ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 22-557

Navigating Remote Delivery of Assessments for Head Start Children During the COVID-19 Pandemic

Leiah Groom-Thomas
Brown University

Monica Lee
Brown University

Cate Smith Todd
Acelero Learning

Kathleen Lynch
University of Connecticut

Susanna Loeb
Brown University

Scott McConnell
University of Minnesota

Lydia Carlis
Acelero Learning

Many preschool agencies nationwide continue to experience closures and/or conversions to virtual or hybrid instruction due to the ongoing COVID-19 pandemic. Despite the importance of understanding young children's learning and development during the COVID emergency, limited knowledge exists on adaptable practices of assessing young children during the pandemic. We detail practices used to assess learning in 336 Head Start children across four states during three different time periods in the 2020-21 school year, using adaptation of traditionally in-person assessments of early numeracy, early literacy, and executive functioning. In doing so, we distill early lessons for the field from the application of a novel, virtual assessment method with the early childhood population. The paper describes adaptations of assessment administration for virtual implementation and incorporation of feedback into continued virtual delivery of assessments. Applications and limitations in broader contexts are discussed.

VERSION: April 2022

Suggested citation: Groom-Thomas, Leiah, Monica Lee, Cate Smith Todd, Kathleen Lynch, Susanna Loeb, Scott McConnell, and Lydia Carlis. (2022). Navigating Remote Delivery of Assessments for Head Start Children During the COVID-19 Pandemic. (EdWorkingPaper: 22-557). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/bfcy-5f07

Navigating Remote Delivery of Assessments for Head Start Children During the COVID-19 Pandemic

Leiah J. G. Thomas[1], Monica G. Lee[1], Cate Smith Todd[2], Kathleen Lynch[3], Susanna Loeb[1], Scott McConnell[4], & Lydia Carlis[2]

[1]Brown University

[2]Acelero Learning

[3]University of Connecticut

[4]University of Minnesota and Renaissance Learning

Author Note

Leiah J.G. Thomas https://orcid.org/0000-0001-9988-3285

Monica G. Lee

Cate Smith Todd

Kathleen Lynch https://orcid.org/0000-0001-6344-7430

Susanna Loeb

Scott R. McConnell https://orcid.org/0000-0003-0897-9236

Lydia J. Carlis

Scott McConnell and others developed assessment tools and related resources known as *Individual Growth & Development Indicators,* described here. This intellectual property has

been licensed by the University of Minnesota to Renaissance Learning, and McConnell, his colleagues, and University have equity and/or royalty interests in Renaissance Learning. These relationships have been reviewed and are being managed by the University of Minnesota in accordance with its conflict of interest policies.

The authors thank Sarah Barton, Stacy Calabretta, Demetra Kalogrides, Jennifer Lin, and Camille Maddox for their data assistance. The authors acknowledge the contributions of Drs. Alisha Wackerle-Hollman and Erin Lease of IGDILab at the University of Minnesota, and Heidi Lund, Sara Salewske, Hailey Lucas, and colleagues at Renaissance Learning for their advice, guidance, and assistance in designing procedures for remote administration of Individual Growth and Development Indicators.

Correspondence concerning this article should be addressed to Leiah J.G. Thomas, Annenberg Institute, Brown University, Providence, RI 02912. Email: leiah_groom-thomas@brown.edu.

Abstract

Many preschool agencies nationwide continue to experience closures and/or conversions to virtual or hybrid instruction due to the ongoing COVID-19 pandemic. Despite the importance of understanding young children's learning and development during the COVID emergency, limited knowledge exists on adaptable practices of assessing young children during the pandemic. We detail practices used to assess learning in 336 Head Start children across four states during three different time periods in the 2020-21 school year, using adaptation of traditionally in-person assessments of early numeracy, early literacy, and executive functioning. In doing so, we distill early lessons for the field from the application of a novel, virtual assessment method with the early childhood population. The paper describes adaptations of assessment administration for virtual implementation and incorporation of feedback into continued virtual delivery of assessments. Applications and limitations in broader contexts are discussed.

The COVID-19 pandemic poses an unprecedented challenge for tracking early skill development among preschool-aged children. At the onset of the pandemic, preschool and kindergarten enrollment declined substantially, decreasing 50% relative to pre-pandemic levels (Delap et al., 2020; Weiland et al., 2021), and 60% of child care centers closed nationwide due to operating issues (Procare Solutions, 2021). Nearly two years into the pandemic, resurging virus outbreaks and subsequent closures continued to delay important opportunities for young children nationwide. Consequences of the pandemic in early education likely mirror effects on academic learning loss for K-12 children (Kuhfeld et al., 2020; Wyse et al., 2020): Preschoolers are at risk of falling behind in crucial foundations for academic success such as early literacy and early mathematics, in addition to core executive-functioning skills (Meloy et al., 2019). Continued delays in learning opportunities further compound the issue of irregular access to quality educational opportunities and exacerbate pre-existing disparities, leading to increasing concerns about how preschool-aged children are faring (Dorn et al., 2020).

Under these circumstances, measuring young children's progress in early skill development is critical. In early learning programs such as Head Start, formal assessments of literacy, numeracy, and executive function skills typically play a pivotal role in identifying children at risk for developmental and learning delays and catalyze planning for learning supports, including supplemental intervention services. Because early skills are foundational for kindergarten readiness and are linked to children's longer-run academic performance and educational attainment outcomes (Bartik, 2014; Karoly & Auger, 2016; Meloy et al., 2019), identifying lags in skills early on is especially important. Snapshots of children's skill

development during the pandemic are also essential because they can serve as critical guideposts for planning post-pandemic early behavioral and/or academic interventions, especially for children who struggle as a result of trauma or other COVID-related issues.

Additionally, children from historically marginalized populations may face additional challenges that are currently veiled by the pandemic. Race and income disparities in access to high-quality preschool precede the pandemic: Even before stay-at-home orders were in place, low-income families were less likely than higher-income families to enroll their children in preschool (Koball & Jiang, 2018) and Black and Hispanic children were less likely to access high-quality preschool programs relative to their White peers (Rothwell, 2016). These inequalities persisted during the pandemic: Fewer low-income families enrolled children in preschool after the onset of the pandemic compared to higher-income families, and families of color were more likely to be enrolled in virtual learning or to experience more disruptions in their young child's learning compared to White families (Weiland et al., 2021). These challenges, coupled with pre-existing racial and income-based gaps in early education, highlight the need for a greater understanding of child development within the context of the pandemic.

