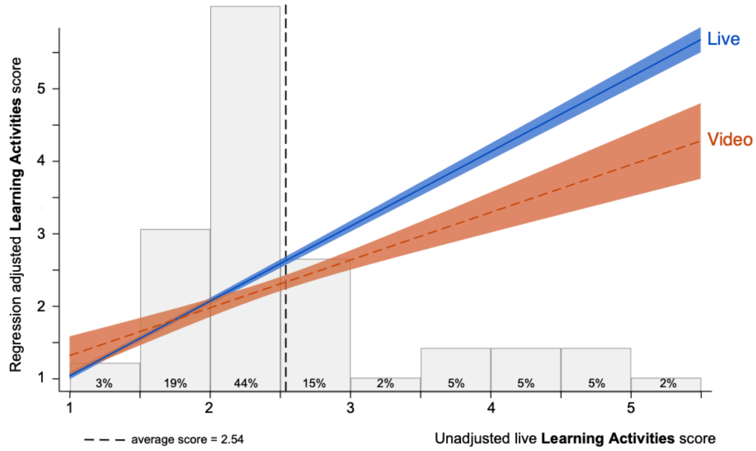
Instructional Support



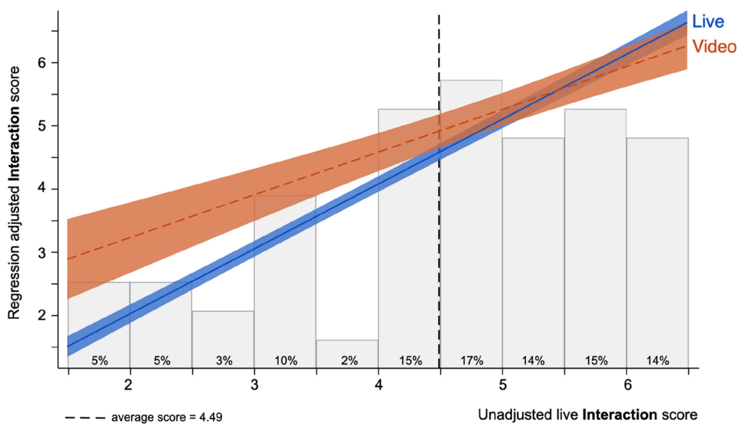
Note. *Graph displays interactions between unadjusted live score and video, overlayed on a histogram of the distribution of unadjusted live scores. To display regression adjusted means for the interaction terms, the graph displays results of between observation models, but the results are substantively unchanged from the within-observation models (see Appendix F, Exhibits F3–F4 for detailed regression results). The same pattern was also found in models using categorical variables for live quality level (Exhibits F5–F6). Note the large differences between live and video Instructional Support scores in high quality classrooms reflect few cases.*

Exhibit 10. ECERS live and video domain scores vary differentially by quality level

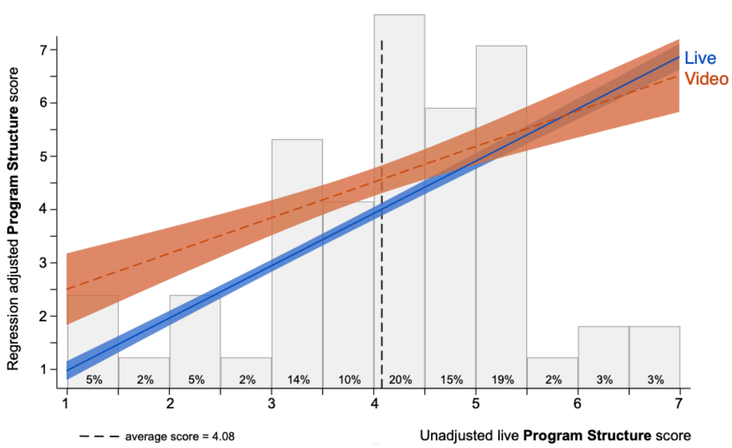
Learning Activities



Interaction



Program Structure



Note. *Graph displays interactions between unadjusted live score and video, overlayed on a histogram of the distribution of unadjusted live scores. To display regression adjusted means for the interaction terms, the graph displays results of between observation models, but the results are substantively unchanged from the within-observation models (see Appendix G, Exhibits G8–G9 for detailed regression results). The same pattern was also found in models using categorical variables for live quality level (Exhibits G10–G11). Note the large differences between live and video Learning Activities scores in high quality classrooms reflect few cases.*

Differences between live and video scores are mostly consistent across program type and language models.

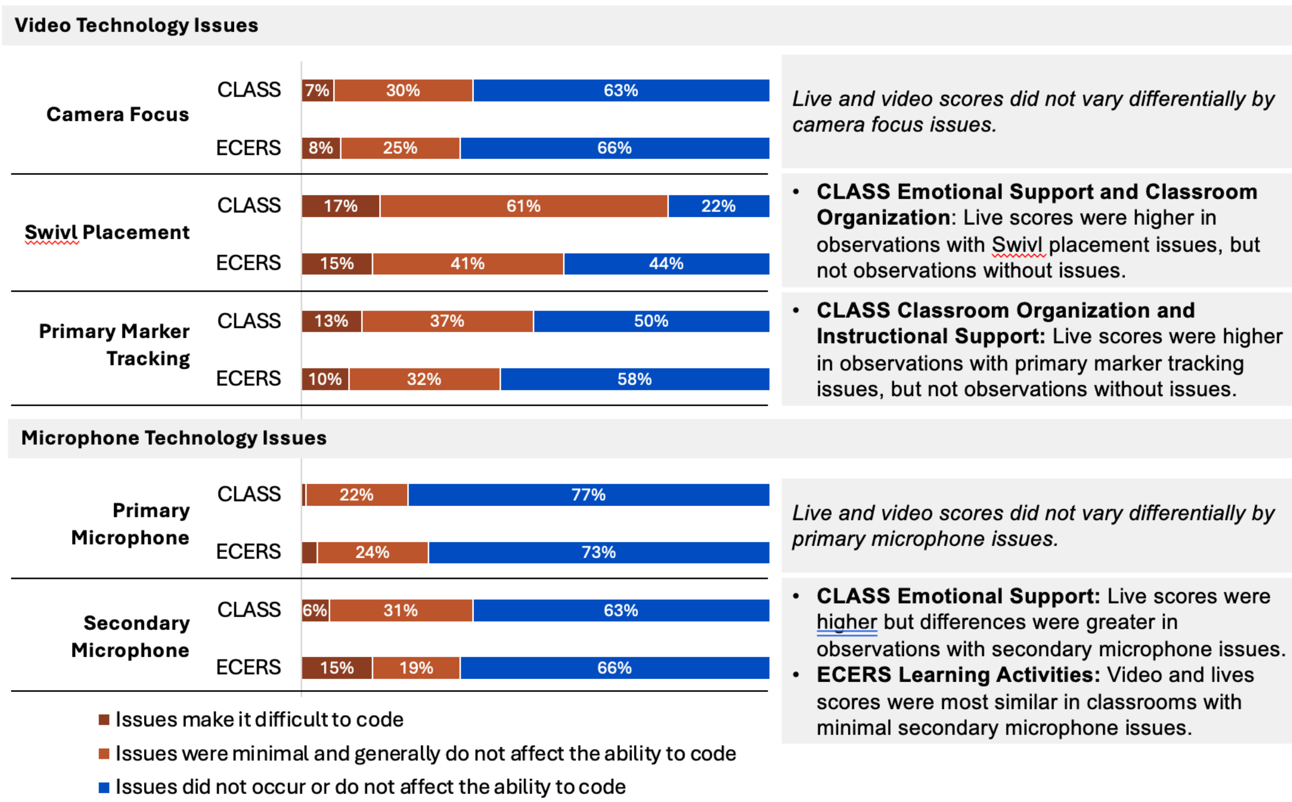
For both CLASS and ECERS-3, live and video scores did not vary differentially across program type (public vs. nonpublic programs), suggesting that video observations function similarly across program auspices (see Appendix F, Exhibit F7; Appendix G, Exhibit G12). Additionally, scores on all ECERS-3 subscales and most CLASS domains did not vary differentially across monolingual English-speaking and multilingual English-Spanish observations (see Appendix F, Exhibit F8; Appendix G, Exhibit G13). The one exception is the CLASS Instructional Support domain: In monolingual observations, live scores were somewhat higher than video scores. In multilingual observations, however, video scores were higher than live scores.

Live and video CLASS scores vary differentially by challenges with video and microphone technology.

We also found that live and video scores varied differentially by video and microphone technology challenges, especially for CLASS (Exhibit 11). For both the Emotional Support and Classroom Organization domains of CLASS, live scores were higher than video scores in observations that were categorized as having challenges with Swivl placement (as most observations did), but not in observations that had no issues with Swivl placement. Likewise, live scores were higher than video scores on both the Classroom Organization and Instructional Support domains in observations that had issues with primary marker tracking, but not in classrooms without such issues.

With respect to microphone technology, although CLASS Emotional Support live scores were higher than video scores across all observations, these differences were greater for observations that had issues with the secondary microphone. However, we observed the opposite pattern with the ECERS-3 Learning Activities subscale, for which scores were most similar across live and video observations in classrooms with secondary microphone issues.

Exhibit 11. Summary of moderation by video and microphone technology issues



Note. *See Appendix F (Exhibits F9–F10) for CLASS models and Appendix G (Exhibits G14–G15) for ECERS-3 models. Additional regression models for analyses of moderation by environmental and technical constraints not displayed here are available in Exhibits F11–F12 and G16–G17.*

Summary: Differences in live and video CLASS and ECERS-3 scores varied by the quality level of the observation and technology challenges. While live and video scores did not vary differentially across program types for CLASS or ECERS-3, CLASS scores varied differentially by languages spoken.

Across both CLASS and ECERS-3, live and video scores varied differentially by the level of the live scores, with higher scores on video for the lowest scoring observations and higher scores live for the highest scoring observations. One explanation for this pattern is that although video may be sufficient to observe the level of interactions needed to score an average level of quality, live observations may better capture the highest quality interactions. Negative practices that may result in a classroom scoring at the lowest level of quality may also be easier to observe live. Alternatively, video observations may simply have a more restricted range, if observers are less likely to provide scores at the extremes, acknowledging that they may have imperfect information.

For the CLASS Instructional Support domain, live scores were higher, on average, in monolingual English observations, but video scores were higher in multilingual (English-Spanish) observations. Although all CLASS observers in multilingual classrooms were fluent in both English and Spanish, the cognitive demands of language-switching during observations may have presented additional challenges. The enhanced audio of the microphone recordings, however, may have facilitated observers' ability to process language switches and consequently identify more high-quality interactions. This may be particularly relevant for the Instructional Support domain, given that research indicates Spanish-English bilingual teachers tend to use Spanish more frequently for informal interactions and English more frequently for instruction, suggesting teachers may switch languages when engaging in the instructionally supportive practices (Franco et al., 2019; Sawyer et al., 2018).

Analyses of how live and video scores varied differentially by video and microphone challenges highlight how, for CLASS in particular, differences between live and video scores may be partially explained by limitations of current recording technology. If video observers are unable to see or hear all high-quality practices due to suboptimal camera placement, challenges with primary marker tracking, or secondary microphone issues, they may not assign scores as high as live observers do. The one exception is the ECERS-3 Learning Activities subscale, for which scores were actually more similar in classrooms with secondary microphone issues—specifically, minor issues that coders noted were unlikely to affect coding. Thus, these issues may be correlated with other factors that made these classrooms score higher on video. For example, observations with minor secondary microphone issues may be ones in which the assistant teacher wearing the secondary microphone was highly active and, as a result, occasionally had microphone issues. Overall, these findings suggest care may be needed to use video technology efficiently and effectively to avoid disadvantaging classrooms receiving video observations, although high-quality audio recordings may counterbalance some of this disadvantage.

The appropriateness of video coding for different uses of observation scores (RQ 2c)

In addition to comparing scores across live and video, the study team conducted analyses to delve deeper into what these differences mean for each of the key use cases for classroom observation: coaching, accountability, and research.

Video coding appears appropriate to inform teacher coaching for CLASS but not ECERS-3.

To understand the implications of using video observations for coaching, we explored whether a coach would make the same decision about a teacher's high-need areas for coaching when observations were conducted live and on video. Although coaches typically use a range of factors to determine the focus of their work with a teacher, for the purposes of these analyses we sought to determine the extent to which live and video scores identified the same components of need within each domain or subscale—or coaching focus—by identifying the dimension or item that scored the lowest within each domain or subscale.

For CLASS, the most common coaching focus (i.e., lowest scoring dimension) within each domain was generally consistent across live and video observations. For Emotional Support and Classroom Organization, the lowest scoring dimension was the same in 93% and 84% of observations, respectively. Instructional Support, however, was less consistent, with 69% of observations having the same lowest scoring dimension across live and video.

For ECERS-3, the most common coaching focus (i.e., lowest scoring item) within a subscale was more variable, with alignment between coaching focus on live and video observations ranging from 29% to 44% across ECERS-3 subscales. Fewer than a third of observations were aligned for Interaction (22%), Program Structure (29%), and Space and Furnishings (29%). Alignment was higher for Language and Literacy (42%), Learning Activities (42%), and Personal Care Routines (44%), but still fewer than half were aligned.

Video scoring of CLASS observations for accountability purposes may produce different ratings for classrooms at the low and high ends of the quality distribution.

To understand the implications of using video observations for accountability, we explored whether live and video scores would result in a classroom being assigned to the same level of quality for Head Start monitoring (CLASS) or QRIS star-level ratings (ECERS-3). For both CLASS and ECERS-3, main effect models found no significant differences in the likelihood that an observation met the Head Start or QRIS star-level thresholds based on whether a live or video score was used (see Appendix F, Exhibit F13; Appendix G, Exhibit G18).

However, in models that include interactions by the live score, differences emerge for CLASS (Exhibit F14) but not ECERS-3 (Exhibit G19). Specifically, high-scoring CLASS observations were less likely to reach quality thresholds with video scores compared to live scores. On the other hand, low-scoring CLASS observations were more likely to reach quality thresholds if scored on video compared to live.

Video scoring appears appropriate for use in research, although ECERS-3 may be best used as a total score rather than as subscales.

Collecting consistent and reliable information is critical for education researchers who use classroom observation data. To delve deeper into this use case, we examined the proportion of variance explained by conducting an observation live or on video and compared it to a typical source of variance in classroom observation data—variance explained by the observers.

For CLASS, we found that the proportion of within-observation variance in scores explained by the video indicator was small, with video explaining at most 4% additional variance beyond what is explained by observers (which ranges from 43% to 55%; Exhibit 12). For the Classroom Organization and Instructional Support domains, the video method added no additional variance explained. For ECERS-3 Interaction, Program Structure, and total scores, the findings are similar, with video explaining an additional 3%–4% of the variance beyond that explained by observers. The one exception is the Learning Activities subscale, in which the video method explained an additional 12% of the within-observation variance.

Exhibit 12. Variance in scores explained by observers and video

	Variance explained by:	
	Observers	**Observer + Video**
CLASS Domains		
Emotional Support	0.43	0.47
Classroom Organization	0.55	0.55
Instructional Support	0.53	0.53
ECERS-3 Subscales		
Learning Activities	0.08	0.20
Interaction	0.25	0.28
Program Structure	0.02	0.07
ECERS-3 Total	0.12	0.15

Note. *Variance reflects adjusted R-squared values from within-observation fixed effect regression models predicting scores from observer indicators and the video indicator, respectively. Results are robust to sensitivity tests that examined the residual variance of random effects multilevel models with observations nested in classrooms.*

Summary: Live and video scores do not appear to vary within an observation in ways that are meaningful for research use cases, but CLASS scores vary in meaningful ways for accountability use cases, as do ECERS-3 scores for coaching use cases.

Coaching focus areas identified as the lowest scoring dimension within a domain were mostly consistent across live and video CLASS scores. However, the coaching focus varied widely across live and video ECERS-3 scores, suggesting that coaches may reach different conclusions about where to target teacher support depending on whether they conducted ECERS-3 live or on video.

With respect to accountability, live and video ECERS-3 scores did not vary with respect to whether a classroom met commonly used quality thresholds, suggesting that ECERS-3 can be used over video in accountability contexts. However, whether a CLASS observation meets quality thresholds did vary across live and video scores, particularly for observations with the highest and lowest scores. For a high-scoring observation, live scores were more likely to meet the Head Start quality thresholds; however, video scores were more likely to meet thresholds for low-scoring observations. Given the funding implications of not meeting the Head Start competitive threshold, the differences across live and video scores raise caution for using CLASS across live and video observations in accountability use cases.

Finally, with respect to research uses, video scores did not introduce much more variance than was explained by the observer, a commonly accepted source of variance in classroom observation data. The one exception was the ECERS-3 Learning Activities subscale, for which video accounted for more variance than did the observer. This finding, along with the coding challenges detailed in Appendix B and insufficient internal consistency for several ECERS-3 subscales, suggests using only the ECERS-3 total, rather than subscales, when conducting the ECERS-3 video observations for research.

