methods, data, analyses | Vol. 20(1), 2026, pp. 87-107 DOI: 10.25521/mda.657

Video Interviewing for Standardized Achievement Assessments with Children

Paula Fomby[1], Narayan Sastry[2], Helen Johnson[2], Sarah Crane[2] & Veronica Connors-Burge[2]

[1] *University of Pennsylvania*
[2] *University of Michigan*

Abstract

We describe the design, execution, and results of a pilot study with a convenience sample ($N = 32$ children aged 5-17 years in 15 families) to evaluate the operational feasibility of video interviewing as a mode to administer standardized assessments of math, reading, and vocabulary skills to children aged 5-17 years in the United States. We review the videoconferencing interface from the perspective of the test examiner and examinee, describe protocols to provide cellular-enabled electronic tablets for examinees' use at home, summarize the fieldwork experience, and describe lessons learned. We conclude that video interviewing is a viable alternative to in person administration from an operational standpoint and can be useful when efforts to reach study participants in-person are cost-prohibitive or ineffective.

Keywords: remote survey administration, standardized testing, children, pilot study, survey operations

Standardized Testing with Children in Survey Research

Children's demonstrated skill in literacy and math is an important indicator of well-being and development for individuals and populations (Organisation for Economic Cooperation and Development, 2020). Consequently, many large-scale surveys of child development include standardized assessments of children's performance in core academic subjects. Raw scores on these assessments are age-standardized against national norms to provide metrics that may be used to describe and explain population variation in learning achievement and associated outcomes.

To date, standardized assessments with children in community-based samples have been conducted in-person during field interviewers' visits to families' homes or other settings where the interviewer can manage the testing environment. In-person administration is attractive because it allows the interviewer to identify the optimal test setting within a child's home, focus children's attention on test stimuli, deflect interference from other family members, and prevent children's efforts to seek outside help to answer questions. However, for at least four reasons, in-person assessment has become increasingly challenging. First, home visits in national samples are expensive compared to telephone and web interviewing (Schober, 2018), and in longitudinal studies, home visits become more expensive over time as participants move and become geographically dispersed. Second, survey respondents have become increasingly resistant to home visits in response to time constraints and preferences regarding privacy. Further, there is some evidence of declining fieldwork effort to recruit resistant respondents to in-person interviews (Bates et al., 2008; Brick & Williams, 2012; Vercruyssen et al., 2011; Williams et al., 2018). Third, scheduling changes for in-person appointments can be costly and inefficient. Especially when interviewers need to travel some distance to a respondent's home, broken appointments and last-minute rescheduling complicate the logistics of completing a home visit. Fourth, the Covid-19 pandemic demonstrated that public health emergencies or other extreme events can make travel or physical proximity unjustifiably risky. These factors have led to the consideration of a range of new approaches to the remote collection of various measures of children's development (see Tsuji et al., 2022).

Video interviewing potentially offers a cost-effective, unintrusive, and safe option to conduct standardized assessments with children at home, but to date there is no published empirical evidence of the operational feasibility of this approach. Video interviewing with respondents in their homes has been adopted successfully as a replacement or complement to face-to-face survey interviewing

Direct correspondence to
Paula Fomby, University of Pennsylvania, Department of Sociology, 3718 Locust Walk, Philadelphia, Pennsylvania 19130, USA
E-mail: pfomby@sas.upenn.edu

with adults (Centeno et al., 2024; Hanson et al., 2026). And prior work has demonstrated that video interviewing is an efficient, valid, and reliable method to conduct assessments when children and adolescents are tested in schools under tightly controlled circumstances (Wright, 2018). We offer the first evaluation of whether such success can translate to standardized assessment with children in home environments. A similar approach has recently been pilot tested for the first time for the remote collection of anthropometric assessments for adolescents and adults (Ghosh-Dastidar et al., 2020) and to assess preschool children's anthropometrics and motor skills (Button et al., 2023). This approach could also be used for other measures that require standardized measurement and supervision by or interaction with a trained interviewer.

Purpose and Aims of the Pilot Study and This Paper

The purpose of the pilot study and this paper is to evaluate the operational feasibility of video interviewing to administer standardized achievement assessments at home with school-aged children in the United States. The primary motivation is to ascertain whether this mode of data collection could be adopted in future waves of the Panel Study of Income Dynamics Child Development Supplement (CDS), a US national study of the well-being of minor children and their caregivers conducted by telephone and in-person interview approximately every five years since 1997 (McGonagle & Sastry, 2015). For several reasons, CDS is well-positioned to adopt video interviewing. First, the CDS sample is geographically diffuse. CDS is embedded in the Panel Study of Income Dynamics (PSID), a household-based panel study started in 1968 that has followed original household members and their descendants to the present day. Sampled households were geographically clustered when PSID began, but families have since become nationally dispersed. Second, CDS includes families living in rural or otherwise geographically remote areas where it is not economically efficient to hire local interviewers. Field interviewers' travel from more densely populated areas to rural or remote areas is expensive and must be carefully coordinated to permit visits to multiple participating households over a few days. Third, families with children are participating in CDS during a busy life stage, when parents may be juggling their own employment and the needs of preschool and school-aged children (Vercruyssen et al., 2011). Beyond the logistical challenge of scheduling a home visit, such a request can risk families' eventual attrition from the study if they perceive continued participation as a significant burden (Sastry et al., 2021). Giving families the choice to participate by video interview or in person could improve participation and retention (see Conrad et al., 2017).