With the ever-growing need to measure skills among children during widespread school closures, researchers are still developing ways to conduct virtual implementation of traditionally in-person assessments and effective approaches to assess children in these conditions. Virtual assessment is not impossible. At the early onset of the pandemic, for instance, one prominent assessment organization successfully administered over 400,000 tests to elementary and secondary school children virtually, with plans to scale virtual testing across broader settings in the following academic year (Northwest Evaluation Association, 2020). However, more research is needed to explore virtual assessments for young children who would typically be assessed in

person in their preschool setting and to better understand implementation challenges that may happen in virtual settings (Farmer et al., 2021; Stifel et al., 2020). Similar gaps exist in policy: While the Office of Head Start recommends that programs should assess young children despite the pandemic (U.S. Department of Health & Human Services, 2021), at the time of this writing, it has not yet offered clear guidelines for how to approach direct or observation-based assessments in a virtual learning setting for this population.

These factors – the need to monitor preschool children's development, the challenges of doing that without face-to-face interaction, and special concerns for assessment and intervention for children from historically marginalized populations and in under-resourced communities – implied a need for rapid adaptation of practices when schools began closing in Spring 2020. This paper provides insights learned from an all-virtual implementation of typically in-person assessments for preschool children enrolled in Head Start programs. We conducted one-on-one virtual assessments with 336 preschool children previously enrolled in in-person (i.e., attending in brick-and-mortar centers) and virtual (i.e., fully virtual; attending from home) classroom models in 44 Head Start centers across four states on domains of print knowledge, early numeracy, and executive functioning. To minimize in-person interactions in line with state and federal health guidelines and to maximize consistency of implementation across assessment locations, we conducted the assessments completely virtually for both in-person and virtual program models.

Contributions of This Paper

Our paper adds two important contributions to research on young learners during the COVID-19 pandemic: 1) detailed information on virtually implementing traditionally in-person assessments within the preschool population, and 2) the ability to observe this process across

three time points during the 2020-21 school year in multiple states/contexts. While several studies have examined the Head Start population throughout the pandemic on a national or multi-state level, they focus on market perspectives (Ali et al., 2021) or children's safety protocols (Coronado et al., 2020) rather than the actual process of measuring child learning during the pandemic. For example, two studies surveyed child care providers to investigate experiences during the pandemic for dual language learners, but results are limited to information about services offered and enrollment, and direct assessments of children's learning are lacking (Quick et al., 2020; White et al., 2021). The current study details the process of collecting such data across the 2020-21 school year. We collect large-scale data on early literacy, numeracy, and executive functioning skills among Head Start preschool children during the COVID-19 pandemic, and conduct large-scale virtual assessments of preschoolers' learning in these domains. Here, we share insights garnered from implementing a fully-virtual data collection model. We address the question: What lessons for the field can be gleaned from the application of a novel, virtual assessment method with the early childhood population? This model will likely be of interest to researchers and practitioners who seek to measure young children's learning during the immediate global health emergency and has implications for a broader range of contexts and settings where in-person assessment is not feasible in the future.

The primary objective of this paper is to describe and reflect upon the process of virtual administration of early education assessments. First, we describe in detail the strategies used to prepare for virtual assessment administration and responsibilities of various stakeholders, providing a process model that can be emulated at an equivalent or larger scale across other early education agencies. Second, we summarize lessons learned from virtual assessment delivery. While many practices were preserved from conventional assessment administration periods, we

describe the two types of changes that were made – changes decided prior to virtual adaptation and changes adopted after launch – to the assessment administration process to better adapt to virtual delivery. Lastly, we discuss limitations to our work and applications to broader settings.

Method

We conducted the virtual assessments three times during the 2020-21 school year (Fall, Spring, and Winter) as part of a broader research-practice partnership between a practitioner organization that operates and/or partners with local Head Start programs (Early Learning Inc[1]) and a private research university on the east coast. Research-practice partnerships are "long-term, mutualistic collaborations between practitioners and researchers that are intentionally organized to investigate problems of practice and solutions for improving [partnering organizations'] outcomes" (Coburn et al., 2013, p. 2). The partnership was an in-depth collaboration responding to the pandemic in order to address the needs of the practitioner organization for information on young children's learning. As a result of the COVID-19 pandemic, we shared a mutual interest in exploring the progress of skill development in virtual models of teaching and learning, which led to the virtual implementation described below.

Sample

Virtual assessments were conducted with a sample of 336 enrolled Head Start children across the four locations of Early Learning Inc: Nevada, New Jersey, Pennsylvania, and Wisconsin. The partnering organization separated enrolled Head Start children into two groups by program option (those enrolled in either the in-person or virtual learning model). Next, the partnering organization used the random function on Excel to select children from each of the four locations such that roughly half the sample consisted of children enrolled in the in-person

[1] Early Learning Inc is a pseudonym for the name of the partnering organization.

learning model and the rest in the virtual learning model. Table 1 presents average demographic characteristics for all children assessed and those enrolled in each learning model. The majority of children enrolled in both in-person and virtual learning models were four years old. Nearly half of the sample reported as Hispanic, and about 28 percent spoke Spanish in the home; these proportions do not vary substantively by learning model. One notable difference between the children in the in-person and virtual learning models is that a higher proportion of virtual learning children (76%) were returning to an Early Learning Inc. Head Start center relative to those in in-person learning models (67%). Additionally, children enrolled in the virtual learning model had fewer sociodemographic factors associated with risk and were less likely to have an active Individualized Education Program or Individualized Family Service Program (IEP/IFSP) compared to children in the in-person learning model.

We initially aimed to assess up to 30% (i.e., 600) of all children across the four locations, with half of the target number attending in-person and half in virtual learning models, although the number of collected assessments was notably fewer than what we aimed for (see Table 2 for a detailed breakdown). This was due to several reasons. First, repeated closures of centers and frequent child absences induced by local COVID-19 outbreaks created added barriers for scheduling virtual assessment appointments, reducing participation rates. Additionally, the number of virtual assessments declined throughout the course of the school year due to opt-outs and withdrawal from Head Start, which are issues typically observed in other research settings. Caregivers of enrolled children (who opted into assessments during enrollment) had the option to opt-out at any time during the school year, and continued opt-outs led to a decrease in assessments administered throughout the year. Additionally, if children participating in virtual

assessments withdrew from Early Learning Inc. centers, they were no longer assessed in subsequent assessment periods as they were no longer participating in the Head Start program.