Teachers' Perceptions of Video Observations (RQ 3)

The study team surveyed lead teachers and interviewed teachers and program leaders to learn more about their observation experiences and to gather their ideas about conducting video observations. Teachers described several ways that the use of video uniquely impacts their observation experiences, including some perceived challenges, and they also shared strategies for addressing these concerns. However, most teachers envisioned ways that video observations could support their professional development, especially when observations were perceived as accurate and when they engaged teachers as partners.

Teachers' experiences with video observations (RQ 3a)

Regardless of format, many pre-K teachers find classroom observations stressful (Lasagabaster & Sierra, 2011). For a few of the teachers in this study, this stress came from the disruptions inherent in having an unfamiliar adult in their classroom. As one teacher noted, "When there are a lot of grown-ups in the classroom, it kind of [changes] the dynamic of the classroom … [the kids] behave differently when there are other adults in the classroom. So they're curious, they are nervous, so are we." Another teacher reported feeling additional pressure when observers join her classroom, stating that "it's just funky, a little uncomfortable … because you want to do well. You want to show how great of a teacher you are, and then somebody comes in. It feels like pressure."

Pre-K teachers also sometimes express concerns that that a single observation may not be enough to capture a complete picture of their day-to-day teaching (Lasagabaster & Sierra, 2011). One participating teacher commented, "I feel like if somebody is not coming in all the time on a regular basis, you may not be able to observe everything that happens. So I think sometimes you're in observations, unless you're doing it consistently, little things can be lost from it." This can be stressful when scores are subsequently used for accountability measures of classroom and program quality.

Most pre-K teachers were comfortable with video observations, although some expressed concerns.

Survey data and teacher testimonials indicate that participating educators generally felt comfortable with classroom observations using video. Results of the teacher survey indicate that only 11% of teachers disagreed or strongly disagreed they were comfortable being observed via video. Moreover, only 17% agreed or strongly agreed that they find being observed via video stressful. Speaking about her experience participating in video observations, one teacher explained, "We were comfortable. We're a group of teachers that know what we're doing, and that's one of the most important parts." Another teacher echoed these sentiments:

> **"[The video observation] felt like it was really unobtrusive, very low stress. It was the easiest, most comfortable thing for me."**

At the same time, many teachers in our study expressed reservations about the use of video observations. Only 65% of teachers agreed or strongly agreed that video observations are generally accurate, compared to 76% who agreed or strongly agreed that live observations are generally accurate. Additionally, 82% of teachers agreed or strongly agreed that video observations are useful, compared to 98% who agreed that live observations were useful. Given their understanding of the current state of video observations, more teachers preferred live observations (50%) than video observations (22%), although some (28%) were undecided.

For some participating pre-K teachers, observation concerns were heightened with the presence of a video camera. As one teacher noted, "Sometimes it's not even necessarily that a teacher is nervous, so to speak, of their skill on camera. Sometimes people are just not comfortable on camera in general."

These concerns may be especially acute for teachers in programs where video observations are not part of the established culture. Many of the teachers surveyed (53%) had not experienced video observations prior to their participation in this study. For teachers unfamiliar with regular use of video in the classroom, the sudden introduction of video for observations could induce anxiety. As one teacher shared,

> **"We've never been recorded with a camera or in this case the iPad. And so, it was a little different. It was more on the lines of funny, superficial, how's my hair, how's my clothes, what do I look like?"**

Even in programs with pervasive video use where teachers feel comfortable being recorded, observation-specific recordings can evoke a distinct reaction. A teacher working in a program with security cameras in classrooms explained, "The cameras that we have in our classroom, we know it's doing 24/7 – it's working and sometimes you don't even think about it ... but when you see the camera in front of you, okay, then you are ... always aware of it in front of you."

In some cases, concerns about video observations were rooted in teachers' linguistic, cultural, or religious practices. One program director explained, "We have some of the teachers who are Orthodox, religious. ... Sometimes they sing, sometimes they ... sit in the positions that they don't want other people to see it." When discussing her experience wearing microphones during video observations, one bilingual teacher commented, "We were very quiet. ... English [is] our second language, so we were kind of, 'Okay, am I speaking right?' Even if I'm reading for the kids, I was like, okay, I was very focused. 'Am I reading it right? Am I using the right sentences? Am I using the right grammar?'"

Nevertheless, many teachers found ways to adapt. As one teacher noted, "I try to forget the camera is there and just get into my groove." Another participant commented, "I think that everybody feels comfortable, and honestly, they remember that we have cameras when they need to see the cameras, but other than that, it is very natural to them."

Others found that hesitations about using video for observations dissipated quickly. One teacher shared, "It's a little bit uncomfortable at first. I'm sure I got kind of used to it a little bit. Sometimes, you tend to forget that they're there." Another teacher admitted, "In the beginning, it's hard, like, for me, I don't like to see my face in the video, my voice, I say, 'Oh.' And then, after that, you get used to, you don't care. It's what it is."

Many participating teachers worried that video distracts students, but some have found ways to limit the extent of the distraction. Some teachers also expressed concerns that the presence of video equipment heightened levels of classroom disruption associated with observations. Almost three quarters (72%) of teachers we surveyed worried about the impact of video observations on classroom disruption. In our study, disruption was primarily manifested through student distraction, either from curiosity or excitement. One teacher humorously recounted, "I think one of the challenges, we had some of our students jump in [front of] the camera to try to look at themselves. I thought that was hilarious."

Another teacher noted the impact on students' focus and behavior, sharing that "I saw the difference for the kids. On that day, my kids are going there and saying even they know that we have a camera … it's something new for them. Oh, 'Miss [redacted], what's this? Why we have it here? Why it's doing?' You know, it's just like … I see a lot of distraction. I have some curious kids."

However, teachers were divided on how much these distractions disrupted the normal flow of classroom activities, and several teachers shared strategies to overcome these hiccups. Only 19% of surveyed teachers agreed or strongly agreed that the video equipment was distracting to their students. Many reported that any initial disruption was short-lived. As one teacher put it, "It was definitely an adjustment … I think kids are super adaptable, though, so after the initial shock or excitement of it, they kind of just move as normal." Others found ways to prepare their students, with one teacher explaining,

> **"I was a little bit nervous about having the technology in my class … but we did have some conversations about how to be safe around it, no running, and they ran with this idea, and, all of a sudden, came up with a million things not to do. They're like, 'Don't kick it. Don't bite it. Don't throw things at it. Don't smash it.' I think it helped."**

A few teachers even found video observations less distracting than live observations. One teacher described the benefit of having fewer adults in the classroom for video observations, noting that "sometimes when there are too many adults in the room, it can be overwhelming for teachers and the students. But if we needed an observation, and they just set up a camera and left, I think that could be an alternative." Another emphasized that "the kids love [video] more. Then, it's not two or three strangers just sitting down in the classroom writing stuff, writing things down." Others discussed the benefit of being able to ignore video equipment more easily than they might ignore an observer. Speaking to this experience, one teacher commented,

> **"I'd rather do video than live because the pressure to keep looking at the person – what's her face or what's his face looking like? Am I doing this right? It's the constant pressure … they're watching you like a hawk. Whereas recording … my camera's watching me, but I can still be my regular self."**

Data privacy was a concern for some teachers. Video recordings of students and classrooms raise important questions about who can access videos, where and how videos are stored, and for what purpose videos are recorded. One teacher shared, "So not everybody, like I said, is comfortable with being recorded and what's going to happen with the usage of this, or who will eventually get their hands on it." These concerns were reflected in our survey, in which many of the teachers surveyed (56%) reported being somewhat or very concerned that videos might be accessed by an unauthorized person.

Teachers also shared worries about student privacy that they heard from parents. One teacher reported taking additional steps to accommodate parents' concerns for participation in our study: "We have a list of the kids that the parent, they don't want them to be video-recorded or anything. … So [on days when study observations took place] they had to put them in different classrooms, because the parent, they … don't want them to be photographed."

Some teachers underscore the need to select flexible instrumentation that minimizes disruption. The success of video observations is dependent on the quality of the video captured. Without specific guidelines, teachers experience a lot of trial and error in determining where and how to place technology. As one teacher explained, "Even if I get a really wide shot of the classroom, say the one that I was talking about where I duct-taped my phone to the corner of the classroom … you end up not hearing things that are happening far away in the footage." This can be a time-consuming and frustrating ordeal for teachers. As one teacher explained, "That's usually what we're struggling with in the classroom with video stuff. I would say I've spent way more time on trying to figure out the camera stuff than I should be."

Teachers emphasized the importance of considering the placement and features of technology chosen for video observations. Highlighting the importance of camera placement, one teacher shared,

> **"I did appreciate the Swivl on this camera just because as an early childhood teacher, we do move around a lot. We move around the room about, and I do feel like when the observation where they had just the pod, if the observer didn't get up and move the camera, sometimes they miss some really great stuff."**

Other teachers underscored the need to keep technology out of the way of students. One remarked, "I would say I would prefer a setup the way your observers had, where it's on a tripod, and it can rotate, and not in the way of children. Because when I use my laptop, it's at their level, and they're all in the camera."

Teachers also reflected on the benefit of selecting technology that is familiar for students to avoid heightened levels of distraction. Two teachers explained that their students were surprisingly disinterested by the use of iPads to record observations for our study. One commented, "I was very surprised that they did not acknowledge the, I think it was the iPad that was used on the observation, at all. I was so proud of them. I was like, yes, my friends are focused on me." The other teacher reflected,

> **"If it was a regular camera as if one of the news anchors or the big industrial camera, then it might have [been distracting], but it was an iPad and because it was an iPad, my children play with iPad, so they pretty much knew … These kids are really smart. And so it wasn't, 'Ooh, this is new' or anything like that. It was just, 'Okay. So she has an iPad. We have an iPad, too.'"**

These experiences highlight instrumentation selection as an opportunity for teachers and programs to consider how best to meet their needs for observations while balancing the role that video might have in impacting classroom routine.

Summary: Pre-K teachers' experiences with video observations reveal both challenges and opportunities for implementation in ECE settings. While most teachers reported feeling comfortable with video observations, some expressed concerns about the accuracy, usefulness, and potential disruption of video observations when compared to live observations.

Some teachers found video observations less stressful than live observations, while others experienced heightened anxiety from being on camera. Cultural and linguistic factors may also play a role in teachers' comfort, with some teachers worried about how and by whom their practices might be perceived on video. Concerns about student distraction were common, although many teachers found that any disruption was short-lived, especially when they enacted strategies to prepare students and minimize distractions. Some teachers even found video observations to be less disruptive than having an additional adult in the classroom.

The type and placement of video equipment greatly impacted teachers' experiences. Many emphasized the importance of using flexible, unobtrusive technology that can capture classroom activities effectively without disrupting the learning environment. Similarly, data privacy also emerged as a significant concern, with many teachers worried about unauthorized access to recordings. This highlights the need for clear protocols and communication about video usage.

Despite reservations, many teachers recognized the potential of video observations for supporting their professional development, especially when implemented thoughtfully and collaboratively. These findings suggest that successful implementation of video observations may require gradual introduction, flexible protocols to accommodate diverse practices, careful selection of recording equipment, clear data privacy guidelines, and strategies to prepare both teachers and students for the presence of video technology in the classroom.

Teachers' ideas about potential benefits of video observations (RQ 3b)

Most participating teachers identified potential benefits to video observations. Even though video impacted teacher experiences with observations, most teachers acknowledge benefits to using video, made possible by the capabilities afforded by video instrumentation. Teachers envisioned themselves using video observations as an opportunity for reflecting on their teaching, understanding classroom dynamics, and sharing exemplars with others.

Most pre-K teachers want to use video to reflect on their practice and identify concrete steps for professional growth

Most teachers acknowledged the potential benefit of using video observations to reflect on their teaching practices. Regarding how video can aid in reflection, one teacher explained, “It is helpful because I think I can see some things that I missed when I was in the moment. It’s helpful to think back, and reflect, and process things.” The reflective nature of video observations enables teachers to revisit their classroom experiences and gain new insights.

Some teachers also discussed how video observations can make the results feel more personal and relevant by grounding their scores in specific video segments. One teacher noted, “I personally think they are very much so helpful. I think that it allows teachers to be in that moment, even after the fact, and you can visually see what ... If it’s being used to be rated against something, you can visually see like, okay, this is what they mean when they say this.” Another teacher explained,

> **“...[M]aybe if you watch yourself back, and have the person explain, it would be better versus just seeing your scores and then reading a bunch of paragraphs as to what you could have done when you probably should be watching yourself, and your coach or whoever is providing you feedback, stop and pause video thing type of conversation to ... help myself get better. *Show me what I’m doing, show me, me.”***

Unlike traditional methods that provide only written feedback or numerical scores, video recordings allow teachers to more easily connect feedback to specific moments and actions, making suggestions for improvement more tangible and actionable.

Many teachers think video can provide a more comprehensive view of classroom dynamics, but some caution against overinterpretation.

Some teachers remarked that video recordings offer a window into student behavior and engagement, revealing insights that might otherwise go unnoticed in the busy classroom environment. This aspect is particularly valuable in ECE settings, where children may express themselves in subtle ways that are easily missed during live observation. One teacher highlighted this benefit:

> **“I think that there are some kids that fly under the radar, and maybe those video recordings would be great to capture that. Those children that are a little bit on the quiet side, what are they working on? If I’m looking back at those recordings, what kind conversations can I hear that I wasn’t hearing because I wasn’t standing next to them?”**

By reviewing these recordings, teachers can gain a more comprehensive understanding of their students’ needs, interests, and progress, allowing for more targeted and effective instruction.

Some teachers communicated that they believed video observations may be more accurate or objective than live observations. Teachers recognized that video recordings can capture a more comprehensive view of the classroom than a single observer can. As one teacher noted, “I think it might be beneficial if, say for example, we have a group of kids in the block area, but the teachers are at different areas … [it] might be a good opportunity to pick up on a specific area of the classroom or something of that nature." Another teacher summarized this benefit:

> **"It'd great to be able to see the whole picture. Sometimes when you're in the moment, you miss things. But with a video, you can really see how everything comes together – the interactions, the transitions, the overall flow of the lesson. *It's like getting a bird's eye view of your own teaching.*"**

Teachers also recognized the capabilities afforded by video observations that are limited for live observations. One teacher pointed out, "How fast can you [the observer] write, really? And jot everything down that you're observing." Video recordings address this concern by allowing for multiple viewings and more detailed analysis. Another teacher emphasized this advantage: "So that ability to rewind … and maybe see a little bit more in the classroom that might not get captured when you're only able to focus on so many things at once in person."

Some teachers envision using video for training and learning from others.

Both teachers and program leaders also saw potential for video observations to serve as training tools. Video recordings provide concrete, real-world examples of teaching practices that can be analyzed, discussed, and learned from in a controlled environment. Commenting on this use case, one program director shared, "Well, I think it would be good for trainings. I mean, we've done video recordings before. But again, you can't have the teachers inside the classrooms doing it, constantly taking pictures. Teacher's interaction, children interactions, what projects they're working on. But I think videos would really help the teachers to see themselves." One teacher anticipated benefits through instructional coaching:

> **"I think it's helpful as being a mentor teacher, that we can stop and talk about the video, and pinpoint certain things. Versus in the moment, we might not be able to always debriefing and explain the why of teaching. So being able to have that video gives us just an opportunity to really expand the learning experience."**

Another teacher expressed excitement about being able to reference the practices of other teachers. This teacher shared, "And if I want to ask somebody else and I want to watch what other teachers are doing, I think those recordings are good mentor for me to learn something from somebody or just as a reference for me to go back and look what I did not do and I should have done in a better way."

Some teachers expressed concern that video observations may miss nuances to interactions or classroom atmosphere that a live observer might catch.

Many (69%) of the teachers we surveyed expressed concerns that video recordings may be inaccurate or incomplete. As one teacher reflected, "I see the social emotions, because as human being, you can see all those expressions. You can see it, you are there. But with computers, I don't think they can see that." Another teacher reflected, "[The video observer] cannot catch all the feeling, the social emotions that human being can see. That's one of the things that I think that video recording can miss." One teacher, who is a certified CLASS observer, also shared words of caution for over-leveraging the affordances of video observations: "When I do my reliability test, so what we do is we watch 20-minute videos, and on those videos we scored the participants. And on those scoring, that's our test, do the reliability test. So those videos we also watch only once. If I'm watching those videos again and again, you keep finding some negativity in those."

Summary: Teachers' perspectives on video observations reveal both potential benefits and limitations. Many teachers believe video can provide a more comprehensive view of classroom dynamics, capturing subtle student behaviors and interactions that might be missed during live observations.

Teachers appreciate the ability to review recordings multiple times, allowing for more detailed analysis and the opportunity to observe areas of the classroom that a single observer might miss. As a result, some teachers perceive video observations as potentially more accurate or objective than live observations, offering a "bird's eye view" of the classroom.

Some teachers also regard video observations as a valuable tool for professional development. Teachers envision using recordings for training, mentoring, and self-reflection. They view the ability to pause, rewind, and discuss specific moments in a controlled environment as particularly beneficial for instructional coaching and learning from peers. Similarly, video has the potential to make observations scores more meaningful by grounding them in specific video segments that allow teachers to view their practices alongside explanations of scoring, which can enhance teachers' observation experiences.