In sum, the use of video interviewing to administer standardized assessments remotely could offer a strategy to control costs and increase efficiency during fieldwork, increase wave-specific response rates, and reduce panel study attri-

tion when offered to families who are unwilling or unable to schedule a home visit. This pilot study is the first step in a larger program to evaluate the merits of this approach.[1] The aims are to assess four aspects of video interviewing to administer standardized assessments: (a) technical development and implementation of a video interface for the standardized tests that was comparable to the in-person format from the test taker's perspective; (b) application development to establish a stable, secure videoconferencing session between the test administrator and test taker; (c) protocols to provide cellular-enabled electronic tablets to test takers for use during the test administration; and (d) data production, as measured by test completeness and perceived test taker burden. The first three aspects of the pilot study are novel to CDS and the survey operations team, and to our knowledge, these protocols have not been tested previously on a sample of children participating at home.

The pilot study was conducted with a small convenience sample (N = 32 children ages 5-17 years in 15 families) in one US state. The use of a small convenience sample allowed the study design team to put additional resources into instrument and protocol development and to evaluate these multiple new technologies and procedures with relatively little risk. The tradeoff is that the findings may underestimate the challenges that could arise from using video interviewing for standardized testing in larger and more diverse samples. At a minimum, though, the outcomes provide useful lessons that permit better-informed development and evaluation for large-scale studies as they move toward integrating video interviewing with other more long-standing modes of data collection.

Below we describe the pilot study design and its implementation. We conclude with an assessment of operational feasibility, a discussion of lessons learned, and attention to the study's limitations.

Pilot study design

This section describes the steps to develop and implement a method to conduct standardized testing with children at home, from technical development through test administration. Development, fieldwork, and post-production processing for the pilot study were conducted by Survey Research Operations (SRO) at the Institute for Social Research, University of Michigan, between July 2021 and April 2022. All study materials and protocols were approved by the Health Sciences and Behavioral Sciences Institutional Review Board at the University of Michigan.

[1] Building on the pilot study, CDS will evaluate mode equivalence and integrate video interviewing with other modes of data collection in a larger and more representative pretest sample before adopting the method for use in the study's active sample.

Technical Development

Instrument Adaptation for Video Interviewing

The Woodcock-Johnson IV (WJ-IV) is a battery of 20 subtests designed to characterize children's learning performance in various domains relative to their age or grade peers or to assess children's learning progress over time (Mather & Jaffe, 2016). The pilot study included three subtests used in prior waves of CDS: (a) Letter-Word Identification (word and letter recognition or pronunciation); (b) Passage Comprehension (making inferences from written passages); and (c) Applied Problems (problem-solving through quantitative reasoning). For the pilot study, subtests were drawn from WJ-IV Form B, and all testing was completed in English.

During in-person administration, the WJ-IV subtests are presented in a standing easel binder referred to as a stimulus book. The test administrator and test taker sit across from each other, and the stimulus book is positioned between them. The test taker sees one page containing a set of stimuli, and the administrator sees correct responses to those stimuli on a page on the back side of the easel. A child begins each subtest at a point determined by his or her current age or grade in school. This starting point is referred to as a basal. If the child responds incorrectly to each of the first several stimuli, he or she is rerouted to an easier starting point, and the basal is reset. With each correct response, the child advances to a more challenging stimulus. The subtest terminates when the child provides multiple consecutive incorrect responses after the basal has been established or responds to the last stimulus in the subtest. Raw scores are calculated as a function of the child's basal and the number of items answered correctly beyond the basal. These raw scores are then transformed into a set of criterion-based scores to measure individual ability and norm-referenced scores to measure performance relative to age or grade peers.

WJ-IV testing rules had been programmed into Blaise survey software previously to support computer-assisted in-person interviewing in CDS. The program uses preloaded information about a test taker's age and grade in school to establish the preliminary basal for each subtest. After each stimulus is presented, the test administrator records whether the child has provided a correct or incorrect response and follows programmed prompts to move through each subtest based on performance. To adapt the in-person administration design for video interviews, each test stimulus page was digitally scanned, formatted, and uploaded into the existing Blaise application to be screenshared by the test administrator to a test taker.[2]

[2] Riverside Insights, the publisher of Woodcock-Johnson IV, granted permission to adapt copyrighted material for electronic use.

Hardware

A laptop computer and external monitor in portrait orientation were provided to each test administrator (Figure 1). The Blaise application for entering test takers' responses to test stimuli was programmed to appear on the laptop monitor and the test stimulus page to appear on the external monitor so that both were visible to the test administrator. A high-resolution webcam and headset with a built-in microphone were also provided.

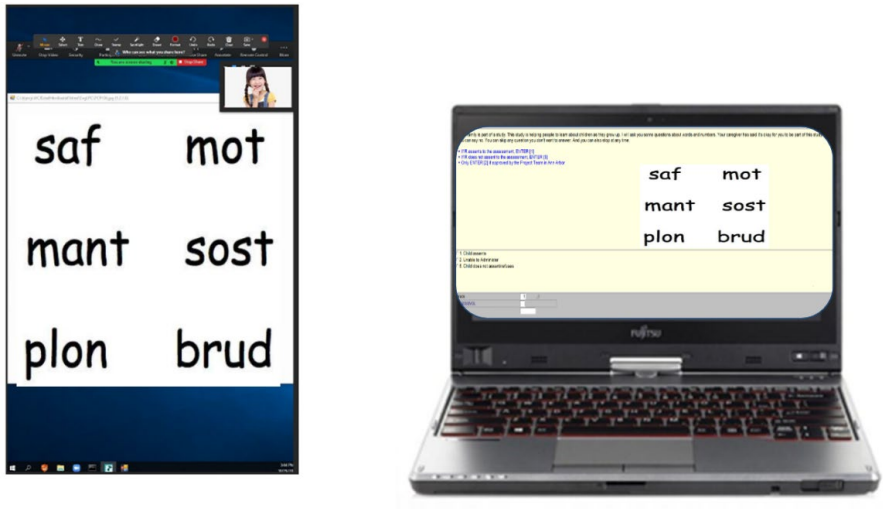

Figure 1 External monitor (left) and laptop computer (right) configured for use by the test administrator during the remote administration of the Woodcock-Johnson IV Letter-Word subtest