As shown in Table 2, more children from the virtual learning model withdrew from assessments compared to those from the in-person model; this difference is statistically significant. Additionally, 52 percent of children originally assigned to the virtual learning model who failed to participate by spring had withdrawn enrollment entirely from the Head Start center where they were based, compared to 65 percent of their counterparts in the in-person learning model. This difference is not statistically significant, however, suggesting that disproportionate withdrawal rates from Head Start were not the driving force behind the difference in assessment collection rate between the in-person and virtual models.

Measures

The assessment battery consisted of standardized measures of print knowledge, early numeracy, and executive function (described below). We selected these measures for three reasons. First, we examined domains of print knowledge, early numeracy, and executive functioning because these skills map onto important constructs foregrounded in Head Start's early learning outcomes framework for school readiness (U.S. Department of Health and Human Services, 2015). Second, we prioritized measures that mapped onto skills assessed in previous years and for which the partnering organization had digital materials ready for use. Third, we attempted to keep virtual administration for the complete battery under 30 minutes (5-7 minutes per assessment) to maintain engagement and participation when working with preschool-aged children in a virtual setting.

We adapted the measures for virtual administration and scoring (detailed below) based on the guidance of assessment vendors. An account representative for each vendor served as the

point of contact for Early Learning, Inc. The research coordinator met individually with account representatives for any feedback on ways to adapt measures for virtual administration (e.g., suggested order of subtests to keep children engaged virtually). These conversations were informal, occurred after purchasing assessments, and the research team was not pressured in any way to adhere to the suggestions made by the vendors themselves. We accessed all virtual materials through digital access codes or password-protected web-based user accounts and created digitized versions of physical materials (i.e., score forms). Table 3 provides a description of the constructs and publisher-reported reliability statistics for the direct child measures. In all cases, we retained existing guidelines for the assessments in the areas of basal/ceiling requirements, practice items, the timing of the assessments, and administrative prompts.

Print Knowledge

We assessed children's understanding of print concepts (e.g., letter and word recognition) using the Print Knowledge subtest of the Test of Preschool Early Literacy (TOPEL; Lonigan et al., 2007; $\alpha = .97$). The Print Knowledge subtest consists of 36 items compiled in a flipbook, for which children are presented with four response options for each item—a mixture of words, letters, and pictures—and asked to select the response that best answers the assessor's prompt (i.e., "Which word can you read? Which is the letter /k/?"). We summed all items for a total score and converted the sum to a standard score for same-age peer comparisons.

Virtual adaptations for the Print Knowledge subtest included using digital versions of the flipbook and score forms. The assessment vendor provided digital flipbook PDFs and the research team converted paper scoring forms into Google Forms. Assessors used the "screen share" feature on Zoom to show digital flipbook pages as they read the respective prompts and recorded child responses in real-time using a Google Form on an iPad. The assessment took 7-10

minutes to administer, which is identical to traditional assessment administration and reflected no alterations in the typical publisher-reported assessment duration.

Early Numeracy

We assessed children's ability to identify and compare numbers using the Oral Counting, Number Naming, and Quantity Comparison subtests of the Individual Growth and Development Indicators of Early Numeracy (IGDI-ENs; Hojnoski & Floyd, 2012; ICCs ranging from .71-.88 and concurrent validity ranges .60-.75 with similar numeracy measures, see subtest description below). Oral Counting asks the child to count aloud, beginning at one and continuing until an error in correct sequencing is made or until one minute elapses, with the child's score being the highest number named in correct sequence either before the first error or at the end of a full minute. Number Naming asks the child to provide the names for the numerals 1 to 20, presented individually and in random order, and the score equals the number of items correctly identified. Quantity Comparison is a subitization task. The child is presented with a series of pictures of two dice faces representing different values from 1 to 6; the child is asked to identify the image "with more." The child's score is the number of correctly identified images in one minute.

Virtual adaptations for the three subtests included using digital versions of placards and score forms. We did not include a fourth and final subtest of the Early Numeracy assessment based on the assessment vendor's suggestion that we would not be able to adapt it for virtual delivery, as there was no way to display this subtest on the computer screen via Zoom. Akin to the administration of TOPEL-PK, assessors used the "screen share" feature on Zoom for each subtest to show digital flipbook pages as they read the respective prompts and recorded child responses in real-time using a Google Form on an iPad. Each subtest was timed at 60 seconds for

a total administration time of roughly 3-5 minutes per subtest, including practice items. We made no alterations in the length of subtests.

Executive Functioning

We assessed children's working memory, inhibitory control, and cognitive flexibility using the Minnesota Executive Function Scale (MEFS; Carlson & Zelazo, 2014; ICC = .93). MEFS is a standardized, child-friendly application designed for children as young as two years old, administered individually on a tablet. The assessment requires children to follow assessor prompts to select one of two given options (i.e., "We are going to play the color game. This is a blue box and this is a green box. All green bicycles go in the green box, all blue bicycles go in the blue box. Where does *this* one go?"). The MEFS application is adaptive (i.e., the starting point depends on child's age, progresses based on child's responses, and increases in difficulty until a child consecutively answers ten items incorrectly). Items were scored, automatically computed, and converted to a standard score for same-age peer comparisons following testing.

Virtual adaptations of MEFS were similar to those for IGDI subtests: assessors used the "screen share" feature on Zoom to show the child each item while reading aloud the respective prompts. Assessors recorded child responses in real-time on the iPad as if the assessor were the child selecting the option. Total administration time ranged from 2-7 minutes. Again, we made no alteration to the length of the assessment.

Procedure

In collaboration with the partnering organization, the research team decided on three assessment periods during the 2020-21 school year. The assessment periods were spaced apart in three-month intervals (i.e., early November 2020, late January 2021, late April 2021). Between each of the assessment windows, the research coordinator (see below) trained new assessors and

staff from the partnering organization shipped tablets to any caregivers who reported that they lacked tech devices. Notably, assessors and staff from the partnering organization participated in implementation meetings after the initial assessment period as a means to collect and apply feedback for future periods. Figure 1 outlines steps that the assessment team took throughout all three assessment periods, split broadly into categories of the research coordinator, virtual assessors, and central staff at each participating Head Start location.