However, some teachers expressed concerns about the limitations of video observations. They worry that recordings may miss nuances in interactions or classroom atmosphere that live observer might catch, particularly regarding social-emotional aspects. Some caution against over-interpretation of video footage, noting that repeated viewings could lead to finding unwarranted negativity.

Taken together, these findings suggest that teachers see the benefit of engaging with video observation for continued growth and learning, but not all teachers agree that video and live observations are equal.

Costs of Video and Live Observations (RQ 4)

Conducting classroom observations is a resource-intensive endeavor. At the time of this writing, participating in CLASS or ECERS training costs between $3,000 and $6,200 per observer. Conducting the observations requires time from trained staff, including time observers spend traveling to and from pre-K programs. Video observations also require additional initial costs in the form of specialized equipment to record and securely store the videos. Video observations may also necessitate that additional staff to travel to programs, set up video cameras, record the observation, and transfer the videos to the trained observers for scoring.

In this section, we estimate the costs for conducting classroom observations over a 3-year period for a hypothetical early childhood system that includes 300 classrooms located in 100 pre-K programs. Within this hypothetical system, we estimate the costs for conducting live and video observations separately for the purposes of accountability and for the purposes of supporting coaching. Within each of these scenarios, we compare the costs to conduct the observations live versus via video recording. It is important to note that the costs for conducting live and video-based observations are highly variable and that the assumptions applied in our analysis may not be appropriate for each context.

In the **accountability** scenario, we assumed that each pre-K classroom is observed once a year and that the live observations are conducted by trained observers who are not employed by the observed schools and programs. For accountability-focused video observations, we assumed that the recordings would be made by audio-video technicians who are not employed by the observed pre-K programs. These audio-video technicians would transport the recording equipment with them to the pre-K programs.

For the **coaching** scenario, we assumed that each pre-K classroom is observed five times a year and that the live observations would be conducted by trained observers who are not employed by the pre-K program. We assumed that video recordings for purposes of coaching could be made by each pre-K program's staff, which would necessitate that each pre-K program have its own set of recording equipment but would minimize setup and travel needs for video recordings.

We summarize the assumptions we use in these estimations in Exhibit 13 and provide additional details in Appendix I.

Exhibit 13. Assumptions for estimating costs for live and video observations

	Accountability		Coaching	
	Live	Video	Live	Video
ECE system structure				
Number of classrooms	300	300	300	300
Number of programs	100	100	100	100
Observations per classroom	1 per year	1 per year	5 per year	5 per year
Cost projections	3 years	3 years	3 years	3 years
Labor cost components				
Number of observers	15	15	50	50
Observation data that can be collected by 1 person in 1 day	1	3 (recorded simultaneously)	1	Videos recorded by program staff
Travel required to conduct observations	90% local, 10% distant	100% local	90% local, 10% distant	No travel required
Staff hours required to collect and score CLASS observations	3.5 hours for trained coder (incl. 1 hour travel/ setup)	3 hours for video recorder, 2.5 hours for trained coder	3.5 hours for trained coder (incl. travel/ setup)	2.5 hours for trained coder
Staff hours required to collect and score ECERS-3 observations	4.5 hours for trained coder (incl. 1 hour travel/ setup)	3 hours for video recorder, 3.5 hours for trained coder	4.5 hours for trained coder (incl. 1 hour travel/setup)	3.5 hours for trained coder
Total number of trips a staff person(s) must make to a program for the purposes of observation	300	105	1,500	0
Equipment cost components				
Recording equipment	N/A	45 Swivl sets (3 per technician)	N/A	100 Swivl sets (1 per site)
Equipment replacement	N/A	20% annually after Year 1	N/A	20% annually after Year 1

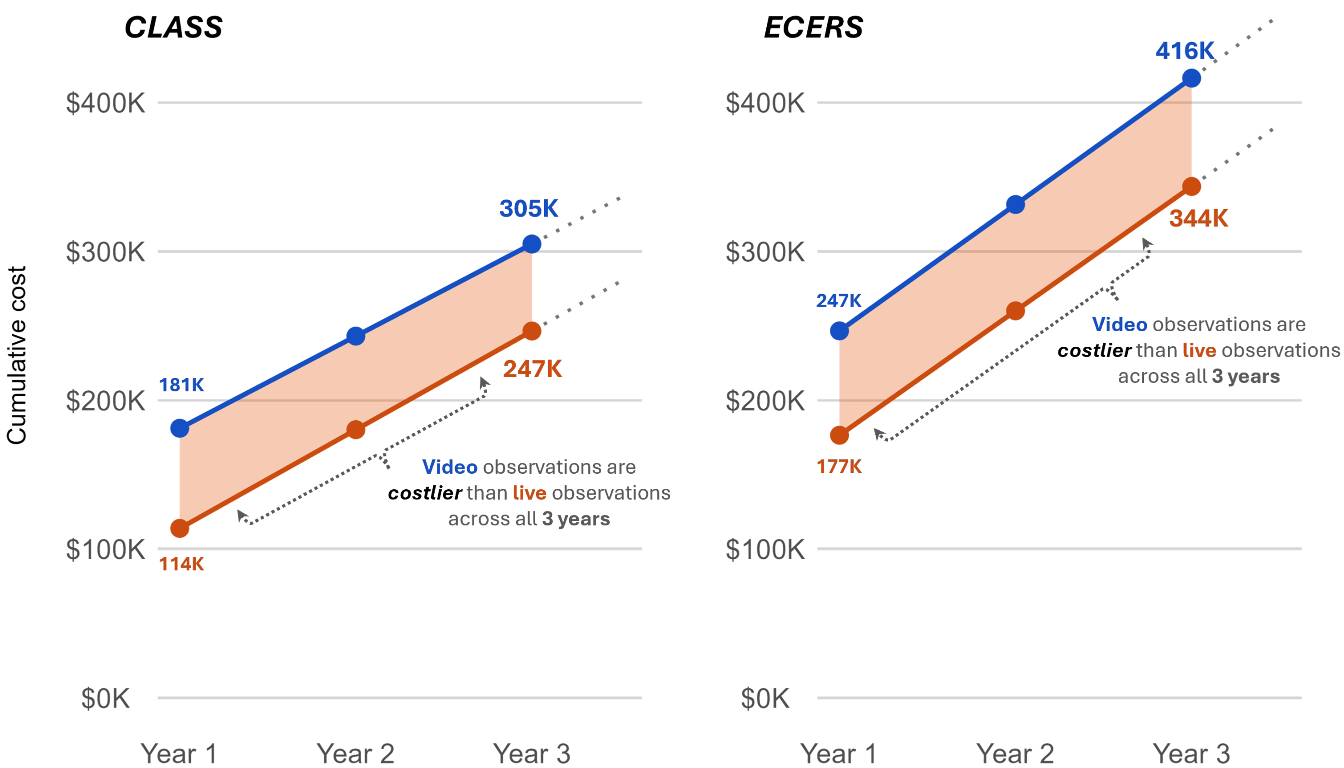
When conducting observations for the purposes of accountability, video observations are <u>more expensive</u> than live observations.

We estimated the costs of live and video observations assuming 15 observers conducting 300 observations in 300 classrooms across 100 pre-K programs (i.e., three classrooms per program). We assumed travel would consist of 90% local travel and 10% distant travel. We defined local travel as being less than 20 miles and not requiring a per diem for meals or a hotel for overnight stay. We defined distant travel as being at least 50 miles and requiring a per diem for meals and an overnight stay at a hotel. One of the potential advantages of video observations is that the recording equipment can be reused in subsequent years. Nevertheless, equipment does break or get lost. Therefore, we estimated that 20% of recording equipment would need to be replaced because of loss or damage each year. We projected costs for both the live and video observations out to 3 years. Detailed estimates are provided in Appendix I, Exhibits I1–I4.

Video observations require a large investment in the initial purchase and periodic replacement of recording equipment. At the same time, video observations offer potential savings in travel and staff labor costs. These savings come from two sources. First, the labor costs associated with the audio-video technician who travels to the pre-K programs to conduct the recordings are lower than the labor costs of the trained observer: a $51,640.00 annual median salary ($24.83 median hourly wage) for an audio-video technician versus a $74,620.00 annual median salary ($35.87 median hourly wage) for an instructional coordinator. The difference in staff salaries reduces the costs associated with getting staff to and from programs. Second, the audio-video technician would record, on average, three classrooms simultaneously each time they travel to a program. This would reduce the total number of trips to programs required for conducting observations.

Despite the lower operational costs in subsequent years, the high initial equipment investment for video observations makes them more expensive than live observations over a 3-year period. For example, in the first year of the 3-year estimate, the costs of conducting all 300 CLASS observations using live observers was $114,000 (average cost of $380 per observation), compared with $181,000 for video observations (average cost of $611 per observation). With the majority of equipment costs paid for in the first year, video observations become less expensive than live observations in the subsequent years (average cost of $221 per live observation vs. average cost of $214 per video observation). However, this moderate reduction in costs per observation after Year 1 is not sufficient to mitigate the initial equipment costs. Similarly, ECERS-3 video observations are substantially more expensive in the first year: total Year 1 cost of $247,000 for video versus $177,000 for live observations. Thus, despite video observations being moderately less expensive in the subsequent years, the savings are not sufficient to account for the equipment costs during the 3-year period (Exhibit 14).

Exhibit 14. Comparison of live and video observation costs in an accountability scenario



When conducting observations for the purposes of coaching, video observations are <u>less expensive</u> than live observations.

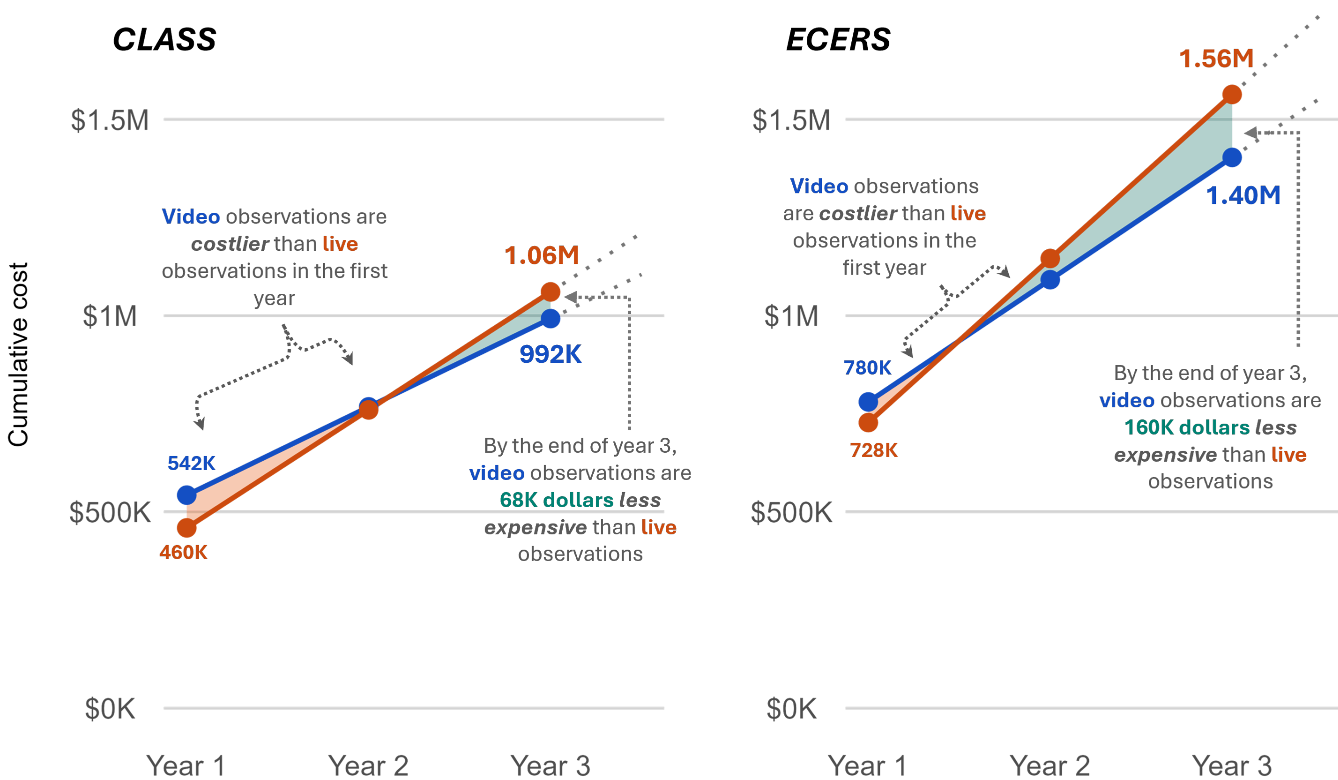
For estimating the costs of coaching, we hypothesized a scenario in which 50 observers conducted five observations for each of the 300 classrooms across the 100 pre-K programs. Live observations would be conducted by trained observers who travel to classrooms. Between travel time, setup, and the observation, live observers would only be able to conduct one CLASS or one ECERS-3 observation per day, which would require 1,500 trips. For the video observations, we assumed it would be most efficient for each pre-K program to have its own set of recording equipment and for its own staff to make the recordings. As with the accountability scenario, the video method would have a large initial investment in recording equipment but provide substantial savings in labor and travel costs. As for the accountability scenario, we assumed 90% local travel and 10% distant travel for coaching, and we projected costs for both live and video observations out to 3 years. Detailed estimates are in Appendix I, Exhibits I5–I8.

Although the cost components are the same across the accountability and coaching scenarios, there are some important differences that lead to significant long-term savings when using video to support coaching-related observations. The initial investment in recording equipment for coaching-related observations is substantially larger than for accountability-related observations. In the coaching scenario, each pre-K program would receive its own recording equipment. This would permit program staff, rather than audio-video technicians, to record the observations. As a result, the substantial reduction in labor and travel costs after Year 1 is sufficient to offset the initial equipment investment within the 3-year period.

For CLASS, the first-year costs for coaching-related observations are again lower for live observations (total Year 1 cost of $459,000 or $282 per observation) than for video observations (total Year 1 cost of $543,000 or $361 per observation), a result of the Year 1 equipment purchase for video observations. The cost dynamics shift dramatically in subsequent years, with video observations becoming significantly less expensive. The estimated 3-year total cost of live observations in the coaching scenario is $1.2 million ($236 per observation), compared with $992,000 ($22 per observation 1) for video observations (Exhibit 15).

Estimates for the costs of ECERS-3 observations for coaching purposes again follow a similar trend to CLASS. The first-year costs show video observations as more expensive: $728,000 for live and $780,000 for video observations. The estimated cost-saving potential of video observations is realized in the second year, with an estimated total reduction in costs of $168,000 for video over live over the 3-year period (see Exhibit 15).

Exhibit 15. Comparison of live and video observation costs in a coaching scenario



Summary: Video observations offer a substantial costs savings when conducting observations for coaching purposes, but not when conducting observations for accountability purposes.

For accountability purposes, in which classrooms are observed once a year, video observations are more expensive than live observations. Although the relative costs of video observations are lower than the costs of live observations after the first year, the savings are not sufficient to account for the substantial initial costs. The break-even point for video observations in the accountability scenario is estimated at approximately 15 years, indicating limited cost-effectiveness in the short to medium term.

By contrast, for coaching purposes, in which classrooms are observed multiple times a year, video observations represent a cost-effective approach. While first-year costs for video observations remain higher because of initial equipment purchases, by the second-year video observations generate considerable savings relative to live observations, with substantial ongoing savings thereafter.

Key factors influencing these cost differences include the initial investment in recording equipment, travel costs for live observations, equipment reusability, and observation time requirements. Video observations allow for simultaneous recording of multiple classrooms and reduce travel time for trained observers, contributing to long-term savings, particularly in coaching scenarios where multiple observations are conducted each year.

These findings suggest that ECE systems should carefully consider their specific needs and observation frequencies when deciding between video and live observations. While video observations may not be cost-effective for scenarios in which a classroom is observed only once per year, they offer significant financial advantages for scenarios that require multiple annual observations per classroom. The decision to implement video observations should weigh initial costs against potential long-term savings, accounting for the intended use case and frequency of observations.

Again, these findings are based on a particular set of assumptions. We encourage ECE systems leaders to model return on investment using cost information relevant to their context prior to making decisions about the use of video-based observations for coaching or accountability purposes.

Conclusion

The results of this study will provide policymakers, early childhood teachers, and researchers with information about the trade-offs for live and video classroom observations. Results suggest that CLASS and ECERS-3 observations conducted on video meet the standards of reliability used to certify observers, although some ECERS-3 subscales did not have acceptable internal consistency. With respect to whether live and video scores vary within an observation, both CLASS and ECERS-3 had minimal score differences for classrooms with typical ratings for quality, but live scores were higher than video scores in the highest quality classrooms and video scores were higher than live scores for lowest quality classrooms. This finding may indicate that observers are less likely to assign scores at the extremes over video, either because highly positive or negative practices may be difficult to observe on video or because observers may be cautious to assign extreme scores when they are aware they have imperfect information.