Test takers used a Samsung Model 5 electronic tablet configured solely to support remote WJ-IV administration on the Zoom video conferencing platform (Figure 2). Electronic tablets were used in order to standardize the test-taking format across home environments and to allow participation where families lacked a home computer or a high-speed internet connection. Each tablet included a built-in video camera and microphone and a 10.5-inch display, consistent with the WJ-IV publisher's guidelines for electronic administration (Riverside Insights, 2020).[3] A push notification application was installed on the home screen so that

[3] Some recommendations provided by the publisher were not incorporated into the pilot study. These include a peripheral camera directed at the examinee's workspace, a headset with two earphones and a microphone for the examinee's use, and a mouse or trackpad for the examinee's use to point to responses on the screen. The last recommendation was obviated by examinees' use of Zoom's annotation function for the same purpose.

test takers could join the test administrator's active video conferencing session. Mobile device management software was installed on each tablet so that its contents could be erased remotely if it were lost or stolen. No personal information was stored on the tablets. Tablets were enabled with a Verizon cellular data plan to establish the video connection between the test administrator and test taker.[4]

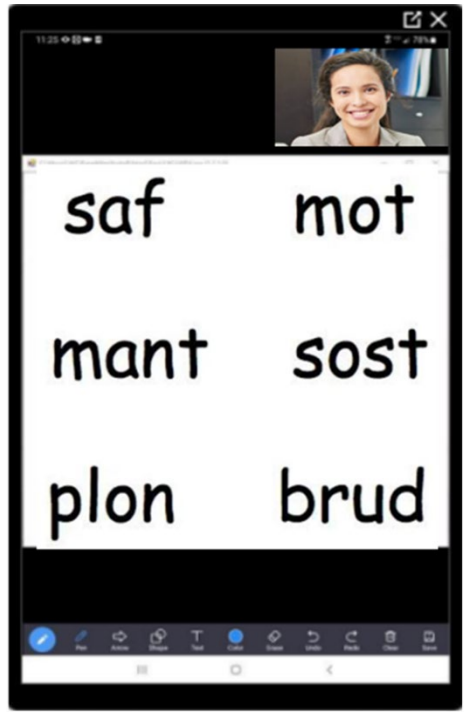

Figure 2 Electronic tablet configuration for use by the respondent during remote administration of the Woodcock-Johnson IV Letter-Word subtest

The video conferencing interface was determined by Zoom's format. During an active session, the test administrator and test taker observed one another through the video feed from their respective web cameras. To minimize distraction, video self-view was disabled for both parties (see Feuer & Schober, 2026, on the effect of self-view on survey response). The test administrator used Zoom's screenshare function to display stimulus pages on the test taker's electronic tab-

[4] As of 2017 (the most recent data available at the time of the pilot study), Verizon mobile wireless coverage reached 97.4% of the US resident population (Federal Communications Commission, 2017).

let screen. During screen sharing, the video tile showed the active speaker in the upper right-hand corner of the screen and was positioned so that it did not block text on the stimulus page.

Most test stimuli required a verbal response, but for young children, some items required the child to point to an object on the screen. To accommodate this, Zoom's annotation function was enabled for both the test administrator and test taker. The annotation feature allows a user to draw a point, line, circle, or other shape on an electronic tablet's touch-enabled screen with their finger. The image appears as a visible overlay on the shared screen, allowing all participants to see the annotation as it is being made. During administration, the test administrator prompted young children to use this feature to select their response on the screenshared stimulus page.

Staffing

Two field interviewers employed by SRO who had experience administering in-person WJ assessments to children were staffed as test administrators. We refer to these staff members as test administrators going forward. In addition to standard training, they received four hours of specialized training in procedures for the remote administration of WJ-IV assessments. This included training in the use of an external monitor and the Zoom interface as well as paired practice with the electronic tablets to experience the test taker's perspective. During fieldwork, test administrators worked from their own homes using computer and communications hardware supplied by SRO.

Sample Recruitment

The planned sample size was 30 children aged 5-17 years in 20 families, which was expected to provide sufficient variation across child age, family socioeconomic status, and geographic location (urban vs. rural) to distinguish contexts in which video interviewing was more or less feasible. Project management and test administrators recruited from their personal social networks a convenience sample of 19 families that included 40 age-eligible children. Sample composition was captured through an electronic screening form programmed in Qualtrics that was completed by the contact person in each recruited family. Four families were unable to participate due to scheduling conflicts during the fieldwork period, and one child was coded out as ineligible due to disability.[5] The final

[5] Criteria for ineligibility are developed by SRO in consultation with study investigators. In the current case, children with significant vision or hearing impairments or a limiting developmental disability were categorized as ineligible.

sample included 32 children in 15 families, and all eligible families participated in the study. All families resided in the same state in the US Midwest.