Assessment Team Roles and Tasks

Research Coordinator

The research coordinator was responsible for leading virtual assessment delivery, which included managing data collection and reporting status updates to the broader team in the research organization. Prior to the start of the assessment window, the research coordinator led preparations for data collection (e.g., allocating technological devices, training assessors), distributed an informational video about staff roles during assessments, and held informational sessions to answer questions and help with scheduling. In addition, the research coordinator engaged Spanish and Arabic-speaking assessors or hired translators to assist with conducting assessments for children whose primary language was not English. Once assessments were taking place, the research coordinator communicated step-by-step instructions on how to manage an assessment, provided progress updates, worked with assessors to mitigate technology issues, noted atypical sample cases (e.g., documenting classroom closures due to COVID), managed data collection and quality assurance measures, and held weekly debriefing meetings with assessors and staff. Upon completion of each assessment window, the research coordinator cleaned and reviewed assessment data in preparation for the next assessment period.

Assessors

The assessment team (n = 10) completed all one-on-one assessments via Zoom. Assessors were individual contractors (i.e., not a part of Head Start center staff) who had: (1) prior data collection experience, particularly in field assessments and primary data collection; (2) prior experience working with young children; and (3) flexible availability to conduct virtual assessments during the three pre-determined assessment windows.

Assessors completed training sessions via Zoom on administration and scoring of assessments in a whole-group setting with the research coordinator. Similar to assessment training protocol for in-person delivery, this assessment training required assessors to review technical and administration manuals and associated scoring procedures. Training sessions included basal/ceiling reminders, an overview of administrative prompts, and dry run rehearsals of administration and scoring of assessments using Zoom in conjunction with other required technology (e.g., digital flips books and online score forms). Prior to the start of the assessment window, assessors were required to complete a 10-item administration fidelity checklist across all subtests with the research coordinator to ensure adherence to publisher administration requirements.

Assessors also completed a training session via Zoom on guidelines and considerations for a virtual assessment. This training aimed to familiarize assessors with the following guidelines specific to conducting a virtual assessment:

- An explicit verbal statement to the child and caregiver regarding the purpose and scope of the assessment.
- Strategies on building a positive and supportive environment during the assessment, given that virtual assessors are people with whom the child is unfamiliar. This includes, but is not limited to, engaging in conversation before administering the battery (“I love

the color of your shirt! Is blue your favorite color?"); using encouraging words and phrases ("I appreciate your focus today; keep up the great work!") and using kinesthetic movements to refocus in between assessments ("Reach up to the sky as high as you can and let's shake out or bodies for the last part of this game today").

- Strategies for troubleshooting various technology problems, such as mitigating issues logging into Zoom, accessing digital materials, and providing sufficient wait time for children needing redirection or breaks between assessments.
- Awareness of demands associated with assessing children in their homes, including, but not limited to, requesting that caregivers find a location with as few distractions as possible, making every effort to follow up with caregivers to schedule missed or incomplete assessments, being flexible during assessments, and addressing any issues or concerns with full transparency and respect for the caregiver and the child.

After completing the virtual training and reliability checks with the research coordinator, assessors were allowed to begin testing. During assessments, assessors administered child-directed measures while scoring via Google Forms in real-time. We required a classroom teacher to be present to support child engagement and technology during virtual assessments for children in the in-person learning model; for virtual assessments with children in the virtual learning model, we required a virtual teacher to be present in the Zoom meeting room and a caregiver to be present in person to help manage the child's engagement and technology. Virtual teachers are teachers who regularly taught in-person in Head Start classrooms prior to the pandemic, but were reassigned to teach children in the home-based model during the 2020-21 SY. Virtual teachers were present during assessments because children and parents were familiar with them and it

helped maintain a calm assessing space. Also, many assessments took place during the child's reoccurring weekly Zoom meeting time.

Adults other than the assessor who were present during the Zoom meeting were asked to only provide neutral feedback to responses regardless of accuracy ("thank you for working hard" or "try your best") and cue the assessor to where the child pointed for each response (left, right). For instance, in addition to teachers and caregivers, translators were often present. Translators were trained to provide cues indicating where the child pointed to in their respective language (left, right) and to translate instructions verbatim to the child. They were not allowed to provide hints or rephrase prompts. If a child did not understand the prompt for an item after translation service occurred, assessors continued to the following item in the protocol.

After administering a battery, assessors filled out a live tracking document that included information such as the date of the scheduled assessment, where the assessment took place (i.e., in the classroom or at home), and whether a partial or full battery was completed during the session.

Head Start Central Staff

Central staff included those employed in Head Start centers across the four states in which Early Learning Inc operates, including center directors, virtual teachers, and family advocates (i.e., early childhood specialists who provide services in support of families during virtual learning) responsible for assisting in outreach and engagement efforts. Central staff facilitated scheduling and coordination with in-person teachers and families during data collection, while family advocates coordinated assessments for children in the virtual learning model. Staff also identified families in need of translation services and delivered the informational video to caregivers via email at the onset of each assessment period.

Implementation Meetings

Upon the conclusion of the initial (i.e., fall) virtual assessment period, members of the three groups above (research coordinator, virtual assessors, and Head Start staff) participated in implementation meetings as a means for continuous improvement in subsequent iterations of virtual assessment delivery. During implementation meetings, the research coordinator collected feedback with the understanding that we would apply group feedback in preparation for winter assessments, set to take place shortly after the fall.

The research coordinator worked with Early Learning, Inc. center leaders to comprise a set of questions for participants in each of three groups: virtual assessors, center-based staff, and virtual staff members (such as virtual teachers or family advocates working virtually). We shared questions with participants ahead of time to serve as prompts that would guide our discussions. Each group consisted of 6-8 individuals who volunteered to participate when asked by center directors of the four metropolitan areas in which Early Learning Inc operates. A total of 10 virtual assessors, 11 center-based staff, 7 virtual teachers, and 6 family advocates participated, representing approximately 20% of all Head Start center staff who directly supported virtual assessments during the 2020-21 school year. Each implementation meeting session lasted one hour and was conducted via Zoom within two weeks after initial assessment delivery in the fall. Caregivers were not invited to participate in implementation meetings in effort to reduce additional family burden during the pandemic.