The implications of these score differences depend on the purpose for the classroom observation. For coaching, comparison of lowest scoring dimensions (a proxy for the area a coach and teacher may choose to focus on) suggests consistency across live and video observations for CLASS, but not ECERS-3. For accountability, on the other hand, the likelihood of meeting policy-relevant thresholds of quality was consistent across live and video scores for ECERS-3, but not CLASS. For research, the proportion of variance explained by video is very small compared to the variance explained by observers for CLASS. For ECERS-3, video explains more variance than observers do for the Learning Activities subscale. This finding, coupled with the low internal consistency across several ECERS-3 subscales, suggests that using the ECERS-3 total score, rather than subscale scores, may be most appropriate for video observations in research.

For programs interested in implementing video observations, data suggest that introducing video for observations changes teachers' experiences in several unique ways, but most teachers view these changes to be surmountable. Teachers in this study reported using strategies to prepare students for the presence of video equipment and minimize distraction, become comfortable teaching in front of the camera, and communicate with parents to address student privacy concerns. Additionally, both teacher and program leaders envision a myriad of benefits to engaging with video observations—from having opportunities for reflection, to bringing observation scores to life, to enhancing understandings of classroom dynamics and even learning from the practices of others. Across both video and live formats, teachers reported similar levels of stress when engaging with observations, even though based on their current experiences with video observations, teachers rated video observations slightly lower than live observations in accuracy and usefulness. Overall, most teachers expressed an openness to engaging with video observations, although there are opportunities to finetune instrumentation and other supports to ensure a smooth observation experience.

Finally, we estimate that the costs of video observations are higher when the observations are conducted in a classroom just once per year, such as those conducted for accountability purposes, but lower when the classrooms are observed multiple times per year, such as to support teacher coaching. The long-term cost-effectiveness of video observations for coaching is primarily attributed to the reduction in travel costs and the ability to reuse recording equipment.

Taken together, these findings suggest that video is a viable method of conducting classroom observations, with reliability comparable to live observations and an overall openness among teachers. However, decisions about whether to use video for observations should consider the specific use case for the observations and the observation tool used. Findings generally support the use of video observations for coaching, with evidence that video has cost savings and that coaches are likely to make similar decisions about where to focus their attention when using CLASS, although not ECERS-3. Findings are less promising for the accountability use case, as cost estimates for video observations are higher than those for live observations, for the typical frequency with which observations are conducted for accountability. Additionally, live and video scores differ for the classrooms with the highest and lowest CLASS scores, suggesting that the use of video may differentially affect precisely those classrooms most affected by accountability observations.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bottiani, J. H., Duran, C. A., Pas, E. T., & Bradshaw, C. P. (2019). Teacher stress and burnout in urban middle schools: Associations with job demands, resources, and effective classroom practices. *Journal of school psychology, 77,* 36–51.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. 10.1191/1478088706qp063oa.

Bronfenbrenner, U., & Morris, P. A. (2007). The bioecological model of human development. In R. M. Lerner (Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed.). https://doi.org/10.1002/9780470147658.chpsy0114

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4rs program. *Journal of Educational Psychology, 102*(1), 153–167. https://doi.org/10.1037/a0018160

Build Initiative & Child Trends. (2024). *A catalog and comparison of quality initiatives* [Data system]. https://qualitycompendium.org/

Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives, 12*(1), 3–9. https://doi.org/10.1111/cdep.12260

Bureau of Labor Statistics. (2024). Educational instruction and library occupations. In *Occupational employment and wage statistics*. U.S. Department of Labor. Retrieved November 11, 2024, from https://www.bls.gov/oes/current/oes_nat.htm#25-0000

Casabianca, J. M., Lockwood, J. R., & McCaffrey, D. F. (2015). Trends in classroom observation scores. *Educational and Psychological Measurement, 75*(2), 311–337. https://doi.org/10.1177/0013164414539163

Clark, M., Max, J., James-Burdumy, S., Robles, S., McCullough, M., Burkander, P., & Malick, S. (2022). *Study of teacher coaching based on classroom videos: Impacts on student achievement and teachers' practices* (NCEE 2022-006r). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pubs/2022006/

Curby, T. W., Johnson, P., Mashburn, A. J., & Carlis, L. (2016). Live versus video observations: Comparing the reliability and validity of two methods of assessing classroom quality. *Journal of Psychoeducational Assessment, 34*(8), 765–781. https://doi.org/10.1177/0734282915627115

Curenton, S. M., Iruka, I. U., Humphries, M., Jensen, B., Durden, T., Rochester, S. E., Sims, J., Whittaker, J. V., & Kinzie, M. B. (2020). Validity for the Assessing Classroom Sociocultural Equity Scale (ACSES) in early childhood classrooms. *Early Education and Development, 31*(2), 284–303. https://doi.org/10.1080/10409289.2019.1611331

Curenton, S. M., Rochester, S. E., Sims, J., Ibekwe-Okafor, N., Iruka, I. U., García-Miranda, A. G., & Whittaker, J. (2022). Antiracism defined as equitable sociocultural interactions in prekindergarten: Classroom racial composition makes a difference. *Child Development, 93*(3), 681–698. https://doi.org/10.1111/cdev.13779

Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., & Nehler, C. G. (2018). Factor structure and validity of the Early Childhood Environment Rating Scale – Third Edition (ECERS-3). *Early Childhood Research Quarterly, 44*, 242–256. https://doi.org/10.1016/j.ecresq.2018.04.009

Farley, K. S., Piasta, S., Dogucu, M., & O'Connell, A. (2017). Assessing and predicting small-group literacy instruction in early childhood classrooms. *Early Education and Development, 28*(4), 488–505. https://doi.org/10.1080/10409289.2016.1250549

Farran, D. C., & Anthony, K. (2014). *Child observation in preschools (COP)*. Peabody Research Institute; Vanderbilt University.

Figueras-Daniel, A., & Li, Z. (2021). Evidence of support for dual language learners in a study of bilingual staffing patterns using the Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA). *Early Childhood Research Quarterly, 54,* 271–285. https://doi.org/10.1016/j.ecresq.2020.09.011

Franco, X., Bryant, D. M., Gillanders, C., Castro, D. C., Zepeda, M., & Willoughby, M. T. (2019). Examining linguistic interactions of dual language learners using the Language Interaction Snapshot (LISn). *Early Childhood Research Quarterly, 48*, 50–61. https://doi.org/10.1016/j.ecresq.2019.02.007

Freedson, M., Figueras-Daniel, A., Frede, E., Jung, K., & Sideris, J. (2011). The classroom assessment of supports for emergent bilingual acquisition. In *Dual language learners in the early childhood classroom* (pp. 233–258). Brookes Publishing.

Hamre, B. K. (2014). Teachers' daily interactions with children: An essential ingredient in effective early childhood programs. *Child development perspectives, 8*(4), 223-230.

Harms, T., Clifford, R., & Cryer, D. (2014). *Early Childhood Environment Rating Scale–Third Edition manual*. Teachers College Press.

Hestenes, L. L., Rucker, L., Wang, Y. C., Mims, S. U., Hestenes, S. E., & Cassidy, D. J. (2019). A comparison of the ECERS-R and ECERS-3: Different aspects of quality? *Early Education and Development, 30*(4), 496–510. https://doi.org/10.1080/10409289.2018.1559681

Justice, L. M., Jiang, H., Khan, K. S., & Dynia, J. M. (2017). Kindergarten readiness profiles of rural, Appalachian children from low-income households. *Journal of Applied Developmental Psychology, 50*, 1–14. https://doi.org/10.1016/j.appdev.2017.02.004

Kane, T. J., Blazer, D., Gehlbach, H., Greenberg, M., Quinn, D. M., & Thal, D. (2020). Can video technology improve teacher evaluations? An experimental study. *Education Finance and Policy, 15*(3), 39–427. https://doi.org/10.1162/edfp_a_00289

Lasagabaster, D., & Sierra, J. M. (2011). Classroom observation: desirable conditions established by teachers. *European Journal of Teacher Education, 34*(4), 449–463. https://doi.org/10.1080/02619768.2011.587113

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2018). *Economic evaluation in education:*

Cost-effectiveness and benefit-cost analysis (3rd ed.). Sage Publications.

Meek, S., Iruka, I. U., Soto-Boykin, X., Blevins, D., Alexander, B., Cardona, M., & Castro, D. (2021). *Equity is quality and quality is equity: Operationalizing equity in quality rating and improvement systems.* Children's Equity Project. https://cep.asu.edu/resources/Equity-is-Quality-and-Quality-is-Equity

Montes, G., Reynolds Weber, M., Infurna, C., Van Wagner, G., Zimmer, A., & Hightower, A. D. (2018). Factor structure of the ECERS-3 in an urban setting: An independent, brief report. *European Early Childhood Education Research Journal, 26*(6), 972–984. https://doi.org/10.1080/1350293X.2018.1533712

Office of Child Care. (2018). Standards and criteria. In *QRIS resource guide*. U.S. Department of Health and Human Resources, Administration for Children and Families. https://ecquality.acf.hhs.gov/resource-guide/standards-and-criteria

Office of Head Start. (2016). *Head Start program performance standards*. U.S. Department of Health and Human Services, Administration for Children and Families.

Office of Head Start. (2024a). *Designation Renewal System overview.* U.S. Department of Health and Human Services, Administration for Children and Families. https://headstart.gov/designation-renewal-system

Office of Head Start (2024b). *Use of Classroom Assessment Scoring System (CLASS®) in Head Start.* U.S. Department of Health and Human Services, Administration for Children and Families. https://headstart.gov/designation-renewal-system/article/use-classroom-assessment-scoring-system-class-head-start-programs

Pakarinen, E., Malmberg, L.-E., Poikkeus, A.-M., Siekkinen, M., & Lerkkanen, M.-K. (2023). Investigating applicability of ratings of indicators of the CLASS Pre-K instrument. *International Journal of Research & Method in Education, 46*(3), 231–247. https://doi.org/10.1080/1743727X.2022.2128741

Partika, A., Johnson, A. D., Phillips, D. A., Luk, G., Dericks, A., & The Tulsa SEED Study Team. (2021). Dual language supports for dual language learners? Exploring preschool classroom instructional supports for DLLs' early learning outcomes. *Early Childhood Research Quarterly, 56*, 124–138. https://doi.org/10.1016/j.ecresq.2021.03.011

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., Duncan, G. J., Dynarski, M., Magnuson, K. A., & Weiland, C. (2016). *The current state of scientific knowledge on pre-kindergarten effects*. Brookings. https://www.brookings.edu/articles/puzzling-it-out-the-current-state-of-scientific-knowledge-on-pre-kindergarten-effects/

Pianta, R. C., DeCoster, J., Cabell, S., Burchinal, M., Hamre, B. K., Downer, J., LoCasale-Crouch, J., Williford, A., & Howes, C. (2014). Dose–response relations between preschool teachers' exposure to components of professional development and increases in quality of their interactions with children. *Early Childhood Research Quarterly, 29*(4), 499-508.

Pianta, R. C., & Hamre, B. K. (2022). *CLASS 2nd Edition Pre-K–3rd reference manual.* Teachstone.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) manual, Pre-K*. Paul H. Brookes Publishing.

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*(4), 431–451. https://doi.org/10.1016/j.ecresq.2008.02.001

Roberts, A., LoCasale-Crouch, J., Hamre, B., & DeCoster, J. (2016). Exploring teachers' depressive symptoms, interaction quality, and children's social-emotional development in Head Start. *Early Education and Development, 27*(5), 642–654. https://doi.org/10.1080/10409289.2016.1127088

Sandilos, L. E., & DiPerna, J. C. (2011). Interrater reliability of the Classroom Assessment Scoring System – Pre-K (CLASS Pre-K). *Journal of Early Childhood and Infant Psychology, 7,* 65–85.

Sawyer, B., Atkins-Burnett, S., Sandilos, L., Scheffner Hammer, C., Lopez, L., & Blair, C. (2018). Variations in classroom language environments of preschool children who are low income and linguistically diverse. *Early Education and Development, 29*(3), 398–416. https://doi.org/10.1080/10409289.2017.1408373

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness* (2nd ed.). The New Teacher Project. https://tntp.org/publication/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness/

White, L. J., Fernandez, V. A., & Greenfield, D. B. (2020). Assessing classroom quality for Latino dual language learners in Head Start: DLL-specific and general teacher-child interaction perspectives. *Early Education and Development, 31*(4), 599–627. https://doi.org/10.1080/10409289.2019.1680785

Wright, H. D. (2018). *Preschool teachers' efficacy and beliefs about children's social development: Potential impacts on interactions and structure in the classroom environment* [Doctoral dissertation, Texas Tech University]. Texas Tech University DSpace Repository. https://ttu-ir.tdl.org/items/505041bb-3815-4978-8f0a-0558421f2bd1

Appendix A. Descriptive Statistics

Exhibit A1. Teacher sample characteristics

Characteristic	Frequency (N)	Percent (%)
Gender		
Female	48	87.3%
Male	3	5.5%
Nonbinary	2	3.7%
Prefer not to respond	2	3.7%
Race		
Asian or Asian American	3	5.5%
Black or African American	15	27.3%
Hispanic, Latine, or Spanish origin	10	18.2%
Multiracial	3	5.5%
White	20	36.4%
Prefer not to respond	4	7.3%
Region		
California	7	12.1%
District of Columbia, Maryland, Virginia	32	55.2%
Massachusetts	19	32.8%
Experience		
1–2 years	8	14.3%
3–5 years	19	33.9%
6–10 years	11	19.6%
More than 10 years	18	32.1%
Degree		
Some college or technical school class	6	10.9%
Associate degree or technical degree	10	18.2%
Bachelor's degree	19	34.56%
Graduate or professional degree	17	30.9%
Prefer not to respond	3	5.5%

Exhibit A2. CLASS Descriptives

	Mean	*SD*	Min	Max
Domain Average				
Emotional Support	5.98	0.49	4.13	7.00
Classroom Organization	5.63	0.67	3.58	7.00
Instructional Support	2.73	0.71	1.33	5.67
Dimension Score				
Positive Climate	6.19	0.77	3.75	7.00
Negative Climate	1.13	0.26	1.00	2.50
Educator Sensitivity	5.84	0.72	4.00	7.00
Regard for Child Perspectives	5.01	0.81	3.00	7.00
Behavior Management	5.71	0.79	3.50	7.00
Productivity	6.04	0.75	3.50	7.00
Instructional Learning Formats	5.14	0.90	2.75	7.00
Concept Development	2.23	0.80	1.00	5.25
Quality of Feedback	2.85	0.86	1.25	6.25
Language Modeling	3.10	0.85	1.50	5.50

Exhibit A3. ECERS-3 Descriptives

	Mean	*SD*	Min	Max
Subscale Average				
Learning Activities	2.42	0.84	1.30	5.20
Interaction	4.63	1.26	1.75	6.75
Program Structure	4.22	1.28	1.00	6.67
ECERS-3 Total	3.39	0.71	1.94	5.16
Item Score				
Indoor space	4.26	1.38	1.00	7.00
Furnishings for care, play, and learning	3.63	0.89	1.00	7.00
Room arrangement for play and learning	3.35	1.41	1.00	7.00
Space for privacy	4.48	1.75	1.00	7.00
Child-related display	3.46	1.24	1.00	6.00
Meals/snack	2.36	1.29	1.00	7.00
Toileting/diapering	2.78	1.49	1.00	7.00
Health practices	2.69	1.38	1.00	6.00
Safety practices	4.40	1.72	1.00	7.00
Helping children expand vocabulary	4.58	1.53	1.00	7.00
Encouraging children to use language	4.67	1.53	1.00	7.00
Staff use of books with children	2.79	1.49	1.00	7.00
Encouraging children's use of books	2.92	1.22	1.00	6.00
Becoming familiar with print	3.36	1.18	1.00	6.00
Fine motor	2.74	1.85	1.00	7.00
Art	3.29	1.31	1.00	7.00
Music and movement	2.42	1.01	1.00	6.00
Blocks	1.93	1.08	1.00	5.00
Dramatic play	2.53	1.71	1.00	7.00
Nature/science	1.86	1.06	1.00	6.00
Math materials and activities	1.78	1.13	1.00	6.00
Math in daily events	2.80	1.27	1.00	7.00
Understanding written numbers	1.53	0.78	1.00	5.00
Promoting acceptance of diversity	3.31	1.34	1.00	7.00
Individualized teaching and learning	4.48	1.65	1.00	7.00
Staff-child interaction	5.40	1.77	1.00	7.00
Peer interaction	4.26	1.42	1.00	7.00
Discipline	4.38	1.43	1.00	7.00
Transitions and waiting times	4.27	1.89	1.00	7.00
Free play	4.00	1.65	1.00	7.00
Whole-group activities	4.40	1.55	1.00	7.00

Appendix B. CLASS Domains and ECERS Subscales

Exhibit B1. CLASS 2nd Edition Domains and Dimensions

Emotional Support
Positive Climate
Negative Climate
Educator Sensitivity
Regard for Child Perspectives
Classroom Organization
Behavior Management
Productivity
Instructional Learning Formats