Table 1 presents descriptive statistics for the pilot study sample and for families with minor children (5-17 years old) in the same state in 2022 (American Community Survey, Ruggles et al., 2025).[6] Compared to the state population, children were about two years younger in the pilot study sample (median age = 9 years) compared to children in the state (11 years), and pilot study caregivers were more often non-Hispanic white (78% versus 70% among household heads statewide) and highly educated (69% held a Bachelor's degree or higher versus 37%). About one-fifth of recruited families lived in areas classified as rural (vs. 12% in non-metropolitan areas statewide). To the extent that these demographic, socioeconomic, and geographic indicators are correlated with performance, pilot study results may not be generalizable to other groups.

Appointment Scheduling

Test administrators were assigned to families with whom they had no prior personal contact. During a telephone call, the test administrator verified the eligibility of children rostered on the screening form as potential participants. Test administrators then scheduled appointments based on the family's availability. With the caregiver's permission, the administrator sent a Google Calendar invitation by e-mail so that the caregiver could import the appointment to their own electronic calendar. One day prior to the appointment, the administrator called the caregiver with a reminder about the upcoming appointment and to address any questions.

Appointments were scheduled to take place at least one week after recruitment to allow time to ship study materials to a family's home using United Parcel Service (UPS) two-day delivery. Each shipment included a study information sheet, the electronic tablet in a protective case,[7] a printed return label, and instructions directed to the participating child's caregiver (i.e., a parent or guardian) on how to use and return the tablet. Use instructions included written guidance and accompanying annotated screenshots on how to (a) power the tablet on, adjust the volume, and navigate to the home screen; (b) use the push notification application to enter a Zoom videoconferencing session; (c) turn on the video camera and use the annotation function in Zoom; and (d) exit Zoom. The caregiver was also directed to prepare a quiet space in the home free from

[6] Child's sex was not recorded.

[7] To eliminate the risk of loss, no power cord was included in the shipment. Tablets had an estimated 14.5-15 hours of active battery life before needing to be recharged. When powered off for shipment, the battery can remain charged for several weeks. No instances of data loss or interruption caused by device power issues were reported.

distraction where the child could complete the assessment and, if relevant, to keep other children out of earshot until it was their turn to participate.

Table 1 Descriptive statistics, Woodcock Johnson video interview pilot study sample ($N = 32$ children in 15 families) and same-state families with children age 5-17 years (American Community Survey [ACS], 2022)

	Pilot study proportion (*N*)	State proportion
Number of children[a,d]		
1	0.27 (4)	0.30
2	0.40 (6)	0.40
3+	0.33 (5)	0.30
Child age (years)[e]		
5–8 years	0.28 (9)	0.29
9–12 years	0.38 (12)	0.30
13–17 years	0.34 (11)	0.42
Educational attainment[b]		
Less than high school	0.00 (0)	0.08
High school/GED	0.09 (3)	0.20
Some college	0.06 (2)	0.23
Associate's degree	0.16 (5)	0.11
Bachelor's degree	0.22 (7)	0.22
Master's degree	0.31 (10)	0.12
Professional/doctoral degree	0.16 (5)	0.03
Race and Hispanic ethnicity[b]		
Non-Hispanic white	0.78 (25)	0.70
Non-Hispanic Black	0.09 (3)	0.15
Non-Hispanic Asian	0.06 (2)	0.04
Non-Hispanic other	0.00 (0)	0.05
Hispanic (any race)	0.06 (2)	0.06
Rurality[c]		
City or township	0.82 (26)	0.88
Rural	0.19 (6)	0.12

[a] Refers to the number of minors participating in the pilot study.
[b] Refers to caregiver characteristics in pilot study and to householder characteristics in ACS.
[c] Measured by metropolitan status in the ACS.
[d] Number of children in Pilot: $M = 2.13$, $SD = 0.92$; Number of children in State population: $M = 2.16$, $SD = 1.09$.
[e] Child age in Pilot: Median = 9, $M = 9.5$, $SD = 3.41$; Child age in State population Median = 11, $M = 11.2$, $SD = 3.72$.

Test Administration

At the appointed time, the test administrator telephoned the caregiver to confirm that the shipment had arrived,[8] the tablet was powered on, the caregiver was familiar with the tablet's functionality as described in the user guide. and participating children were available and prepared to begin. The administrator then instructed the caregiver to initiate the video conferencing session via the push notification application on the tablet's home screen, after which the telephone connection between the test administrator and caregiver was terminated. In the videoconferencing session, the test administrator obtained informed consent from the caregiver and verbal assent from the child for the child's participation. The test administrator also requested the caregiver's consent to record the session for quality control.[9] The test administrator then launched the first subtest (Letter-Word Identification).[10] The caregiver was asked to remain nearby and to allow the child to complete each subtest independently.

In families with more than one participant, children completed the assessments back-to-back.[11] When all participants had finished, the test administrator reminded the caregiver to package and return the electronic tablet and terminated the video conference session. The test administrator contacted UPS to schedule prepaid pickup of the packaged tablet, usually for the following day. Each child who completed the assessments received a check for $50 as a token of appreciation after the electronic tablet was received at SRO. Blaise data were transferred to SRO using a secure file transfer protocol and removed from the test administrator's laptop daily. SRO technical support staff recharged and sanitized each electronic tablet and checked device settings before shipping the tablet to the next participating family.

Fieldwork occurred over three weeks beginning in the second week of February 2022. At the conclusion of fieldwork, raw data from the assessments were batch uploaded to the proprietary WJ-IV online scoring system with individual records distinguished by a unique numeric identifier. No personally identifying information was included in the upload. Calculated scores were returned from the scoring system in a comma-delimited text file.