Results

Although conducting virtual assessments was a novel approach, many of the procedures pre-, mid-, and post-assessments stayed largely the same as the ones done for in-person assessments in an attempt to preserve the testing experience and scoring process as much as

possible. There were two types of changes made that are particular to virtual delivery (and not necessarily to the pandemic): One is a series of pre-determined changes that intentionally prepared for the digital administration of assessments (see Table 3), given that direct assessments were originally created for in-person delivery; the other were changes adopted in the middle of the first assessment window as a result of unexpected challenges, or due to feedback from assessment team members participating in implementation meetings, after the end of the first assessment window (discussed below).

Practices Preserved from Traditional Assessments

While administering assessments virtually, training pre-administration, the actual test-taking experience, and scoring largely remained the same as in traditional, in-person assessments, as the goal was to simulate traditional assessments as best as possible.

Regardless of the context, assessors were trained to maintain the same best practices that they would have otherwise in direct, in-person assessments of young children. For example, assessors received training on building rapport with the child to establish an environment of trust. The training encouraged assessors to create an effective testing environment by working with the teacher/caregiver to find a quiet, distraction-free space. Assessors knew to use redirection strategies and breaks in between assessments when necessary. All of these strategies were the same ones that would have been used to promote the child's focus and engagement in an in-person setting.

Lastly, we adhered to the same protocol as we would when conducting and scoring traditional assessments. During the test-taking period, we adhered to the original guidelines provided by each assessment vendor to the extent possible. Within each subtest administered, we retained the same items from the original assessment and administered the assessments one-on-

one. Assessors followed administrative prompts, feedback, basal/ceilings per manual requirements, and real-time scoring procedures. In the rare case that the integrity of the data was compromised (i.e., a caregiver provided corrective feedback; unsure of child's true response), we documented this and discussed it before making a decision on whether or not to count items as correct.

Changes to Assessment Administration for Virtual Adaptation

One key change made for virtual adaptation of assessments was the conversion of all physical materials and resources to digital versions for online access. For example, assessors used publicly-available digital testing materials such as PDF placards and flipbooks provided by assessment vendors during Zoom meetings via the screenshare feature. For scoring, we replaced paper scoring sheets with digital forms and sheets on Google for real-time, item-level scoring. For a complete list of assessment adaptations for virtual administration, refer to Table 3.

While the training, testing, and scoring largely stayed the same, the back-end management of assessments changed substantially to adhere to virtual settings. Prior to the assessment window, we established new ways of using technology, tracking data and progress, and methods of communication in order to effectively manage virtual assessments. For example, we created an online scheduling portal such that teachers and caregivers could schedule assessments asynchronously and independently of research staff; 100% of families were required to use the online scheduling portal to make their assessment appointments, as all assessments were conducted virtually. Confirmations and appointment reminders were sent via email. A high number of no-shows to testing appointments among children in the virtual learning model posed a challenge. However, virtual assessment of children at home also afforded flexibility in assessment administration – for example, assessors were able to meet with children at various

hours of the day, including evenings and weekends, and Early Learning, Inc. was able to hire assessors who did not reside in the local area.

Changes Adopted After Launch

Navigating virtual assessments with young children during a pandemic presented several challenges. We encountered challenges particular to scheduling and participation rates while data collection activities were taking place and made three significant modifications to mitigate their impact on the assessment progress during the school year.

First, we extended each data collection timeframe by one week, increasing the total duration of assessments from three to four weeks due to the difficulty of scheduling children in the virtual learning model. At the halfway point of the assessment window, less than 50% of children were scheduled, and even fewer were assessed. This was in part due to unfamiliarity with the online scheduling portal, although spikes in COVID-19 and related center closures and child absences further compounded the issue.

Additionally, we decided to engage virtual teachers to help schedule and manage assessments for children in virtual learning models who had considerably lower participation rates from the beginning compared to their in-person counterparts. During the fall assessment period, only 25% of the children and caregivers who scheduled an appointment attended it at their designated time. This was likely because caregivers were overwhelmed with other duties or life circumstances, which may have been exacerbated by the pandemic. We decided to leverage the weekly check-in time that virtual teachers had scheduled with each child in their class to encourage participation in assessments. This improved appointment scheduling rates by nearly 30% and attendance by 56% for children in the virtual learning model throughout the remainder of fall assessments as well as winter and spring assessment periods.

In a similar vein, we established clearer guidelines for rescheduling appointments after unexpectedly encountering a large proportion of no-shows upon the start of the fall assessment window, particularly for children in the virtual learning model. As a result, we established guidelines in order to increase participation rates. For example, we asked assessors to wait for five minutes with a no-show appointment before contacting the teacher/caregiver to confirm attendance. After 20 minutes, assessors were instructed to sign off of Zoom and contact the teacher or caregiver to reschedule the assessment via email, text message, or phone call. After establishing guidelines for communications, teachers/caregivers were more likely to reschedule a missed assessment or notify the research team ahead of time about cancellations. The first round of changes occurred in November 2020.

Changes Adopted After Implementation Meetings

In order to gather recommendations for improvement of assessment procedures in future rounds, the research coordinator held implementation meetings with virtual teachers, assessors, and center staff after the first round of testing, during which members of the Early Learning, Inc. staff took meeting notes. After the implementation meetings were complete, the research coordinator reviewed the meeting notes with the purpose of identifying feedback on what implementation components worked during the first round of testing and suggestions for improving implementation during the next round of testing. The research coordinator compiled recommendations and implemented them in subsequent assessment periods. These recommendations are summarized in the section below.

First, participants in the implementation meetings made the recommendation noted above of leveraging virtual teachers to improve communications in subsequent assessment rounds. We also used pre-existing text messaging and automated phone calling systems used by Early

Learning Inc's Head Start centers to remind caregivers of appointments. Additionally, we adjusted the delivery of information about the assessments in response to feedback from virtual assessors and teachers who reported that caregivers were confused about the purpose of the assessment and/or what their role was during the duration of the assessment.