Instructional Support
Concept Development
Quality of Feedback
Language Modeling

Exhibit B2. ECERS-3 Subscales and items

Space and Furnishings
Indoor space
Furnishings for care, play, and learning
Room arrangement for play and learning
Space for privacy
Child-related display
Space for gross motor play
Gross motor equipment
Personal Care Routines
Meals/snacks
Toileting/diapering
Health practices
Safety practices
Language and Literacy
Helping children expand vocabulary
Encouraging children to use language
Staff use of books with children
Encouraging children's use of books
Becoming familiar with print

Learning Activities
Fine motor
Art
Music and movement
Blocks
Dramatic play
Nature/science
Math materials and activities
Math in daily events
Understanding written numbers
Promoting acceptance of diversity
Appropriate use of technology
Interaction
Supervision of gross motor
Individualized teaching and learning
Staff-child interaction
Peer interaction
Discipline
Program Structure
Transitions and waiting times
Free play
Whole-group activities for play and learning

Appendix C. ECERS-3 Video Scoring Protocol

To transfer ECERS-3 to the video format, the study team developed a coding protocol wherein observers could indicate an indicator as *Cannot Score* if they were unable to score the item because of the video format. Certain indicators or full items that would require an observer to leave the classroom (e.g., space for gross motor, gross motor equipment, supervision of gross motor) were never permitted to be scored on video. However, observers had the option to mark any indicator as *Cannot Score*. Below, we provide a few illustrative examples:

- A video observer could not identify 10 positive examples of diversity in the classroom, as required for Indicator 5.2 in the promoting acceptance of diversity item. However, she could not see every part of the classroom, such as the exact books on the bookshelf or all the toys available in the dramatic play area. Therefore, she did not feel she had sufficient evidence to code this as absent in the classroom and instead coded it as *Cannot Score*.
- A video observer was in a classroom where an assistant teacher spent the entire free play time in the Block center and was the only teacher to engage with the children in that area. However, this assistant teacher did not wear a microphone, as is common when there are more than two teachers in the classroom. Therefore, although the observer could see that the assistant teacher was having many conversations with children, she was not able to toggle the microphone (or, as she would in a live observation, move toward that area) to understand whether the teacher pointed out math concepts during block play, as required for Indicator 7.3 for the Blocks item. Thus, she coded Indicator 7.3 as *Cannot Score* because she was not able to observe those conversations.
- A video observer was in a classroom where whole-group time took place in an area not fully visible on video. She was able to hear the whole-group activities because of the microphones, but she could not see the children. Therefore, she coded *Cannot Score* for Indicator 7.1 (all children in the group are actively engaged in group activities) of the *whole-group activities for play and learning* item because she was not able to see if the children were engaged and paying attention.

After coding videos, our first step for analysis was to review which items and subscales were most likely to have indicators marked as *Cannot Score*. Exhibit C1 presents the average percentage of indicators marked as *Cannot Score* in each item, organized by subscale.

Exhibit C1. Average percentage of indicators marked *Cannot Score* on video by item

ECERS-3 Item	Average percentage of indicators marked *Cannot Score*
Space and Furnishings	**6%***
Indoor space	12%
Furnishings for care, play, and learning	10%
Room arrangement for play and learning	3%
Space for privacy	1%
Child-related display	6%
Space for gross motor play	100%
Gross motor equipment	100%
Personal Care Routines	**26%**
Meals/snacks	21%
Toileting/diapering	36%
Health practices	17%
Safety practices	32%
Language and Literacy	**5%**
Helping children expand vocabulary	0%
Encouraging children to use language	3%
Staff use of books with children	2%
Encouraging children's use of books	12%
Becoming familiar with print	7%
Learning Activities	**7%**
Fine motor	8%
Art	3%
Music and movement	3%
Blocks	7%
Dramatic play	9%
Nature/science	4%
Math materials and activities	7%
Math in daily events	2%
Understanding written numbers	19%
Promoting acceptance of diversity	15%
Appropriate use of technology	4%
Interaction	**1%***
Supervision of gross motor	100%
Individualized teaching and learning	1%
Staff-child interaction	0%
Peer interaction	1%
Discipline	0%
Program Structure	**3%**
Transitions and waiting times	0%
Free play	2%
Whole-group activities for play and learning	7%

***** *Subscale averages omit gross motor items never scored on video (space for gross motor, gross motor equipment, supervision of gross motor).*

To incorporate the *Cannot Score* data into ECERS-3 scoring scheme, we considered three different approaches: (1) treating *Cannot Score* as N/A; (2) treating *Cannot Score* as No; and (3) treating *Cannot Score* as an indication of missing data. Each approach is detailed in Exhibit C2.

Exhibit C2. Pros and cons of ECERS video scoring approaches

Approach	Details	Pros	Cons
1. *Cannot Score* = N/A	Treat indicators scored as *Cannot Score* as "N/A" and follow ECERS-3 scoring guidelines for N/A (i.e., indicators marked N/A are not considered in the denominator of item scoring).	Follows existing ECERS-3 rules for scoring, doesn't penalize indicators that may have been missed due to video format.	May overweight observed indicators; for example, if we *cannot score* 3 of 4 Level 7 indicators, that one indicator will carry all the weight.
2. *Cannot Score* = No	Treat indicators scored as *Cannot Score* as "No" and follow ECERS-3 scoring guidelines for N/A (i.e., if the observer didn't see it on video, the box is unchecked and considered to not be present in the classroom).*	Doesn't require changes to coding practice, ensures all ECERS-3 indicators are scored.	May unfairly deflate scores on video if we consider something a "No" just because we couldn't see it in the video format.
3. *Cannot Score* = Missing	Treat indicators scored as *Cannot Score* as missing data and code the entire item as missing if any indicator that is typically needed to create the score is missing.	Minimizes assumptions made in scoring and only reports scores that we can be confident in.	Could result in high rates of missing data, leading to loss of information on what is happening in the classroom, in addition to what is already lost by not including gross motor.

** There are 7 exceptions to this where indicators captured the lack of some observed behavior or material. These were assigned Cannot Score = Yes to appropriately assign the score for the double negative, and include safety practices 3.1, 3.2, 5.1, and 7.1; discipline 3.1, 5.4; and transitions and waiting times 7.2.*

We first ruled out the *Cannot Score = Missing* approach because it resulted in high rates of missing data (33% of data missing, on average, ranging from 5% missing for Interaction items to 76% missing for Personal Care Routine items).

Next, we explored whether *Cannot Score = N/A* or *Cannot Score = No* resulted in more reliable scores both within and across modalities (Exhibit C3). In general, the *Cannot Score = No* approach resulted in better reliability in scores—particularly with respect to within-condition video inter-rater and test-retest reliability. However, between-condition inter-rater reliability (IRR) was better with the *Cannot Score = N/A* approach.

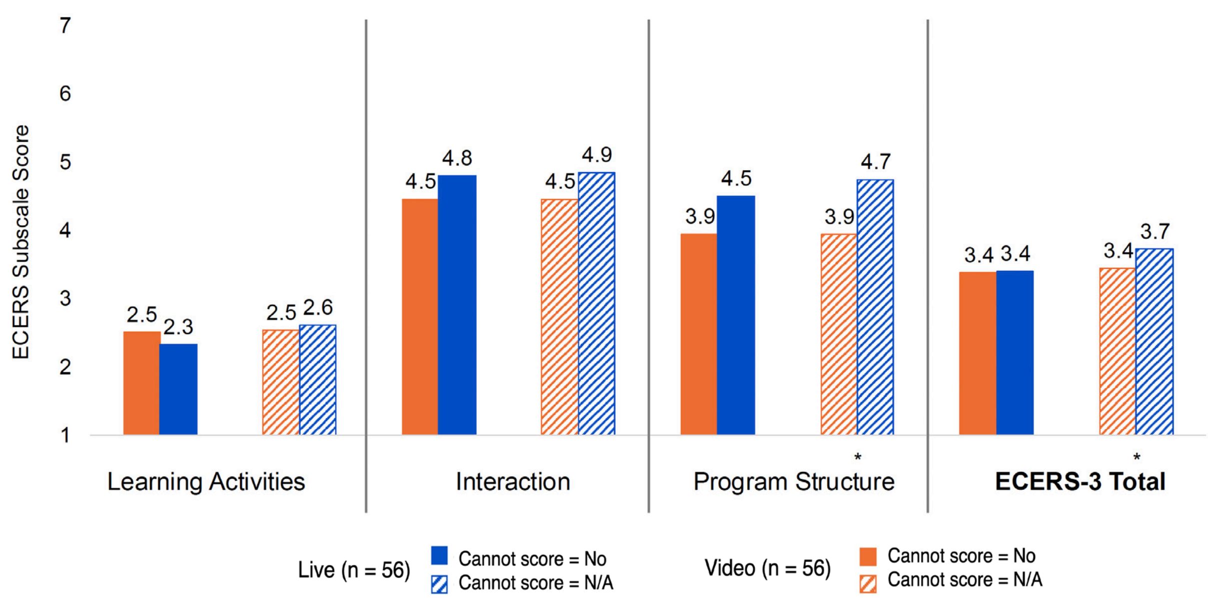
Exhibit C3. Reliability of different ECERS-3 scoring approaches

Subscale	IRR on Video		Video Test-Retest Reliability		Internal Consistency		IRR across Live and Video	
	Cannot Score = N/A	*Cannot Score = No*	*Cannot Score = N/A*	*Cannot Score = No*	*Cannot Score = N/A*	*Cannot Score = No*	*Cannot Score = N/A*	*Cannot Score = No*
Space and Furnishings	0.86	0.90	0.82	0.87	0.55	0.55	0.88	0.83
Personal Care Routines	0.75	0.83	0.72	0.84	0.6	0.58	0.89	0.77
Language and Literacy	0.90	0.90	0.83	0.86	0.49	0.33	0.90	0.85
Learning Activities	0.83	0.90	0.83	0.90	0.85	0.81	0.92	0.88
Interaction	0.89	0.88	0.84	0.85	0.78	0.77	0.94	0.94
Program Structure	0.88	0.85	0.83	0.85	0.75	0.63	0.93	0.92
Total	**0.85**	**0.88**	**0.82**	**0.87**	**0.92**	**0.89**	**0.91**	**0.86**
N	22 observations		24 classrooms		59 observations		59 observations	

Note. *All statistics omit items never scored on video (space for gross motor, gross motor equipment, supervision of gross motor), as well as appropriate use of technology, which was N/A in 66% of observations.*

Finally, we compared scores on the ECERS-3 subscales with sufficient internal consistency (internal consistency for Space and Furnishings [.55], Personal Care Routines [.58–.60], and Language and Literacy [.33–.49] was insufficient to use these subscales in analysis) across both live and video observations using both of these scoring approaches. We found that although the *Cannot Score = NA* approach had better percent-within-one IRR across live and video, differences across live and video within an observation were actually greater than in the *Cannot Score = No* approach (Exhibit C4). For this reason, we use *Cannot Score = No* in all subsequent analyses that use subscale scores.

Exhibit C4. Comparison of ECERS-3 live and video scores with different scoring approaches



Appendix D. Environmental and Technical Constraints Coding Protocol

Watch the first 30 seconds of every 10-minute block, and toggle between primary and secondary markers.

- CLASS: Watch bookmarked cycles only.
- ECERS-3: Watch video(s) in its entirety.

There are limitations to coding observations across all videos. However, when rating video and audio quality, please consider the scale of 1 to 5 as (1) completely uncodable to (5) the best possible quality with the technology used (e.g., a perfect Swivl recording with no issues) rather than comparing to a live observation.

Part 1: Video Quality

1. To what extent do issues with the **physical structure of the classroom** (e.g., multiple rooms; walls, pillars, or doors in the way; classroom is too large to see across) affect coding?
    1. Issues with the classroom structure make the video uncodable.
    2. Issues with the classroom structure make the video very difficult to code.
    3. Issues with the classroom structure make the video somewhat difficult to code.
    4. Issues with the classroom structure were minimal and generally do not affect the ability to code the video.
    5. Issues with the classroom structure either did not occur or do not affect the ability to code the video.
2. To what extent do issues with the **furniture layout** (e.g., bookcases, cubbies, interest centers) affect coding?
    1. Issues with the furniture layout make the video uncodable.
    2. Issues with the furniture layout make the video very difficult to code.
    3. Issues with the furniture layout make the video somewhat difficult to code.
    4. Issues with the furniture layout were minimal and generally do not affect the ability to code the video.
    5. Issues with the furniture layout either did not occur or do not affect the ability to code the video.
3. To what extent do issues with **camera focus** (e.g., blurry, fisheye lens blocking) affect coding?
    1. Issues with camera focus make the video uncodable.
    2. Issues with camera focus make the video very difficult to code.
    3. Issues with camera focus make the video somewhat difficult to code.
    4. Issues with camera focus were minimal and generally do not affect the ability to code the video.
    5. Issues with camera focus either did not occur or do not affect the ability to code the video.

4. To what extent do issues with **Swivl placement** (e.g., against a wall, in a corner far from action) affect coding?
    1. Issues with Swivl placement make the video uncodable.
    2. Issues with Swivl placement make the video very difficult to code.
    3. Issues with Swivl placement make the video somewhat difficult to code.
    4. Issues with Swivl placement were minimal and generally do not affect the ability to code the video.
    5. Issues with Swivl placement either did not occur or do not affect the ability to code the video.
5. To what extent do issues with **primary marker tracking** (i.e., how well the lead teacher was tracked by the camera) affect coding?
    1. Issues with primary marker tracking make the video uncodable.
    2. Issues with primary marker tracking make the video very difficult to code.
    3. Issues with primary marker tracking make the video somewhat difficult to code.
    4. Issues with primary marker tracking were minimal and generally do not affect the ability to code the video.
    5. Issues with primary marker tracking either did not occur or do not affect the ability to code the video.
6. What is your **overall assessment** of video quality for coding?
    1. Completely uncodable
    2. Very difficult to code
    3. Somewhat difficult to code
    4. Generally easy to code
    5. As easy to code as possible with the technology used
7. Would you like to elaborate on any of your prior responses? For example, were there any video quality issues not captured in Items 1–5? [*open-ended*]

Part 2: Audio Quality

1. To what extent do issues with **the primary microphone** (e.g., microphone was disconnected, microphone was hanging too low to hear well) affect coding?
    1. Issues with the primary microphone make the video uncodable.
    2. Issues with the primary microphone make the video very difficult to code.
    3. Issues with the primary microphone make the video somewhat difficult to code.
    4. Issues with the primary microphone were minimal and generally do affect the ability to code the video.
    5. Issues with the primary microphone either did not occur or do not affect the ability to code the

video.

2. To what extent do issues with the **secondary microphone** (e.g., microphone was disconnected, microphone was hanging too low to hear well) affect coding?
    1. Issues with the secondary microphones make the video uncodable.
    2. Issues with the secondary microphone make the video very difficult to code.
    3. Issues with the secondary microphone make the video somewhat difficult to code.
    4. Issues with the secondary microphone were minimal and generally do not affect the ability to code the video.
    5. Issues with the secondary microphone either did not occur or do not affect the ability to code the video.
3. To what extent do issues with the **ability to hear adults *without* microphones** affect coding?
    1. Issues with the ability to hear children make the video uncodable.
    2. Issues with the ability to hear children make the video very difficult to code.
    3. Issues with the ability to hear children make the video somewhat difficult to code.
    4. Issues with the ability to hear children were minimal and generally do not affect the ability to code the video.
    5. Issues with the ability to hear children either did not occur or do not affect the ability to code the video.
4. To what extent do issues with the **ability to hear children speak** affect coding?
    1. Issues with the ability to hear children make the video uncodable.
    2. Issues with the ability to hear children make the video very difficult to code.
    3. Issues with the ability to hear children make the video somewhat difficult to code.
    4. Issues with the ability to hear children were minimal and generally do not affect the ability to code the video.
    5. Issues with the ability to hear children either did not occur or do not affect the ability to code the video.
5. What is your **overall assessment** of audio quality for coding?
    1. Completely uncodable
    2. Very difficult to code
    3. Somewhat difficult to code
    4. Generally easy to code
    5. As easy to code as possible with the technology used
6. Would you like to elaborate on any of your prior responses? For example, were there any video quality issues not captured in Items 1–4? *[open-ended]*

Appendix E. RQ 2 Analysis Details

RQ 2a: Do live and video scores vary systematically within an observation?

To address RQ 2a, we estimated the following within-observation econometric fixed effects models, estimating separate models for each outcome:

$$ObservationScore_{iv} - \overline{ObservationScore_i} = \beta_1 (Video_{iv} - Video_i) + \beta_2(DaysSinceCertification_{iv} + \overline{DaysSinceCertification_i}) + \gamma_t (Observer) + \varepsilon_{im} - \overline{\varepsilon}_i$$

...for observation *i*, where *v* = video observation and *k* = number of observers.