---

8 Delivery was also confirmed independently by the service provider.

9 All caregivers provided consent to record. The recording captured the videoconferencing session as it appeared on the test administrator's external monitor. Consent to record was not required to proceed with the test administration.

10 Prior to launching the first subtest, the caregiver was also advised to provide the child with paper and pencil for use in the Applied Problems section, consistent with guidelines for in-person administration.

11 In one family with three children, one child participated the day after the other two.

Outcomes

In this section, we review the outcomes of the pilot study's implementation. Overall, the study demonstrates the operational feasibility of video interviewing for standardized testing with school-aged children while providing some lessons to improve efficiency and ensure test fidelity. We assess outcomes using six sources of information: a debriefing with test administrators at the end of fieldwork; test administrators' responses to closed-ended survey questions about each test taker's experience with each subtest; test administrators' open-ended written comments in Blaise; a log of shipment and return dates for the electronic tablets; timing data from Blaise; and participants' calculated WJ-IV scores.

Technical Systems

Here we review the implementation of study design choices related to Blaise programming, hardware configuration, cellular coverage, and videoconferencing format. Most worked as planned, with some caveats.

The integrated laptop PC and external monitor allowed test administrators to observe test takers, monitor the screenshared presentation of subtest stimulus pages, and record test takers' responses without toggling between screens. In the post-fieldwork debriefing, both test administrators noted that it was helpful to see the active test stimulus page on the external monitor while the test subject was also looking at it, in contrast to in-person administration with an easel, where the test stimulus page is visible only to the test taker. However, both also stated that additional training (beyond the four hours provided) on the hardware configuration and Zoom's functionality before beginning fieldwork would have increased their familiarity and confidence.

We use the test administrators' debriefing and their open-ended comments to describe participating families' experience with the videoconferencing application. All caregivers successfully used the push notification on the electronic tablet's pared-down home screen to join the videoconferencing session. No information is available about whether caregivers found the printed user guide sent in advance or the real-time guidance from the test administrator to be more helpful for getting started. It is important to note that caregivers and children in this relatively highly educated sample may have been familiar with push notifications and the Zoom interface.

Cellular data coverage on the tablets was mostly stable and reliable and provided clear communication. There were three instances of unstable or lost connections (one each in a city, township, and rural area), but these were resolved and did not prevent completion of the assessments. Had connectivity issues persisted, test administrators were trained to ask families to travel to another location, such as a library or the home of a friend or family member, to complete the

assessment. However, it was never necessary for a family to travel to receive cellular service for the study. Other studies wishing to implement a similar design should review service coverage in the areas where sample members reside and develop protocols for cases where service is unstable or unavailable.

All caregivers were able to turn on the annotation feature at the beginning of the videoconferencing session. In open-ended comments provided after children completed their subtests, the test administrator noted that in one family the caregiver helped younger children (aged 6 and 9 years) to use the annotation function. No other observations of caregiver assistance related to children's use of the annotation function were reported. However, the test administrator's instruction to young children on how to use the feature added administration time. Children in the youngest age group (aged 5-8 years) were the most likely to respond to stimuli in the Letter-Word Identification subtest that required them to use the annotation function (e.g., to identify a letter of the alphabet), and children in that age group needed about twice as much time to complete the subtest (M = 6.66 minutes) compared to test takers in middle childhood (aged 9-12 years, M = 3.33 minutes). The Passage Comprehension test also required early readers to point to their responses for some stimuli (e.g., to identify the shape represented by a printed word), resulting in longer administration times compared to the middle childhood group (M = 11.4 minutes in early childhood vs. 10 minutes in middle childhood; see Table 2).

Test administrators offered two further recommendations based on their fieldwork observations. First, materials provided to participating families should include a stand to hold the tablet and instructions for the child to sit at a table while completing the subtests. In the pilot study, children held the tablet by hand. During the Applied Problems subtest, children frequently put the tablet down in order to solve problems. Because the test administrator could not see the child, it was not possible to know whether they were solving the problems by hand on their own. Second, Zoom features should be locked down. Test administrators observed that some children changed the name field or attempted to alter the annotation functionality.

Hardware Durability and Shipment

Test administrators and project staff reported that the electronic tablet hardware worked reliably in the field. All tablets were fully charged and operational and the screens intact when the devices were delivered to test takers' households, and there was no reported damage when tablets were returned to SRO throughout fieldwork. No tablets were reported missing or stolen.

All tablets shipped from SRO using UPS two-day delivery service reached recipients as scheduled one to two days before the scheduled appointment. Test administrators usually scheduled UPS to pick up the tablet for return the day

after the test administration was completed, and SRO logged receipt two to three days later. In one case, a test administrator drove to a participating family's home to collect the tablet when it had not been picked up by UPS after several days. Upon receipt at SRO, tablets were processed for reuse and available for shipment to the next household one to two business days later. The interval between shipments for a given tablet ranged from seven to 15 calendar days, with a median of 10 days. The log indicates that longer intervals occurred because of lags in scheduled appointments rather than delays in processing or shipping.

The smooth circulation of tablets between SRO and participating families likely benefited from high compliance in this small convenience sample. Participants' willingness to keep their scheduled appointments was another benefit. A request to reschedule after a shipment was already *en route* could have significantly hampered the capacity to send tablets to other families in advance of their scheduled appointments. Yet because tablets were often tied up in shipment or at a participating family's home until the time of a scheduled appointment, the study would have benefitted from having more than five tablets available even in this small sample. Overall, the pilot study achieved the basic objective of demonstrating the feasibility of the shipment and use of electronic tablets in a household-based study with child respondents. But significantly more testing is required to establish a system suitable for larger, more dispersed, and more population-representative samples.