We also made improvements in mid-assessment practices to minimize corrective feedback and other distractions. Five of ten virtual assessors reported that caregivers occasionally provided feedback or assisted their child during administration. Given children's young age coupled with the fact that assessments were being conducted in their homes, it was challenging to prevent all aspects of caregiver involvement mid-assessment. However, beginning with the winter assessment, we requested that caregivers stand *behind* their child for the duration of the assessment to minimize the likelihood of providing corrective feedback or assistance with the assessment. Additionally, participants had reported that caregivers had challenges alleviating distractions in the home during the assessments and that strategies learned in training did not mitigate these issues. Examples of such distractions included siblings or other children playing in the background or other background noises that detracted from the testing experience. To address this issue, in subsequent rounds of testing, we added clarifying language to the informational video we distributed to families highlighting the issue of testing in a quiet space (e.g., "please find a quiet location with as few distractions as possible") and assessors reiterated the importance of a distraction-free environment at the start of each assessment. Lastly, in order to ensure the maintenance of privacy, we advised caregivers to ask their child to be seated against a blank wall during the assessment.[2] Lastly, assessors prioritized the first five minutes of each assessment to be dedicated to technology- or internet-related issues. It is important to note that all adaptations

[2] While the blurred background feature on Zoom was not yet available at the time of the assessments, this could be a useful tool for the future.

for virtual delivery resulted from group feedback or the decision to administer virtually, as direct assessments were originally created for in-person delivery. The second round of changes were set in effect beginning the winter assessments in February 2021.

Discussion

This paper fills gaps in knowledge pertaining to how researchers and practitioners can partner to deliver virtual assessments to inform early childhood education. We highlight the underlying processes, challenges, and improvements made while conducting a fully-virtual delivery of assessments with preschool children during the 2020-21 school year. We conducted implementation meetings midyear to gain feedback and used these recommendations to improve the assessment process in future rounds. Given the lack of precedent on assessing Head Start children virtually, we encountered several challenges related to the logistics of scheduling assessments and effectively communicating to caregivers. Below, we discuss these limitations and considerations for future research.

Implications

This work demonstrates a virtual approach to conducting a program evaluation for children in Head Start. The implications of this work are immediately relevant within the context of the COVID-19 pandemic but likely also extend beyond the pandemic into future applications in research and practice. There are considerable advantages to employing virtual assessment strategies. For example, virtual assessments are cost-effective and scalable. In our experience, we found virtual assessors to be more hirable in the virtual context, eliminating the issue of having an unequal distribution of assessors by location or the potential risk of infections or quarantining. Virtual assessments also reduce costs such as traveling for in-person appointments, so these areas of the budget can then be re-invested in other items, such as technological devices.

Additionally, virtual assessments have a potential scalability advantage over in-person ones, given that the work of training assessors and the general process of virtual assessments can become a centralized, online process. Given the persistence of a digital divide in the United States, a clear, practical barrier to the conduct of virtual assessments with low-income populations is access to technological devices and stable Internet. Early Learning Inc. directly provided tablets and Internet hotspot devices to families who indicated that they needed them. We urge future researchers considering virtual assessments to prioritize a needs assessment for and provision of technological devices and support at the onset of their studies.

Limitations and Future Research

There are several limitations to our work. First, transitioning a traditionally in-person assessment to an online format yields scores that must be interpreted with caution. For example, while assessors attempted to ameliorate circumstances where caregivers provided corrective feedback during an assessment, we do not have a comprehensive understanding of how often or the extent to which this feedback influenced children's responses. Additionally, we do not have sufficient data to indicate whether assessment scores observed during virtual delivery of assessments would be similar to those observed during traditional in-person assessment periods because different assessments (that map on to the same learning constructs in this study) were used by Early Learning Inc in previous years. This especially applies to specific subgroups of children – such as children with disabilities – who may have performed differently purely as a result of the digital nature of administration. Lastly, we do not have item-level data available to calculate sample-specific reliability estimates for each assessment. However, initial publisher-reported reliability estimates of the measures used are relatively high when administered in person, as mentioned previously (see Zelazo & Carlson, 2017; Floyd et al., 2006; Wilson &

Lonigan, 2010). Given that we trained for, administered, and scored the assessments in the exact same way as we would have with in-person administration, we anticipate that the assessments should perform similarly in an online format as they typically do in person, although it serves as a useful avenue for future research on reliability when in-person assessments are administered virtually.

Second, the nature of this paper is to outline the steps to and experience with implementing a novel, virtual delivery of an assessment. We do not examine the impacts of specific early learning program models on children's outcomes. Such evaluations are an important step for future research, and we aim to quantify learning gains observed across the program models in which children were enrolled during the COVID pandemic-affected program years in subsequent studies. In addition, we did not collect detailed information about implementation meetings, create transcriptions of meetings, code said transcripts, or consider aspects such as data saturation or data triangulation. We conducted these meetings to meet needs outside of conventional qualitative research: to immediately troubleshoot challenges in implementation in the midst of a quickly-changing environment. Future work that includes the above processes can provide detailed, rich descriptions of qualitative data collection methods and analysis procedures.

Third, in the interest of reducing the number of requests we made of families and caregivers during a peak phase of the pandemic when many were experiencing stressors relating to potential illness, work, and child care disruptions, we did not solicit active involvement from caregivers. This meant that we did not ask for caregivers' participation in the midyear implementation meetings on assessment procedures. More broadly, all communication with caregivers was channeled through Head Start staff for the same purpose, so we do not have

evidence, anecdotal or statistical, to gauge caregiver involvement in or perceptions of assessment delivery (including whether caregivers watched and understood the information video). We plan to gather this information in the future, and we urge future researchers to experiment with creative means of collecting family and caregiver perceptions and input on their experiences with virtual assessments.

Fourth, while we employed numerous strategies for engagement, such as minimizing the assessment battery to a short time interval to maintain attention and providing redirection strategies during assessor training, we did not systematically monitor children's behavior during virtual assessments. Future work is needed to better understand challenges and potential remediations when engaging children during virtual assessments. In addition, although we were able to measure several important dimensions of children's early literacy, numeracy, and executive function that support school readiness, we could not measure every possible construct. For example, print knowledge is only one dimension of early literacy. An important step for future research is to experiment with administering assessments that capture (1) a broader range of early skills, including a variety of skills in early literacy, and (2) strategies to promote and maintain engagement for a diverse group of learners, including supports for children with disabilities.