In this model, $ObservationScore_{iv} - \overline{ObservationScore_i}$ represents the within-observation deviation from the average score for the observation across live and video; in other words, the observation score for observation *i* on video ($ObservationScore_{iv}$) minus the average score on that skill for observation *i* across live and video ($\overline{ObservationScore_i}$)= β_1 reflects the association between whether the score was assigned over video and the within-observation deviation from the average observation score across live and video. $DaysSinceCertification_{iv} + \overline{DaysSinceCertification_i}$ represents the difference between the days since certification for the video observer and the average number of days since certification across the live and video observers. γ_t *(Observer)* represents observer fixed effects (binary variables for each observer) that control for all things that are different across observers but are constant across observations (e.g., Observer X always assigns lower scores than a typical observer).

Additionally, although we estimate multiple models for this research question, we opt to present our main results without a correction for multiple comparisons (e.g., Benjamini-Hochberg correction). The Benjamini-Hochberg (BH) and similar corrections are common practice to control for the false discovery rate—in other words, the chances that estimates result in a significant finding ($p < .05$) by chance, rather than reflecting a true association—in studies where multiple tests are used to address the same research question (Benjamini & Hochberg, 1995). Such corrections are typically applied to control for the risk of false positive. In this study, however, the risk of a false *negative* (reporting that there are *no* differences across live and video when in fact there are) is more consequential than a false positive. Therefore, we do not correct for multiple comparisons.

RQ 2b: Do live and video scores vary differentially by program and observation characteristics?

To address RQ 2b, we estimated the following within-observation econometric fixed effects models with observation-level interactions, estimating separate models for each outcome and interaction term:

$$ObservationScore_{iv} - \overline{ObservationScore_i} = \beta_1 (Video_{iv} - \overline{Video_i}) + \beta_2(DaysSinceCertification_{iv} + \overline{DaysSinceCertification_i}) + \beta_3 (Video * Moderator_{iv} + \overline{Video * Moderator_i}) + \gamma_t (Observer) + \varepsilon_{im} - \overline{\varepsilon_i}$$

...for observation *i*, where *v* = video observation and *k* = number of observers.

In this model, the main effect, β_1, reflects the within-observation effect of video for observations in the omitted category (see table notes in Appendices F and G for the omitted category for each model). The interaction term, β_3, reflects the difference in the within-observation effect of video on the observation score for the moderator group compared to the omitted category.

Exhibit E1. Quality level cutoffs for CLASS scores

CLASS	Emotional Support	Classroom Organization	Instructional Support
Head Start Threshold			
Does not meet competitive threshold	1.00 – 4.99	1.00 – 4.99	1.00 – 2.49
Meets competitive but not quality threshold	5.00 – 5.99	5.00 – 5.99	2.50 – 2.99
Meets quality threshold	6.00 – 7.00	6.00 – 7.00	3.00 – 7.00
Sample-Specific Quintiles			
Bottom quintile	4.81 – 5.75	3.58 – 5.17	1.33 – 1.99
Middle quintile	5.81 – 6.38	5.25 – 6.33	2.00 – 3.17
Top quintile	6.48 – 7.00	6.33 – 7.00	3.25 – 5.67

Exhibit E2. Quality level cutoffs for ECERS-3 scores

ECERS-3	Learning Activities	Interaction	Program Structure	ECERS-3 Total
QRIS Threshold				
Does not meet 3-star level	N/A	N/A	N/A	1.00 – 3.49
Meets 3-star level	N/A	N/A	N/A	3.50 – 7.00
Data Quintiles				
Bottom quintile	1.40 – 1.90	1.75 – 3.25	1.00 – 3.33	1.94 – 2.87
Middle quintile	2.00 – 2.90	3.50 – 5.50	3.67 – 5.00	2.97 – 4.03
Top quintile	3.40 – 5.20	5.75 – 6.50	5.33 – 6.67	4.06 – 5.16

Appendix F. CLASS Analysis Tables

Exhibit F1. Within-observation fixed effect regression models predicting CLASS domain scores

	Emotional Support	Classroom Organization	Instructional Support
Video	-0.15*	-0.08	-0.10
	(0.07)	(0.07)	(0.10)
N	202	202	202

p* < .05, ** *p* < .01, * *p* < .001

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors.*

Exhibit F2. Within-observation fixed effect regression models predicting CLASS dimension scores

	Positive Climate	Negative Climate	Educator Sensitivity	Regard for Child Perspectives	Behavior Management	Productivity	Instructional Learning Formats	Concept Development	Quality of Feedback	Language Modeling
Video	-0.20*	0.03	0.02	-0.39***	-0.05	-0.13	-0.05	-0.22	-0.00	-0.09
	(0.09)	(0.03)	(0.11)	(0.10)	(0.10)	(0.09)	(0.10)	(0.11)	(0.11)	(0.12)
N	202	202	202	202	202	202	202	202	202	202

p* < .05, ** *p* < .01, * *p* < .001

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors.*

Exhibit F3. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and unadjusted live score

	Emotional Support	Classroom Organization	Instructional Support
Video	-0.16*	-0.12	-0.05
	(0.06)	(0.08)	(0.09)
Video x Live Score	-0.19***	-0.30***	-0.41***
	(0.04)	(0.06)	(0.06)
N	202	202	202

p* < .05, ** *p* < .01, * *p* < .001

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors.*
Unadjusted live scores are mean-centered.

Exhibit F4. Between-observation regression models predicting CLASS domain scores from interactions between video and unadjusted live score

	Emotional Support	Classroom Organization	Instructional Support
Video	-0.18**	-0.17*	-0.01
	(0.06)	(0.08)	(0.08)
Live Score	0.38***	0.56***	0.61***
	(0.02)	(0.04)	(0.04)
Video x Live Score	-0.21***	-0.36***	-0.42***
	(0.04)	(0.07)	(0.06)
N	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. Unadjusted live scores are mean-centered.*

Exhibit F5. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and live-score Head Start quality threshold

	Emotional Support	Classroom Organization	Instructional Support
Video	-0.26***	-0.37**	-0.47**
	(0.06)	(0.12)	(0.14)
Video x Does not meet HS competitive threshold	0.70***	0.64**	0.75***
	(0.13)	(0.20)	(0.16)
Video x Meets HS competitive threshold but not quality threshold	0.20*	0.38**	0.28
	(0.09)	(0.13)	(0.16)
N	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations that did not meet the competitive threshold. HS = Head Start.*

Exhibit F6. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and live-score quality quintiles

	Emotional Support	Classroom Organization	Instructional Support
Video	-0.17*	-0.09	-0.05
	(0.07)	(0.08)	(0.10)
Video x Bottom quintile	0.27*	0.33*	0.51**
	(0.12)	(0.13)	(0.16)
Video x Top quintile	-0.28**	-0.55***	-0.53**
	(0.10)	(0.14)	(0.15)
N	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with live scores in the middle three quintiles.*

Exhibit F7. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and program type

	Emotional Support	Classroom Organization	Instructional Support
Video	-0.21*	-0.14	-0.14
	(0.10)	(0.11)	(0.15)
Video x Public program	0.09	0.09	0.06
	(0.11)	(0.14)	(0.18)
N	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors.*

Exhibit F8. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and language

	Emotional Support	Classroom Organization	Instructional Support
Video	-0.17*	-0.13	-0.18
	(0.07)	(0.08)	(0.11)
Video x Multilingual	0.13	0.32	0.50**
	(0.11)	(0.17)	(0.15)
N	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors.*

Exhibit F9. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and video technology issues

Technology Issue	Camera Focus Issues			Swivl Placement Issues			Primary Marker Tracking Issues		
CLASS Domain	*ES*	*CO*	*IS*	*ES*	*CO*	*IS*	*ES*	*CO*	*IS*
Video	-0.15*	-0.03	-0.10	0.09	0.14	-0.10	-0.05	0.05	0.06
	(0.07)	(0.09)	(0.11)	(0.12)	(0.14)	(0.16)	(0.08)	(0.09)	(0.11)
Video x issues that affect coding	-0.11	-0.42	-0.21	-0.42**	-0.26	0.06	-0.22	-0.39**	-0.09
	(0.17)	(0.22)	(0.31)	(0.15)	(0.21)	(0.17)	(0.13)	(0.14)	(0.18)
Video x minor issues	0.03	-0.07	0.03	-0.28*	-0.28*	-0.02	-0.18	-0.20	-0.37**
	(0.09)	(0.13)	(0.10)	(0.12)	(0.14)	(0.15)	(0.10)	(0.14)	(0.13)
N	202	202	202	202	202	202	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with no issues. ES = Emotional Support, CO = Classroom Organization, IS = Instructional Support.*

Exhibit F10. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and microphone technology issues

Technology Issue	Primary Microphone Issues			Secondary Microphone Issues		
CLASS Domain	*ES*	*CO*	*IS*	*ES*	*CO*	*IS*
Video	-0.13	-0.08	-0.17	-0.13	-0.06	-0.17
	(0.07)	(0.08)	(0.12)	(0.07)	(0.09)	(0.11)
Video x issues that affect coding				-0.24*	-0.13	0.20
				(0.12)	(0.26)	(0.18)
Video x minor issues	-0.09	-0.06	0.15	-0.03	-0.04	0.19
	(0.09)	(0.11)	(0.14)	(0.11)	(0.14)	(0.12)
N	200	200	200	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with no issues. Primary microphones issues that affect coding were only present in one observation and, as such, were omitted from the model. ES = Emotional Support, CO = Classroom Organization, IS = Instructional Support.*

Exhibit F11. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and challenges with classroom layout

Classroom Challenge	Challenges With Classroom Structure			Challenges With Furniture Layout		
CLASS Domain	*ES*	*CO*	*IS*	*ES*	*CO*	*IS*
Video	-0.17*	-0.09	-0.10	-0.10	0.08	-0.37**
	(0.07)	(0.08)	(0.10)	(0.09)	(0.10)	(0.11)
Video x Challenges that affect coding	0.10	0.03	-0.31	-0.21	-0.44*	0.28
	(0.13)	(0.15)	(0.21)	(0.16)	(0.19)	(0.23)
Video x Minor challenges	0.07	0.05	0.09	-0.04	-0.15	0.38***
	(0.09)	(0.11)	(0.14)	(0.10)	(0.12)	(0.11)
N	202	202	202	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with no issues. ES = Emotional Support, CO = Classroom Organization, IS = Instructional Support.*

Exhibit F12. Within-observation fixed effect regression models predicting CLASS domain scores from interactions between video and other audio challenges

Audio Challenge	Challenges Hearing Adults Without Microphones			Challenges Hearing Children		
CLASS Domain	*ES*	*CO*	*IS*	*ES*	*CO*	*IS*
Video	-0.18*	-0.09	-0.19	-0.17*	-0.10	-0.25*
	(0.09)	(0.09)	(0.11)	(0.07)	(0.10)	(0.11)
Video x Challenges that affect coding	0.00	0.03	0.02	0.04	0.02	0.14
	(0.17)	(0.18)	(0.17)	(0.12)	(0.19)	(0.14)
Video x Minor challenges	0.08	0.03	0.27*	0.03	0.04	0.30*
	(0.11)	(0.13)	(0.12)	(0.08)	(0.12)	(0.14)
N	202	202	202	202	202	202

**p < .05, ** p < .01, *** p < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with no issues. ES = Emotional Support, CO = Classroom Organization, IS = Instructional Support.*

Exhibit F13. Within-observation fixed effect regression models predicting Head Start quality thresholds

	Meets Head Start Quality Threshold			Meets Head Start Competitive Threshold		
	ES	*CO*	*IS*	*ES*	*CO*	*IS*
Video	-0.07	-0.06	0.03	-0.06	-0.01	-0.06
	(0.08)	(0.07)	(0.06)	(0.05)	(0.06)	(0.07)
N	202	202	202	202	202	202

p* < .05, ** *p* < .01, * *p* < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. ES = Emotional Support, CO = Classroom Organization, IS = Instructional Support.*

Exhibit F14. Within-observation fixed effect regression models predicting Head Start quality thresholds from interactions between video and unadjusted live scores

	Meets Head Start Quality Threshold			Meets Head Start Competitive Threshold		
	ES	*CO*	*IS*	*ES*	*CO*	*IS*
Video	-0.08	-0.06	-0.09	-0.02	0.05	-0.04
	(0.07)	(0.05)	(0.07)	(0.06)	(0.05)	(0.07)
Video x Unadjusted live score	-0.17***	-0.07	-0.19**	-0.07	-0.14**	-0.17**
	(0.05)	(0.07)	(0.07)	(0.06)	(0.05)	(0.06)
N	202	202	202	202	202	202

p* < .05, ** *p* < .01, * *p* < .001*

Note. *Models control for observer and days since observer's most recent CLASS certification and use classroom-level cluster-robust standard errors. ES = Emotional Support, CO = Classroom Organization, IS = Instructional Support.*

Appendix G. ECERS-3 Analysis Tables

Exhibit G1. Within-observation regression models predicting ECERS-3 subscale averages

	Learning Activities	Interaction	Program Structure	ECERS-3 Total
Video	-0.18	0.34	0.56	0.03
	(0.13)	(0.26)	(0.31)	(0.13)
N	118	118	118	118

**p < .05, ** p < .01, *** p < .001*
Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G2. Within-observation regression models predicting ECERS-3 Space and Furnishings items

	Indoor Space	Furnishings for Care, Play, and Learning	Room Arrangement for Play and Learning	Space for Privacy	Child-Related Display
Video	-0.24	-0.81**	0.41	-0.45	0.06
	(0.52)	(0.28)	(0.38)	(0.48)	(0.43)
N	118	118	118	118	118

**p < .05, ** p < .01, *** p < .001*
Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G3. Within-observation regression models predicting ECERS-3 Personal Care Routines items

	Meals/Snacks	Toileting/Diapering	Health Practices	Safety Practices
Video	-0.38	0.36	0.19	1.70***
	(0.35)	(0.50)	(0.34)	(0.39)
N	118	118	118	118

**p < .05, ** p < .01, *** p < .001*
Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G4. Within-observation regression models predicting ECERS-3 Language and Literacy items

	Helping Children Expand Vocabulary	Encouraging Children to Use Language	Staff Use of Books With Children	Encouraging Children's Use of Books	Becoming Familiar With Print
Video	0.40	0.09	0.21	-0.96**	-1.00*
	(0.31)	(0.30)	(0.32)	(0.32)	(0.38)
N	118	118	118	118	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G5. Within-observation regression models predicting ECERS-3 Learning Activities items

	Fine Motor	Art	Music and Movement	Blocks	Dramatic Play	Nature/ Science	Math Materials/ Activities	Math in Daily Events	Understanding Written Numbers	Promoting Acceptance of Diversity
Video	-0.62	-0.16	0.16	-0.15	-0.20	-0.35	-0.14	-0.07	0.08	-0.35
	(0.48)	(0.35)	(0.25)	(0.23)	(0.30)	(0.28)	(0.16)	(0.29)	(0.19)	(0.32)
N	118	118	118	118	118	118	118	118	118	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G6. Within-observation regression models predicting ECERS-3 Interaction items

	Individualized Teaching and Learning	Staff-Child Interaction	Peer Interaction	Discipline
Video	-0.54	0.91	0.51	0.47
	(0.37)	(0.53)	(0.45)	(0.35)
N	118	118	118	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G7. Within-observation regression models predicting ECERS-3 Program Structure items

	Transitions and Waiting Times	Free Play	Whole-Group Activities for Play and Learning
Video	1.04	-0.12	0.76*
	(0.61)	(0.50)	(0.36)
N	118	118	118

p* < .05, ** *p* < .01, * *p* < .001

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G8. Within-observation regression models predicting ECERS-3 subscales from interactions between video and unadjusted live scores

	Learning Activities	Interaction	Program Structure	ECERS-3 Total
Video	-0.42**	0.16	0.37	-0.13
	(0.13)	(0.23)	(0.29)	(0.14)
Video x Unadjusted live score	-0.36***	-0.48***	-0.43**	-0.23**
	(0.06)	(0.10)	(0.14)	(0.07)
N	118	118	118	118

p* < .05, ** *p* < .01, * *p* < .001

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors. Unadjusted live scores are mean-centered.*

Exhibit G9. Between-observation regression models predicting ECERS-3 subscales from interactions between video and unadjusted live scores

	Learning Activities	Interaction	Program Structure	ECERS-3 Total
Video	-0.28***	0.33*	0.55***	-0.02
	(0.06)	(0.15)	(0.15)	(0.07)
Unadjusted live score	0.92***	1.31***	1.19***	0.73***
	(0.02)	(0.03)	(0.03)	(0.02)
Video x Unadjusted live score	-0.33***	-0.45***	-0.38**	-0.21**
	(0.07)	(0.10)	(0.13)	(0.07)
N	118	118	118	118

p* < .05, ** *p* < .01, * *p* < .001

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors. Unadjusted live scores are mean-centered.*

Exhibit G10. Within-observation regression models predicting ECERS-3 subscales from interactions between video and live-score quality quintiles