Data Production

All testing was completed with children in their homes. Thirty-one children completed all subtests. One child (age 6 years) did not complete the Passage Comprehension subtest.

Test administrators described all participants as cooperative, but children varied in their level of engagement with the subtests. Two children asked to skip some questions, and one child required frequent redirection to focus on the test stimuli. In one case, a caregiver tried to intervene to help the child answer a question, and in another the child turned to their caregiver for assistance. In both cases, the test administrator advised the caregiver to allow the child to answer each question on their own and reported no further interference. Test administrators described the effort to focus children's attention and redirect caregivers during the video interview as no more challenging than what they had previously encountered during in-person administration of the same subtests.

It was not always possible to keep study participants separated from other family members during testing. Test administrators reported three instances of temporary distractions when family members passed through the room where children were completing the assessment and one instance where a parent was

speaking on the telephone in another room nearby. One related concern is that disruption and distraction could have negatively affected the results for the child being tested. Test administrators' ability to monitor the child's environment through the video connection allowed them to assess whether the conditions made the test infeasible to administer, which never occurred.

Another potential consequence is that children who were in line for the assessments may have seen or overheard portions of a subtest's content and the test taker's responses. In theory, there should be little associated risk. Children of different ages would be expected to complete different segments of each subtest, and test takers (and others around them) do not know whether they have provided a correct response. Still, based on this experience, we advise providing headphones or earbuds to test takers to prevent the test administrator's instruction from being overheard and to block out background noise. A tradeoff is that such a strategy would require the child to remove the audio device should the test administrator and caregiver need to communicate during the test administration.

Test Performance

This pilot study was designed to assess whether video interviewing to administer the WJ-IV to children was operationally feasible rather than to establish mode equivalence compared to in-person administration. We present results from the WJ-IV subtests as descriptive. The key takeaways are that nearly all children completed all subtests (with the exception of one incomplete subtest noted above) without greater difficulty than during in-person administration and the distribution of test scores does not suggest systematic upward or downward bias. We describe three outcomes indicative of test performance: time to subtest completion in minutes, proportion of respondents who had some or a lot of difficulty completing the subtest, and age-standardized scores (Table 2). All statistics are reported overall and by age group (5-8, 9-12, and 13-17 years). For comparison, we report weighted standardized scores and frequency of any testing difficulty among same-age participants in the 2019 PSID Child Development Supplement (CDS). Timings for individual subtests in CDS were not available.

Table 2 Standardized scores, time to completion in minutes, and proportion experiencing some or a lot of difficulty, WJ-IV Letter-Word Identification, Passage Comprehension, and Applied Problems, overall and by age group with comparison to the 2019 Panel Study of Income Dynamics Child Development Supplement (CDS)

Subtest	Metric	Data	Overall (*N* = 32)	5-8 years (*N* = 9)	9-12 years (*N* = 12)	13-17 years (*N* = 11)
			Mean (*SD*) or proportion	Mean (*SD*) or proportion	Mean (*SD*) or proportion	Mean (*SD*) or proportion
Letter-Word Identification	Standardized score	Pilot	101.7 (13.7)	101.3 (13.8)	104.1 (12.2)	99.4 (15.9)
		CDS	100.5 (15.9)	97.3 (18.2)	104.0 (15.4)	100.8 (14.8)
	Time	Pilot	3.9 (2.6)	6.7 (2.9)	3.3 (1.4)	2.2 (0.9)
	Any difficulty	Pilot	0.09	0.11	0.17	0.00
		CDS	0.49	0.58	0.43	0.38
Passage Comprehension	Standardized score	Pilot	97.2 (13.4)	100.5 (10.3)	99.1 (12.0)[a]	92.8 (16.4)
		CDS	94.0 (13.9)	96.1 (14.5)	94.9 (14.0)	91.6 (12.8)
	Time[b]	Pilot	9.8 (3.4)	11.4 (3.9)	10.0 (3.11)	8.2 (2.8)
	Any difficulty	Pilot	0.47	0.67	0.42	0.36
		CDS	0.63	0.75	0.64	0.54
Applied Problems	Standardized score	Pilot	98.6 (18.0)	99.6 (20.2)	96.5 (18.4)	100.0 (17.4)
		CDS	99.0 (17.0)	96.4 (15.4)	100.5 (16.8)	99.3 (17.8)
	Time	Pilot	15.9 (4.5)	13.8 (5.5)	16.0 (3.6)	17.5 (4.1)
	Any difficulty	Pilot	0.34	0.56	0.25	0.27
		CDS	0.66	0.72	0.63	0.56

[a] *N* = 11
[b] Excludes one outlier whose test administration was suspended overnight.
CDS estimates are survey-weighted to represent the population of US children.
Time to module completion is not available for CDS.

The median time to complete all of the assessments was 29.5 minutes, which is similar to in-person administration times in the CDS (Panel Study of Income Dynamics, 2022). The Letter-Word Identification was the shortest (3.9 minutes

overall), followed by Passage Comprehension (9.8 minutes)[12] and Applied Problems (15.9 minutes). As noted above, the youngest children required more time than older children to complete the Letter-Word and Passage Comprehension assessments. Older children required the most time to complete the Applied Problems assessment.