Conclusion

Our experience in virtual implementation of assessments suggests that, in spite of potential limitations of results and learnings offered here, the urgency and importance of adapting to a rapidly and radically changed educational environment must be considered. At the time of writing this paper, many children have experienced disruptions from continuous in-person schooling for nearly two years. While this has likely been a challenge for children of all

ages, it is particularly important for younger learners; if we fail to find ways to adapt to virtual assessment and intervention, preschool children like those participating here run the risk of losing critical support for the development of foundational skills, and it is possible that this additional loss of opportunity would produce disproportionate negative consequences throughout these children's educational careers. Therefore, adapting our practices to accommodate the changes COVID-19 produced is an essential element of our ongoing commitment to education for all.

References

Ali, U., Herbst, C. M., & Makridis, C. A. (2021). The impact of COVID-19 on the U.S. child care market: Evidence from stay-at-home orders. *Economics of Education Review, 82*. https://doi.org/10.1016/j.econedurev.2021.102094

Bartik, T. J. (2014). From preschool to prosperity: The economic payoff to early childhood education. Kalamazoo, MI: W. E. *Upjohn Institute for Employment Research.* https://doi.org/10.17848/ 9780880994835

Carlson, S. M., & Zelazo, P. D. (2014). *Minnesota executive function scale: Technical report.* St. Paul, MN: Reflection Sciences.

Coburn, C.E., Penuel, W.R., & Geil, K.E. (2013, January). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts.* New York, NY: William T. Grant Foundation.

Coronado, F., Blough, S., Bergeron, D., Proia, K., Sauber-Schatz, E., Beltran, M., Rau, K.T., McMichael, A., Fortin, T., Lackey, M., Rohs, J., Sparrow, T., & Baldwin, G. (2020). Implementing mitigation strategies in early care and education settings for prevention of SARS-CoV-2 transmission — Eight States, September–October 2020. *Morbidity and Mortality Weekly Report, 69*, 1868-1872. doi: http://dx.doi.org/10.15585/mmwr.mm6949e3

Delap, S., Franko, M., Hasan, N., McGee, A., & Thornton, C. (2020). *Impact of COVID-19 on early childhood care and education providers*. Denver, CO: Early Milestones Colorado.

Dorn, E., Hancock, B., Sarakatsannis, J., and Viruleg, E. (2020, June). *COVID-19 and student learning in the United States: The hurt could last a lifetime*. New York, NY: McKinsey.

Farmer, R.L. McGill, R. J., Dombrowski, S. C., Benson, N. F., Smith-Kellen, S., Lockwood, A. B., Powell, S., Pynn, C., & Stinnett, T. A. (2021). Conducting psychoeducational assessments during the COVID-19 crisis: The danger of good intentions. *Contemporary School Psychology*, *25*(1), 27-32. https://doi.org/10.1007/s40688-020-00293-x

Floyd, R. G., Hojnoski, R. L., & Key, J. (2006). Preliminary evidence of technical adequacy of the preschool numeracy indicators. *School Psychology Review*, *35*(627–644). https://doi.org/10.1080/02796015.2006.12087967

Hojnoski, R., & Floyd, R. G. (2012). *Individual growth and development indicators – early numeracy*. Bloomington MN: Early Learning Labs/Renaissance Learning.

Karoly & Auger (2016). *Informing investments in preschool quality and access in Cincinnati: Evidence of impacts and economic returns from national, state, and local preschool programs.* Santa Monica, CA: RAND Corporation.

Koball, H., & Jiang, Y. (2018). *Basic facts about low-income children: Children under 9 years, 2016.* New York, NY: National Center for Children in Poverty, Columbia University Mailman School of Public Health.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher, 49*(8), 549565. doi: 10.3102/0013189X20965918

Lonigan, C. J., Wagner, R. K., & Torgesen, J. K. (2007). *TOPEL: Test of preschool early literacy*. Austin, TX: Pro-Ed.

Meloy, B., Gardner, M., & Darling-Hammond, L. (2019*). Untangling the evidence on preschool effectiveness: Insights for policymakers.* Palo Alto, CA: Learning Policy Institute.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the national early literacy panel.* Washington, D.C.: National Institute for Literacy.

Northwest Evaluation Association. (2020, June). *As schools restart, NWEA is ready for remote testing.* Portland, OR: NWEA.

Procare Solutions. (2021, February). *Tracking the impact of COVID-19 on the child care industry.* Denver, CO: Procare Software, LLC.

Quick, H., White, L. J., Brodziak de los Reyes, I., Bergey, R., & Carbuccia-Abbott, M. (2020*). A system in jeopardy: California's early learning system and its dual language learners during the COVID-19 pandemic.* Washington DC: American Institutes for Research. Retrieved from: https://www.air.org/sites/default/files/A-System-in- Jeopardy-COVID-DLLs-September-2020.pdf.

Rothwell, J. (2016, June 29). *Black and hispanic kids get lower quality pre-K.* Washington D.C.: Brookings Institution.

Stifel, S. W., Feinberg, D. K., Zhang, Y., Chan, M., & Wagle, R. (2020). Assessment during the COVID-19 pandemic: Ethical, legal, and safety considerations moving forward. *School Psychology Review*, *49*, 438-452. https://doi.org/10.1080/2372966X.2020.1844549.

U.S. Department of Health and Human Services, Administration for Children & Families, Office of Head Start (OHS). (2015). Head Start early learning outcomes framework: Ages birth to five. Retrieved from https://eclkc.ohs.acf.hhs.gov/interactive-head-start-early-learning-outcomes-framework-ages-birth-five

U.S. Department of Health and Human Services, Administration for Children & Families, Office of Head Start (OHS). (2020). *OHS COVID-19 Updates*. Retrieved from https://eclkc.ohs.acf.hhs.gov/about-us/coronavirus/ohs-covid-19-updates

Weiland, C., Greenberg, E., Bassok, D., Markowitz, A., Rosada, P. G., Luetmer, G., Abenavoli, R., Gomez, C., Johnson, A., Jones-Harden, B., Maier, M., McCormick, M., Morris, P., Nores, M., Phillips, D., & Snow, C. (2021). *Historic crisis, historic opportunity: Using evidence to mitigate the effects of the COVID-19 crisis on young children and early care and education programs.* University of Michigan Gerald R Ford School of Public Policy: Education Policy Initiative.