	Learning Activities	Interaction	Program Structure	ECERS-3 Total
Video	-0.36**	0.12	0.22	-0.13
	(0.12)	(0.28)	(0.36)	(0.13)
Video x Bottom quintile	0.44**	0.75**	0.72	0.57**
	(0.13)	(0.24)	(0.36)	(0.18)
Video x Top quintile	-0.58**	-0.43*	-0.61	-0.01
	(0.20)	(0.20)	(0.41)	(0.17)
N	118	118	118	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with live scores in the middle three quintiles.*

Exhibit G11. Within-observation regression models predicting ECERS-3 total score from interactions between video and live QRIS quality threshold

	ECERS-3 Total
Video	0.01
	(0.13)
Video x QRIS 3-star quality level	-0.29
	(0.16)
N	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G12. Within-observation regression models predicting ECERS-3 subscales from interactions between video and program type

	Learning Activities	Interaction	Program Structure	ECERS-3 Total
Video	-0.34	0.69*	0.92**	0.10
	(0.23)	(0.29)	(0.32)	(0.17)
Video x Public program	0.19	-0.41	-0.42	-0.09
	(0.17)	(0.22)	(0.26)	(0.10)
N	118	118	118	118

*$p < .05$, ** $p < .01$, *** $p < .001$

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G13. Within-observation regression models predicting ECERS-3 subscales from interactions between video and classroom language

	Learning Activities	Interaction	Program Structure	ECERS-3 Total
Video	-0.22	0.35	0.56	0.02
	(0.12)	(0.27)	(0.32)	(0.13)
Video x Multilingual	-0.37	0.04	0.01	-0.09
	(0.23)	(0.22)	(0.36)	(0.15)
N	118	118	118	118

*$p < .05$, ** $p < .01$, *** $p < .001$

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G14. Within-observation regression models predicting ECERS-3 subscales from interactions between video and video technology challenges

Technology Issue	Camera Focus Issues				Swivl Placement Issues				Primary Marker Tracking Issues			
ECERS-3 Subscale	*LA*	*I*	*PS*	*Total*	*LA*	*I*	*PS*	*Total*	*LA*	*I*	*PS*	*Total*
Video	-0.22	0.29	0.51	0.02	-0.15	0.35	0.54	0.02	-0.18	0.40	0.54	0.03
	(0.16)	(0.23)	(0.26)	(0.13)	(0.13)	(0.30)	(0.41)	(0.13)	(0.13)	(0.28)	(0.35)	(0.13)
Video x issues that affect coding	-0.15	-0.18	-0.36	-0.31	0.24	0.15	-0.10	0.13	0.15	-0.39	0.12	-0.05
	(0.28)	(0.28)	(0.43)	(0.29)	(0.13)	(0.33)	(0.39)	(0.13)	(0.18)	(0.35)	(0.53)	(0.21)
Video x minor issues	0.12	0.17	0.20	0.06	-0.07	-0.02	0.04	-0.00	-0.02	-0.15	0.07	0.00
	(0.12)	(0.28)	(0.36)	(0.13)	(0.16)	(0.21)	(0.35)	(0.14)	(0.16)	(0.27)	(0.36)	(0.13)
N	118	118	118	118	118	118	118	118	118	118	118	118

p* < .05, ** *p* < .01, * *p* < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust Note. Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with no issues. LA = Learning Activities, I = Interaction, PS = Program Structure, Total = ECERS-3 Total.*

Exhibit G15. Within-observation regression models predicting ECERS-3 subscales from interactions between video and microphone technology challenges

Technology Issue	Primary Microphone Issues				Secondary Microphone Issues			
ECERS-3 Subscale	*LA*	*I*	*PS*	*Total*	*LA*	*I*	*PS*	*Total*
Video	-0.23	0.35	0.57	0.01	-0.20	0.34	0.57	0.03
	(0.14)	(0.25)	(0.31)	(0.14)	(0.13)	(0.26)	(0.30)	(0.13)
Video x issues that affect coding					-0.05	-0.27	0.27	-0.02
					(0.19)	(0.39)	(0.52)	(0.25)
Video x minor issues	0.19	-0.15	-0.15	-0.00	0.30*	0.13	-0.20	-0.01
	(0.10)	(0.27)	(0.44)	(0.14)	(0.14)	(0.28)	(0.42)	(0.12)
N	114	114	114	114	118	118	118	118

p* < .05, ** *p* < .01, * *p* < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with no issues. Primary microphones issues that affect coding were only present in two observations and, as such, were omitted from the model. LA = Learning Activities, I = Interaction, PS = Program Structure, Total = ECERS-3 Total.*

Exhibit G16. Within-observation regression models predicting ECERS-3 subscales from interactions between video and challenges with classroom layout

Classroom Challenge	Challenges With Classroom Structure				Challenges With Furniture Layout			
ECERS-3 Subscale	*LA*	*I*	*PS*	*Total*	*LA*	*I*	*PS*	*Total*
Video	-0.25	0.41	0.39	-0.04	-0.13	0.37	0.74*	-0.05
	(0.15)	(0.26)	(0.28)	(0.11)	(0.13)	(0.29)	(0.34)	(0.12)
Video x Challenges that affect coding	0.46**	-0.33	0.35	0.20	-0.22	0.11	0.20	0.09
	(0.15)	(0.23)	(0.40)	(0.16)	(0.25)	(0.27)	(0.29)	(0.16)
Video x minor issues	0.06	-0.10	0.39	0.12	-0.05	-0.07	-0.34	0.11
	(0.15)	(0.27)	(0.43)	(0.14)	(0.14)	(0.24)	(0.30)	(0.13)
N	118	118	118	118	118	118	118	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with no issues. LA = Learning Activities, I = Interaction, PS = Program Structure, Total = ECERS-3 Total.*

Exhibit G17. Within-observation regression models predicting ECERS-3 subscales from interactions between video and other audio challenges

Audio Challenge	Challenges Hearing Adults Without Microphones				Challenges Hearing Children			
ECERS-3 Subscale	*LA*	*I*	*PS*	*Total*	*LA*	*I*	*PS*	*Total*
Video	-0.15	0.34	0.45	0.01	-0.19	0.39	0.52	0.02
	(0.13)	(0.25)	(0.31)	(0.12)	(0.13)	(0.24)	(0.30)	(0.11)
Video x Challenges that affect coding	0.03	-0.07	0.51	0.11	0.15	0.03	0.21	0.11
	(0.17)	(0.48)	(0.62)	(0.31)	(0.13)	(0.29)	(0.43)	(0.14)
Video x minor issues	-0.10	0.02	0.26	0.04	-0.16	-0.61*	-0.01	-0.16
	(0.16)	(0.26)	(0.32)	(0.11)	(0.25)	(0.24)	(0.37)	(0.17)
N	118	118	118	118	118	118	118	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors. The omitted category for the interaction term reflects observations with no issues. LA = Learning Activities, I = Interaction, PS = Program Structure, Total = ECERS-3 Total.*

Exhibit G18. Within-observation regression models predicting typical ECERS-3 QRIS 3-star quality level

	QRIS 3-Star Quality Level
Video	0.12
	(0.09)
N	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Exhibit G19. Within-observation regression models predicting meeting typical ECERS-3 QRIS 3-star quality level from interactions between video and unadjusted live scores

	QRIS 3-Star Quality Level
Video	0.09
	(0.10)
Video x Unadjusted live score	-0.05
	(0.04)
N	118

**p < .05, ** p < .01, *** p < .001*

Note. *Models include observation-level fixed effects and control for observer and days since observer's most recent ECERS-3 certification, and use classroom-level cluster-robust standard errors.*

Appendix H. Reliability Analysis Details

To address our first research question whether CLASS and ECERS-3 observations yield comparable scores across live and video observations, the study team first compared the reliability of observation scores over video. Both CLASS and ECERS-3 had sufficient reliability on video, though some ECERS-3 subscales may not be appropriate given low internal consistency (Exhibit H1).

Exhibit H1. Video reliability statistics

	Video Reliability Statistics		
	Inter-Rater Reliability	**Test-Retest Reliability**	**Internal Consistency**
Class			
Emotional Support	90%	91%	0.72
Classroom Organization	87%	82%	0.74
Instructional Support	87%	88%	0.75
CLASS Total	**88%**	**88%**	**0.86**
N	*205 cycles*	*40 classrooms*	*101 observations*
ECERS-3			
Space and Furnishings	90%	87%	0.55
Personal Care Routines	83%	84%	0.58
Language and Literacy	90%	86%	0.33
Learning Activities	90%	90%	0.81
Interaction	88%	85%	0.77
Program Structure	85%	85%	0.63
ECERS-3 Total	**88%**	**87%**	**0.89**
N	22 observations	24 classrooms	59 observations

CLASS had sufficient inter-rater reliability (IRR), test-retest reliability, and internal consistency on video, both in total and within each subscale. Specifically, percent-within-one IRR for CLASS was 88%, meaning that when two observers score the same observation, they assign scores within 1 point 88% of the time. This level of reliability is over the cutoff of 80% required to become certified in CLASS, and is comparable to prior studies that use CLASS, both live (e.g., Brown et al., 2010; Pakarinen et al., 2023; Sandilos & DiPerna, 2011). Percent-within-one test-retest reliability was also 88%, meaning that when the same classroom is observed twice, observers assign the same score within 1 point 88% of the time, a rate comparable to test-retest reliability for live observations in this study (see Exhibit H2) and in past studies that use CLASS live (e.g., Sandilos & DiPerna, 2011). Finally, internal consistency, measured using Cronbach's alpha, was above the .70

cutoff for good reliability, ranging from .72 for Emotional Support to .75 for Instructional Support. The internal consistency for Emotional Support and Classroom Organization was similar on video compared to live but was higher live for Instructional Support (.87)

Exhibit H2. Live reliability statistics

	Live Reliability Statistics	
	Test-Retest Reliability	**Internal Consistency**
CLASS		
Emotional Support	95%	0.71
Classroom Organization	89%	0.77
Instructional Support	97%	0.87
CLASS Total	**94%**	**0.86**
N	*40 classrooms*	*101 observations*
ECERS-3		
Space and Furnishings	87%	0.44
Personal Care Routines	88%	0.66
Language and Literacy	83%	0.58
Learning Activities	90%	0.88
Interaction	88%	0.83
Program Structure	86%	0.62
ECERS-3 Total	**88%**	**0.91**
N	24 classrooms	59 observations

Likewise, ECERS-3 had sufficient IRR and test-retest reliability, although internal consistency varied across subscales. Specifically, percent-within-one IRR for ECERS-3 was 88%, meaning that when two observers score the same observation, they assign scores within 1 point 88% of the time. This level of reliability is over the cutoff of 80% required to become certified in ECERS-3 and is comparable to prior studies that use ECERS-3 live (e.g., Hestenes et al., 2019; Wright, 2018). Percent-within-one test-retest reliability was 87%, meaning that when the same classroom is observed twice, observers assign the same score within 1 point 87% of the time, a rate comparable to test-retest reliability for live observations in this study (see Exhibit H2) and in past studies that use ECERS-3 live (e.g., Hestenes et al., 2019). Finally, internal consistency, measured using Cronbach's alpha, was above the .70 cutoff for acceptable reliability for the total ECERS-3 score, Learning Activities, and Interaction, and over the .60 cutoff for acceptable reliability for Program Structure. For Space and Furnishings and Language and Literacy, however, the alphas reflect unacceptable levels of internal consistency. This pattern was consistent across live and video observations. This pattern is also consistent with past research with the ECERS-3 in which a confirmatory factor analyses revealed a weak model fit for the six published subscales and subsequent exploratory factor analyses revealed different factor structures than the published subscales (Early et al., 2018; Montes et al., 2018). Since factor analysis was outside the scope of this project, we opted to use the original subscales but exclude the Space and Furnishings, Personal Care Routines, and Language and Literacy subscales due to their low internal consistency.

Next, we examined whether live and video observations had sufficient between-condition IRR; in other words, whether observers assign similar ratings to the same observation period scored on video and live. Both CLASS and ECERS-3 had sufficient (over 80%) percent-within-one IRR between live and video observations (Exhibit H3). However, reliability was lower for items in the ECERS-3 Personal Care Routines subscale, which had 77% reliability.

Exhibit H3. Inter-rater reliability across live and video

	IRR across live and video
CLASS	
Emotional Support	90%
Classroom Organization	84%
Instructional Support	91%
CLASS Total	**88%**
N	*399 cycles*
ECERS-3	
Space and Furnishings	83%
Personal Care Routines	77%
Language and Literacy	85%
Learning Activities	88%
Interaction	94%
Program Structure	92%
ECERS-3 Total	**84%**
N	59 observations

Appendix I. Cost Analysis Details

Exhibit I1 presents detailed cost estimates for the first year of the accountability scenario using CLASS and comparing the video and live methods. This scenario hypothesizes 15 observers conducting 300 observations in 300 classrooms across 100 sites. The 15 observers require training, and the estimate includes both training fees and the labor hours required to complete the training. For video observations, an audio-video technician can set up equipment in multiple classrooms and conduct recordings simultaneously. In this hypothetical scenario, we assumed the technician can record three classrooms at one site simultaneously. Thus, each technician has three sets of equipment, equaling 45 sets of equipment for the 15 technicians. Based on SRI staff experiences with video recordings, we assumed it would take 3 hours to make each recording, which includes setup time. Regarding travel, the 15 technicians would only need to make a total of 105 trips because each of them can make three recordings at a time. Once the recordings are made, an observer must code the observations, which based on our experience takes an estimated 2.5 hours.

The live observation method does not have the costs for recording equipment, but the travel costs are higher. Once we account for travel time and setup, we estimate that each observation will take 3.5 hours. This would most likely occur before lunch. Many preschool classrooms have afternoon activities that take place outside of the classroom or with a professional other than their typical teacher (e.g., art, games), so we assumed that observers will only be able to conduct one observation per classroom per day. Thus, the observers will need to make a total of 300 trips.

The bottom of Exhibit I1 shows costs if 100% of travel is local, if 100% of travel is distant, and if 90% of travel is local and 10% is distant. When comparing costs between live and video observations, we chose the 100% local travel for the video method and the 90% local and 10% distant for the live method (see Costs of Video and Live Observations [RQ 4] for details). Because audio-video technicians can make the recordings, they can come from many programs or districts around a state, so it is likely they only must travel local distances. Live observers require specialized training and will have a more specialized skill set compared to the technicians. Because there are fewer observers, they are more likely to have to travel further distances to complete the observations.

Exhibit I1. Year 1 costs for CLASS video and live observations: accountability scenario

Ingredient	CLASS Video			CLASS Live		
	Unit Price	Unit Quantity	Cost	Unit Price	Unit Quantity	Cost
Personnel						
Training (fees & labor hours)	$3,041.90	15	$45,628.44	$3,041.90	15	$45,628.44
Observer labor per 1 observation	$92.80	300	$27,840.00	$129.92	300	$38,976.00
A/V technician labor per 3 observations	$28.49	100	$2,849.00	-		-
Travel						
Local	$39.19	105	$4,114.95	$39.19	300	$11,757.00
Distant	$251.69	105	$26,427.45	$251.69	300	$75,507.00
Equipment						
Microlens	$54.57	45	$2,455.65			
Swivl	$1,691.82	45	$76,131.90			
Swivl stand	$186.31	45	$8,383.95			
Tripod carrying case	$31.90	45	$1,435.50			
Lanyard	$28.23	45	$1,270.35			
CLASS score sheet	$28.23	300	$8,469.00	$28.23	300	$8,469.00
CLASS field guide	$129.08	15	$1,936.20	$129.08	15	$1,936.20
Clipboard	$4.60	0	$0.00	$4.60	15	$69.00
File folder	$2.32	300	$696.00	$2.32	300	$696.00
Screen protector	$2.23	0	$0.00	$2.23	15	$33.45
Subtotal: Local			**$181,210.94**			**$107,565.09**
Subtotal: Distant			**$203,523.44**			**$171,315.09**
Subtotal: 90% local and 10% distant			**$183,442.19**			**$113,940.09**

Exhibit I2 presents annual Year 2 and Year 3 costs for the accountability scenario using CLASS for both the video and live methods. We assumed that the 15 staff who were trained to administer CLASS during the first year will continue to do so and that there will be no additional training costs. We also assumed that most of the recording equipment can be reused, but that 20% of it will need to be replaced due to loss or damage. This results in a substantial decrease in the costs to complete the video recordings. On the contrary, there are no decreases in travel for live observations, which results in video observations being cheaper during Year 2 by $4,402.03. The costs for Year 3 are the same as for Year 2.