Test administrators reported in post-test observations whether a participant had no difficulty, some difficulty, or a lot of difficulty with each subtest. These reports are intended to assess the child's evident cognitive burden when evaluating the test stimuli, but here the items might also have captured difficulty specific to the modality. Fewer than 10% of children had any difficulty with the Letter-Word Identification subtest, but nearly half had at least some difficulty with the Passage Comprehension subtest, and one-third had at least some difficulty with the Applied Problems subtest. The youngest children (aged 5-8 years) had difficulty more often than their older peers on the latter two subtests, while children aged 9-12 years more often had trouble with the Letter-Word Identification subtest (17%). The same age patterns were reported for these subtests when they were administered in person during the 2019 wave of CDS (Panel Study of Income Dynamics, 2022), but the frequency of reported difficulty overall was lower in the pilot study sample, perhaps reflecting its relatively educationally advantaged composition.

For each subtest, the standardized score has an expected mean of 100 and a standard deviation of 15. Overall, the average score on the Letter-Word Identification subtest (101.7) is slightly higher than the expected mean, and the average scores on the Applied Problems (98.6) and Passage Comprehension (97.2) are slightly lower. The magnitude of difference compared to the expected values is consistent with those observed in CDS, and pilot study sample means were not significantly different from the expected mean in a series of one-sample hypothesis tests ($p < .05$). The same pattern holds for each age group except for the Passage Comprehension subtest for adolescents age 13 years and older, where the sample mean (92.8) is about half of a standard deviation below the expected value. This relatively large difference may be attributable to chance, to unobserved characteristics of the sample ($N = 11$), or to some interaction between the subtest content and the video format that disproportionately influenced scores on one subtest for one age group. It is worth noting that the sample mean of the standardized score on this subtest for same-age adolescents in CDS was 91.6 (Table 2), suggesting that the lower-than-expected average performance on this subtest may be attributable to factors other than mode of administration.

[12] Statistics exclude one outlier case (age 9 years) for whom administration of the Passage Comprehension subtest was suspended overnight.

Conclusion

The purpose of the pilot study and the objective of this paper were to assess the operational feasibility of video interviewing to administer standardized achievement tests at home with school-aged children. This study represents the first step in a larger program of development and evaluation to inform the adoption of video interviews in large-scale survey research with US children and their families.

We assessed four aspects of video administration. The first two were technical development and implementation of a secure screen-based interface for administering the standardized tests. This effort was largely successful. The integration of a laptop personal computer and external monitor allowed test administrators to observe the test taker's environment, the test stimulus page, and the data collection instrument simultaneously. The annotation function was effective for children with early reading and vocabulary skills, but training in how to use that feature appeared to contribute substantially to test administration time. As an alternative, Riverside Insights, the publisher of the WJ-IV battery of tests, recommends using a mouse or trackpad for test takers to point to their responses. Future work could compare these approaches in terms of their intuitiveness and time requirements.

The use of the screen-based interface relies upon two other aspects of video interviewing that the pilot study assessed: development of a push notification to establish a stable, secure connection between the test administrator and the test taker and protocols to provide cellular-enabled electronic tablets to children in all participating families. The push notification worked as expected. The cellular-enabled internet connection was mostly reliable, but some sessions were dropped due to lost connectivity. In each case, the test administration continued after the connection was re-established, but in a larger sample drawn by other means, participants may be unable or less willing to re-establish a lost connection. The availability of more reliable cellular coverage would improve continuity.

The third aspect was the protocols developed to manage the distribution of electronic tablets with a home-based sample. This poses a potential sticking point for scaling up. In many ways, the pilot study was a best-case scenario because the tablets had relatively little "down time." They were either in shipment, in processing for reuse, or in receipt at the test taker's household for a short time before the scheduled appointment. Further, no tablets were lost, stolen, or damaged, so they all remained in circulation during the short study period. And yet the shipment log and SRO personnel reports suggest that it would have been helpful to have more tablets on hand. Managing similar logistics in a large-scale data collection effort may be untenable or prohibitively costly if the same protocol were established for all participants. However, the approach may be more feasible where video interviews are selectively offered as an option to a subset of

participants, such as those who are geographically isolated or strongly resistant to an in-person visit.

Finally, we considered the data production experience. Here we asked whether children as young as 5 years old would be able to focus on the test-taking task at home when they were engaged with a remote test administrator in a videoconferencing session. Our conclusion is that they can. Although test administrators reported occasional background noise or disruption and described some children as less engaged, their overall assessment was that children's attention to the test and apparent ease or difficulty was comparable to what testers had observed during previous in-person administrations. Test administrators offered several suggestions to improve the data production experience: provide a stand for the electronic tablet so that the testing environment remains visible and children may use the tablet hands-free; provide earphones or earbuds to block out distractions and prevent other test takers from overhearing test content; lock down Zoom features to prevent manipulation; and increase training time and practice for test administrators who are new to the video interviewing format.

An important caveat to these conclusions is that the pilot study was conducted with a small convenience sample of families recruited from the personal networks of study staff. These personal connections likely resulted in greater compliance than would be observed in samples drawn by other methods. Further, caregivers in the convenience sample were highly educated compared to the general population. This may have contributed to caregivers' and children's ease with the technology used in this study and may not be generalizable to more population-representative samples. Finally, the small sample size limits the reliability of our conclusions.

Despite these limitations, the results of the current study suggest that video administration is a feasible and attractive alternative when in-person administration is challenging due to high travel costs or in cases where respondents are hard to reach in-person but amenable to remote contact. Future work to establish mode equivalence compared to in-person administration and to identify the circumstances under which video administration is most cost-effective will further inform decision-making about the appropriate use of this innovative approach for standardized testing with children in survey research.