White, L., Quick, H., Bergey, R., delos Reyes, I.B., & Carbuccia-Abbott, M. (2021). *Supporting child care providers and the dual language learners they serve during the COVID-19 crisis: Insights from a statewide survey*. Washington DC: Early Edge California and American Institutes for Research. Retrieved from: https://earlyedgecalifornia. org/supporting-child-care-providers-and-the-dual-language-learners-they-serve-during-the-covid-19-crisis/.

Wilson, B., Lonigan, C. J. (2010) Identifying preschool children at risk of later reading difficulties: Evaluation of two emergent literacy screening tools. *Journal of Learning Disabilities, 43*(1) 62–76. doi: 10.1177/0022219409345007

Wyse, A. E., Stickney, E. M., Butz, D., Beckler, A., & Close, C. N. (2020). The potential impact of COVID-19 on student learning and how schools can respond. *Educational Measurement, Issues and Practice, 39*(2), 60-64. https://doi.org/10.1111/emip.12357

Zelazo, P. D., & Carlson, S. M. (2017). Embracing complexity in the study of executive function and its development. In *Advancing developmental science* (pp. 110-118). Routledge.

Table 1

Average Descriptive Statistics of Children in Virtual Assessment Sample

	Population	Random Sample	Final Sample
White (Non-Hispanic)	0.07	0.10	0.08
Black (Non-Hispanic)	0.36	0.42	0.37
Hispanic (Any Race)	0.51	0.41	0.48
Asian/PI (Non-Hispanic)	0.04	0.04	0.05
Other Race (Non-Hispanic)	0.02	0.03	0.02
Home Language - English	0.62	0.69	0.66
Home Language - Spanish	0.32	0.23	0.28
Home Language - Other	0.07	0.08	0.07
First Year in Program	0.47	0.30	0.29
Second or Third Year in Program	0.53	0.70	0.71
Age at K Cutoff - 3 Years Old	0.43	0.37	0.29
Age at K Cutoff - 4 Years Old	0.57	0.63	0.71
Number of Days Absent	--	--	--
Number of Days Absent or Late	--	--	--
Number of Virtual Meetings Attended	--	--	--
IEP/IFSP	0.11	0.11	0.12
Sociodemographic Risk Factors			
SSI Recipient	0.09	0.13	0.10
TANF Recipient	0.07	0.08	0.08
WIC Recipient	0.49	0.52	0.52
SNAP Recipient	0.52	0.57	0.53
Primary Parent w/out High School Degree	0.20	0.17	0.19
Single Parent Household	0.60	0.64	0.64
Primary Parent Unemployed	0.42	0.41	0.42
Foster Child	0.02	0.03	0.02
Homeless Flag	0.09	0.08	0.11
Count of Sociodemographic Risk Factors	2.49	2.63	2.61
Parent-Teacher Communication 1x+ Weekly	0.75	0.77	0.77
Parent Satisfied with Communication Frequency	0.91	0.92	0.90
Enrolled in Virtual Learning Model	0.26	0.50	0.42
N	1823	600	336

Note. Table shows means of nonmissing demographic variables for the entire population of Head Start children attending Early Learning Inc (Column 1), Head Start children randomly selected for virtual assessments (Column 2), and Head Start children who were randomly selected *and* completed at least one assessment (i.e., final sample; Column 3) during the 2020-21 SY. Sociodemographic risk factors include count of the following child-level flags: Recipient of government subsidy or services (i.e., SSI, WIC, or SNAP); primary parent does not have a high school diploma; child resides in a single parent household; primary parent is unemployed; foster child; family is homeless. Proportions shown for race/ethnicity, language, age or year in program, and sociodemographic risk factors (e.g., 0.08 for White in Column 3 indicates 8 percent of children in the final sample were White). Means shown for count variables (attendance and total number of sociodemographic risk factors). Parent teacher communication measures are derived from responses to end-of-year online caregiver surveys.

Table 2

Number of Assessments by Region, Assessment Period and Learning Model

	Fall		Winter		Spring	
	In-person	Virtual	In-person	Virtual	In-person	Virtual
Nevada	70	56	66	50	61	44
Pennsylvania	57	41	54	39	51	32
New Jersey	61	30	59	29	53	26
Wisconsin	8	3	6	2	5	0
Total N	196	130	185	120	170	102

Note: Count includes children with at least one nonmissing assessment score in each assessment period, region, and learning model (in-person or virtual). Assessments took place in the fall (November), winter (January), and spring (April) of the 2020-21 SY.

Table 3

Measures, Constructs, and Considerations for Virtual Delivery in 2020-21 School Year

Measure	Respective Construct	Psychometric Properties for the Instruments	Administration Adaptations for Virtual Delivery
Test of Preschool Early Literacy – Print knowledge (TOPEL-PK)	Print knowledge (print concepts, letter discrimination and identification, letter-sound ID, word discrimination)	Test-retest reliability = .73, internal consistency = .88; concurrent validity with TOPEL early literacy index = .71-.72, sensitivity = .97, specificity = .59	• Digital flip book PDF • Digital Google Form score sheet • Used "screen share" feature on Zoom to present items to child
Individual Growth and Development Indicators (myIGDIs)	Oral counting, number naming, and quantity comparison	Test-retest reliability = .71-.88; concurrent validity ranges .60-.75 with other standardized measures	• Digital flip book PDF • Digital Google Form score sheet • Used "screen share" feature on Zoom present items to child
Minnesota Executive Function Scale (MEFS)	Working memory, inhibitory control, and cognitive flexibility	Test-retest reliability = .93; convergent validity ranges .73 - .77 with other standardized measures	• Used "screen share" feature on Zoom with a separate iPad for scoring, where the assessor selected the appropriate answers (as if he/she was the child responding directed on the iPad) • Caregivers/teachers accompanied via Zoom and provided verbal cues on which box children selected for each item (i.e. "left" or "right")

Note: Psychometric properties reported in this table are publisher-reported and derived from assessments administered in in-person conditions.

Figure 1

Process Diagram for Virtual Assessment Delivery Across Assessment Team.