Exhibit I2. Annual Year 2 and 3 costs for CLASS video and live observations: accountability scenario

Ingredient	CLASS Video			CLASS Live		
	Unit Price	Unit Quantity	Cost	Unit Price	Unit Quantity	Cost
Personnel						
Training (fees & labor hours)	$3,041.90	0	$0.00	$3,041.90	0	$0.00
Observer labor per 1 observation	$92.80	300	$27,840.00	$129.92	300	$38,976.00
Recorder labor per 3 observations	$28.49	100	$2,849.00	-		-
Travel						
Local	$39.19	105	$4,114.95	$39.19	300	$11,757.00
Distant	$251.69	105	$26,427.45	$251.69	300	$75,507.00
Equipment						
Microlens	$54.57	9	$491.13			
Swivl	$1,691.82	9	$15,226.38			
Swivl stand	$186.31	9	$1,676.79			
Tripod carrying case	$31.90	9	$287.10			
Lanyard	$28.23	9	$254.07			
CLASS score sheet	$28.23	300	$8,469.00	$28.23	300	$8,469.00
CLASS field guide	$129.08	0	$0.00	$129.08	0	$0.00
Clipboard	$4.60	0	$0.00	$4.60	0	$0.00
File folder	$2.32	300	$696.00	$2.32	300	$696.00
Screen protector	$2.23	0	$0.00	$2.23	15	$33.45
Subtotal: Local			**$61,904.42**			**$59,931.45**
Subtotal: Distant			**$84,216.92**			**$123,681.45**
Subtotal: 90% local and 10% distant			**$64,135.67**			**$66,306.45**

Exhibit I3 presents detailed cost estimates for the first year of the accountability scenario using ECERS-3 and comparing the video and live methods. This scenario uses the same assumptions as for the CLASS observations above with respect to number of observers, audio-video technicians, sets of recording equipment, and travel. Although ECERS-3 observations take longer to complete than CLASS observations, there is no effect on travel for either the video or live methods. This is because audio-video technicians can make multiple recordings simultaneously, similar to CLASS recordings. Similarly, live observers will only be able to conduct one observation per classroom per day, which is the same as using CLASS.

Exhibit I3. Year 1 costs for ECERS-3 video and live observations: accountability scenario

Ingredient	ECERS-3 Video			ECERS-3 Live		
	Unit Price	Unit Quantity	Cost	Unit Price	Unit Quantity	Cost
Personnel						
Training (fees & labor hours)	$6,188.63	15	$92,829.45	$6,188.63	15	$92,829.45
Observer labor per 1 observation	$129.92	300	$38,976.00	$167.04	300	$50,112.00
Recorder labor per 3 observations	$56.98	100	$5,698.00	-		-
Travel						
Local	$39.19	105	$4,114.95	$39.19	300	$11,757.00
Distant	$251.69	105	$26,427.45	$251.69	300	$75,507.00
Equipment						
Microlens	$54.57	45	$2,455.65			
Swivl	$1,691.82	45	$76,131.90			
Swivl stand	$186.31	45	$8,383.95			
Tripod carrying case	$31.90	45	$1,435.50			
Lanyard	$28.23	45	$1,270.35			
ECERS-3 rating scale	$48.84	300	$14,652.00	$48.84	300	$14,652.00
Clipboard	$4.60			$4.60	15	$69.00
File folder	$2.32	300	$696.00	$2.32	300	$696.00
Screen protector	$2.23			$2.23	15	$33.45
Subtotal: Local			**$246,643.75**			**$170,148.90**
Subtotal: Distant			**$268,956.25**			**$233,898.90**
Subtotal: 90% local and 10% distant			**$248,875.00**			**$176,523.90**

Exhibit I4 presents annual Year 2 and Year 3 costs for the accountability scenario using ECERS-3 for both video and live observations. We assumed that the 15 staff who were trained to administer ECERS-3 during the first year will continue to do so and that there will be no additional training costs. We also assumed that most of the recording equipment can be reused, but that 20% of it will need to be replaced due to loss or damage. This results in a substantial decrease in the costs to complete the video recordings; however, this decrease is not enough to make it cheaper than live observations with 90% local and 10% distant travel. The costs for Year 3 are the same as those for Year 2.

Exhibit I4. Annual Year 2 and 3 costs for ECERS-3 video and live observations: accountability scenario

Ingredient	ECERS-3 Video			ECERS-3 Live		
	Unit Price	Unit Quantity	Cost	Unit Price	Unit Quantity	Cost
Personnel						
Training (fees & labor hours)	$6,188.63	0	$0.00	$6,188.63	0	$0.00
Observer labor per 1 observation	$129.92	300	$38,976.00	$167.04	300	$50,112.00
Recorder labor per 3 observations	$56.98	150	$8,547.00	-		-
Travel						
Local	$39.19	105	$4,114.95	$39.19	300	$11,757.00
Distant	$251.69	105	$26,427.45	$251.69	300	$75,507.00
Equipment						
Microlens	$54.57	9	$491.13			
Swivl	$1,691.82	9	$15,226.38			
Swivl stand	$186.31	9	$1,676.79			
Tripod carrying case	$31.90	9	$287.10			
Lanyard	$28.23	9	$254.07			
ECERS-3 rating scale	$48.84	300	$14,652.00	$48.84	300	$14,652.00
Clipboard	$4.60			$4.60	0	$0.00
File folder	$2.32	300	$696.00	$2.32	300	$696.00
Screen protector	$2.23			$2.23	0	$0.00
Subtotal: Local			**$84,921.42**			**$77,217.00**
Subtotal: Distant			**$107,233.92**			**$140,967.00**
Subtotal: 90% local and 10% distant			**$87,152.67**			**$83,592.00**

Exhibit I5 presents detailed cost estimates for the first year of the coaching scenario using CLASS and comparing the video and live methods. This scenario hypothesizes 50 observers conducting 1,500 observations in 300 classrooms across 100 sites; classrooms are observed five times each year so coaches can help teachers improve their practice. The 50 observers require training, and the estimates include both training fees and the labor hours required to complete the training. Because these observations occur five times a year in each classroom, for the video observations it would be more efficient for each site to have its own set of recording equipment and have its own staff conduct the recordings; thus, there are 100 sets of recording equipment. Further, there are no travel costs because each site would have its own staff make the recordings. Based on our experience, we estimate that the observers/coders (i.e., coaches) would need 2.5 hours to code each observation, but no travel is required for them to do this.

For the live observations, we assumed observers/coders will require 3.5 to complete an observation. This would most likely occur before lunch. Many preschool classrooms have afternoon activities that take place outside of the classroom or with a professional other than their typical teacher (e.g., art, games), so we assumed that observers will only be able to conduct one observation per classroom per day. Thus, the observers will need to make a total of 1,500 trips.

The bottom of Exhibit I5 presents subtotals for each observation method given travel costs. Note there is no difference in travel costs for video observations because we assumed no travel would be necessary. Although the video observations do not require travel, the initial investment in recording equipment makes it more expensive than live observations when travel is local or 90% local and 10% distant. When comparing costs between the methods, we chose 90% local and 10% distant for live observations because we assumed that most districts would have coaches locally available, but that a few would require visits from a coach located elsewhere, such as a state department of education.

Exhibit I5. Year 1 costs for CLASS video and live observations: coaching scenario

Ingredient	CLASS Video			CLASS Live		
	Unit Price	**Unit Quantity**	**Cost**	**Unit Price**	**Unit Quantity**	**Cost**
Personnel						
Training (fees & labor hours)	$3,041.90	50	$152,094.80	$3,041.90	50	$152,094.80
Observer labor per 1 observation	$92.80	1,500	$139,200.00	$129.92	1,500	$194,880.00
Recorder labor per 3 observations	$28.49	0	$0.00	-		-
Travel						
Local	$16.38	0	$0.00	$16.38	1,500	$24,570.00
Distant	$251.69	0	$0.00	$251.69	1,500	$377,535.00
Equipment						
Microlens	$54.57	100	$5,457.00			
Swivl	$1,691.82	100	$169,182.00			
Swivl stand	$186.31	100	$18,631.00			
Tripod carrying case	$31.90	100	$3,190.00			
Lanyard	$28.23	100	$2,823.00			
CLASS score sheet	$28.23	1,500	$42,345.00	$28.23	1,500	$42,345.00
CLASS field guide	$129.08	50	$6,454.00	$129.08	50	$6,454.00
Clipboard	$4.60	0	$0.00	$4.60	50	$230.00
File folder	$2.32	1,500	$3,480.00	$2.32	1,500	$3,480.00
Screen protector	$2.23	0	$0.00	$2.23	50	$111.50
Subtotal: Local			**$542,856.80**			**$424,165.30**
Subtotal: Distant			**$542,856.80**			**$777,130.30**
Subtotal: 90% local and 10% distant						**$459,461.80**

Exhibit I6 presents annual Year 2 and Year 3 costs for the coaching scenario using CLASS for both the video and live observations. We assumed that the 50 staff who were trained to administer CLASS during the first year will continue to do so and that there will be no additional training costs. We also assumed that most of the recording equipment can be reused, but that 20% of it will need to be replaced due to loss or damage. This results in a substantial decrease in the costs to complete the video recordings. By contrast, there are no decreases in travel for live observations, which results in the video observations being cheaper during Year 2 by $75,689.90. The costs for Year 3 are the same as those for Year 2.

Exhibit I6. Annual Year 2 and 3 costs for CLASS video and live observations: coaching scenario

Ingredient	CLASS Video			CLASS Live		
	Unit Price	Unit Quantity	Cost	Unit Price	Unit Quantity	Cost
Personnel						
Training (fees & labor hours)	$3,041.90	0	$0.00	$3,041.90	0	$0.00
Observer labor per 1 observation	$92.80	1,500	$139,200.00	$129.92	1,500	$194,880.00
Recorder labor per 3 observations	$28.49	0	$0.00	-		-
Travel						
Local	$16.38	0	$0.00	$16.38	1,500	$24,570.00
Distant	$251.69	0	$0.00	$251.69	1,500	$377,535.00
Equipment						
Microlens	$54.57	20	$1,091.40			
Swivl	$1,691.82	20	$33,836.40			
Swivl stand	$186.31	20	$3,726.20			
Tripod carrying case	$31.90	20	$638.00			
Lanyard	$28.23	20	$564.60			
CLASS score sheet	$28.23	1,500	$42,345.00	$28.23	1,500	$42,345.00
CLASS field guide	$129.08	0	$0.00	$129.08	0	$0.00
Clipboard	$4.60	0	$0.00	$4.60	0	$0.00
File folder	$2.32	1,500	$3,480.00	$2.32	1,500	$3,480.00
Screen protector	$2.23	0	$0.00	$2.23	0	$0.00
Subtotal: Local			**$224,881.60**			**$265,275.00**
Subtotal: Distant			**$224,881.60**			**$618,240.00**
Subtotal: 90% local and 10% distant						**$300,571.50**

Exhibit I7 presents detailed cost estimates for the first year of the coaching scenario using ECERS-3 and comparing the video and live methods. This scenario uses the same assumptions as the CLASS observations presented above with respect to number of observers, audio-video technicians, sets of recording equipment, and travel. Although ECERS-3 observations take longer to complete than CLASS observations, there is no effect on travel for either the video or live methods. This is because we assume there will be no travel for the video observations, but travel will be required for the live observations. Additionally, we assume that live observers will only be able to conduct one ECERS-3 observation per classroom per visit.

Exhibit I7. Year 1 costs for ECERS-3 video and live observations: coaching scenario

Ingredient	ECERS-3 Video			ECERS-3 Live		
	Unit Price	Unit Quantity	Cost	Unit Price	Unit Quantity	Cost
Personnel						
Training (fees & labor hours)	$6,188.63	50	$309,431.50	$6,188.63	50	$309,431.50
Observer labor per 1 observation	$129.92	1,500	$194,880.00	$167.04	1,500	$250,560.00
Recorder labor per 2 observations	$56.98	0	$0.00	-		-
Travel						
Local	$39.19	0	$0.00	$39.19	1,500	$58,785.00
Distant	$251.69	0	$0.00	$251.69	1,500	$377,535.00
Equipment						
Microlens	$54.57	100	$5,457.00			
Swivl	$1,691.82	100	$169,182.00			
Swivl stand	$186.31	100	$18,631.00			
Tripod carrying case	$31.90	100	$3,190.00			
Lanyard	$28.23	100	$2,823.00			
ECERS-3 rating scale	$48.84	1,500	$73,260.00	$48.84	1,500	$73,260.00
Clipboard	$4.60			$4.60	50	$230.00
File folder	$2.32	1,500	$3,480.00	$2.32	1,500	$3,480.00
Screen protector	$2.23			$2.23	50	$111.50
Subtotal: Local			**$780,334.50**			**$695,858.00**
Subtotal: Distant			**$780,334.50**			**$1,014,608.00**
Subtotal: 90% local and 10% distant						**$727,733.00**

Exhibit I8 presents annual Year 2 and Year 3 costs for the coaching scenario using ECERS-3 for both the video and live methods. We assumed that the 50 staff who were trained to administer ECERS-3 during the first year will continue to do so and that there will be no additional training costs. We also assumed that most of the recording equipment can be reused, but that 20% of it will need to be replaced due to loss or damage. This results in a substantial decrease in the costs to complete the video recordings, resulting in substantial savings compared to live observations. The costs for Year 3 are the same as those for Year 2.

Exhibit I8. Annual Year 2 and 3 costs for ECERS-3 video and live observations: coaching scenario

Ingredient	ECERS-3 Video			ECERS-3 Live		
	Unit Price	**Unit Quantity**	**Cost**	**Unit Price**	**Unit Quantity**	**Cost**
Personnel						
Training (fees & labor hours)	$6,188.63	0	$0.00	$6,188.63	0	$0.00
Observer labor per 1 observation	$129.92	1,500	$194,880.00	$167.04	1,500	$250,560.00
Recorder labor per 2 observations	$56.98	0	$0.00	-		-
Travel						
Local	$39.19	0	$0.00	$39.19	1,500	$58,785.00
Distant	$251.69	0	$0.00	$251.69	1,500	$377,535.00
Equipment						
Microlens	$54.57	20	$1,091.40			
Swivl	$1,691.82	20	$33,836.40			
Swivl stand	$186.31	20	$3,726.20			
Tripod carrying case	$31.90	20	$638.00			
Lanyard	$28.23	20	$564.60			
ECERS-3 rating scale	$48.84	1,500	$73,260.00	$48.84	1,500	$73,260.00
Clipboard	$4.60			$4.60	0	$0.00
File folder	$2.32	1,500	$3,480.00	$2.32	1,500	$3,480.00
Screen protector	$2.23			$2.23	0	$0.00
Subtotal: Local			**$311,476.60**			**$386,085.00**
Subtotal: Distant			**$311,476.60**			**$704,835.00**
Subtotal: 90% local and 10% distant						**$417,960.00**

Exhibit I9 presents per-observation costs for each of the three years, as well as the average for all three years, for CLASS and ECERS-3, both live and video, and using the Accountability and Coaching use cases. These calculations use the same assumptions detailed in the previous tables. Per-observation costs are calculated by dividing the total cost by the number of observations conducted.

Exhibit I9. Per-observation costs by observation tool, observation modality, and scenario

	Accountability				Coaching			
	Year 1	Year 2	Year 3	Three Year Average	Year 1	Year 2	Year 3	Three Year Average
CLASS								
Local								
CLASS live	$358.55	$199.77	$199.77	$252.70	$282.78	$176.85	$176.85	$212.16
CLASS video	$604.04	$206.35	$206.35	$338.91	$361.90	$149.92	$149.92	$220.58
Distant								
CLASS live	$571.05	$412.27	$412.27	$465.20	$518.09	$412.16	$412.16	$447.47
CLASS video	$678.41	$280.72	$280.72	$413.28	$361.90	$149.92	$149.92	$220.58
90% local & 10% distant								
CLASS live	$379.80	$221.02	$221.02	$273.95	$306.31	$200.38	$200.38	$235.69
CLASS video	$611.47	$213.79	$213.79	$346.35	$361.90	$149.92	$149.92	$220.58
ECERS-3								
Local								
ECERS-3 live	$567.16	$257.39	$257.39	$360.65	$463.91	$257.39	$257.39	$326.23
ECERS-3 video	$822.15	$283.07	$283.07	$462.76	$520.22	$207.65	$207.65	$311.84
Distant								
ECERS-3 live	$779.66	$469.89	$469.89	$573.15	$676.41	$469.89	$469.89	$538.73
ECERS-3 video	$896.52	$357.45	$357.45	$537.14	$520.22	$207.65	$207.65	$311.84
90% local & 10% distant								
ECERS-3 live	$588.41	$278.64	$278.64	$381.90	$485.16	$278.64	$278.64	$347.48
ECERS-3 video	$829.58	$290.51	$290.51	$470.20	$520.22	$207.65	$207.65	$311.84

Gates Foundation SRI

SRI

SRI Education, a division of SRI, is helping federal and state agencies, school districts, major foundations, nonprofit organizations, and international and commercial clients tackle some of the most complex issues in education to help students succeed. Our mission is **to reduce barriers and optimize outcomes for all children, youth, and families.** We do this by conducting high-quality research, supporting use of data and evidence, helping to strengthen state and local systems, and developing tools that improve teaching and accelerate and deepen learning. Our work covers a range of topics, including early learning and development, student behavior and well-being, teaching quality, digital learning, STEM and computer science, literacy and language arts, and college and career pathways.

SRI is a nonprofit research institute whose innovations have created new industries, extraordinary marketplace value, and lasting benefits to society.

Silicon Valley
(SRI Headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
+1.650.859.2000
education@sri.com

Washington, D.C.
1100 Wilson Blvd., Suite 2700
Arlington, VA 22209
+1.703.524.2053

www.sri.com/education-learning/

STAY CONNECTED