Funding

The research described herein was supported by the *Eunice Kennedy Shriver* National Institute for Child Health and Human Development (R01HD104671-01S1, PI: Sastry).

References

Bates, N., Dahlhamer, J., & Singer, E. (2008). Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. *Journal of Official Statistics*, *24*(4), 591.

Brick, J. M., & Williams, D. (2012). Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 36–59. https://doi.org/10.1177/0002716212456834

Button, A. M., Webster, E. K., Kracht, C. L., Hendrick, C., Okely, A., Chong, K. H., Cross, P., & Staiano, A. E. (2023). Validation of remote assessment of preschool children's anthropometrics and motor skills. *Frontiers in Digital Health*, *5*, 1168618. https://doi.org/10.3389/fdgth.2023.1168618

Centeno, L., Arrue, J., Good, C., Edwards, B., & Steiger, D. (2024). Practical considerations for developing and implementing a video mode for survey data collection. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *164*(1), 122–146. https://doi.org/10.1177/07591063241277065

Conrad, F. G., Schober, M. F., Antoun, C., Yan, H. Y., Hupp, A. L., Johnston, M., Ehlen, P., Vickers, L., & Zhang, C. (2017). Respondent mode choice in a smartphone survey. *Public Opinion Quarterly*, *81*(S1), 307–337. https://doi.org/10.1093/poq/nfw097

Federal Communications Commission. (2017). *Implementation of Section 6002(b) of the Omnibus Budget Reconciliation Act of 1993, Annual Report and Analysis of Competitive Market Conditions with Respect to Mobile Wireless, Including Commercial Mobile Services* (FCC-17-126). https://www.fcc.gov/document/fcc-releases-20th-wireless-competition-report-0

Feuer, S., & Schober, M. F. (2026). Effects of the self-view window in live video survey interviews. *Methods, Data, Analyses, 20*(1), 124–178. https://doi.org/10.12758/mda.2025.07

Ghosh-Dastidar, M., Nicosia, N., & Datar, A. (2020). A novel approach to anthropometric assessment for geographically dispersed samples: A pilot study. *Preventive Medicine Reports*, *19*, 101125. https://doi.org/10.1016/j.pmedr.2020.101125

Hanson, T., Briceno-Rosas, R., Ainsaar, M., & Kartau, H. K. (2026). Live online video interviewing as a complementary mode to in-person interviews: Evidence from the European Social Survey. *Methods, Data, Analyses, 20*(1), 179–208. https://doi.org/10.25521/mda.663

Mather, N., & Jaffe, L. E. (2016). *Woodcock-Johnson IV: Reports, recommendations, and strategies.* (pp. ix, 590). John Wiley & Sons, Inc.

McGonagle, K. A., & Sastry, N. (2015). Cohort Profile: The Panel Study of Income Dynamics' Child Development Supplement and Transition into Adulthood Study. *International Journal of Epidemiology*, *44*(2), 415–422. https://doi.org/10.1093/ije/dyu076

Organisation for Economic Cooperation and Development. (2020). *Early Learning and Child Well-being.* https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/03/early-learning-and-child-well-being_74042660/3990407f-en.pdf

Panel Study of Income Dynamics. (2022). *Child Development Supplement to the Panel Study of Income Dynamics, public use dataset* (Survey & I. for S. R. Research Center University of Michigan, Eds.). Survey Research Center, Institute for Social Research, University of Michigan.

Riverside Insights. (2020). *Tips for Remote Assessment Using Riverside Insights Clinical Products*. https://info.riversideinsights.com/hubfs/Remote_Assessment%20ProtocolTSdocx-MAY7.pdf

Ruggles, S., Flood, S., Sobek, M., Backman, D., Cooper, G., Rivera Drew, J. A., Richards, S., Rodgers, R., Schroeder, J., & Williams, K. C. W. (2025). *IPUMS USA: Version 16.0* [Dataset]. IPUMS. https://doi.org/10.18128/D010.V16.0

Sastry, N., Fomby, P., & McGonagle, K. A. (2021). Effects on Panel Attrition and Fieldwork Outcomes from Selection for a Supplemental Study: Evidence from the Panel Study of Income Dynamics. In P. Lynn (Ed.), *Advances in Longitudinal Survey Methodology* (pp. 74–99). Wiley. https://doi.org/10.1002/9781119376965.ch4

Schober, M. F. (2018). The future of face-to-face interviewing. *Quality Assurance in Education, 26*(2), 290–302. https://doi.org/10.1108/QAE-06-2017-0033

Tsuji, S., Amso, D., Cusack, R., Kirkham, N., & Oakes, L. M. (2022). Empirical research at a distance: New methods for developmental science. *Frontiers in Psychology, 13*, 938995. https://doi.org/10.3389/fpsyg.2022.938995

Vercruyssen, A., Van de Putte, B., & Stoop, I. A. (2011). Are they really too busy for survey participation? The evolution of busyness and busyness claims in Flanders. *Journal of Official Statistics, 27*(4), 619–632.

Williams, D., Brick, J. M., Williams, D., & Brick, J. M. (2018). Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort. *Journal of Survey Statistics and Methodology, 6*(2), 186–211. https://doi.org/10.1093/jssam/smx019

Wright, A. J. (2018). Equivalence of Remote, Online Administration and Traditional, Face-to-Face Administration of Woodcock-Johnson IV Cognitive and Achievement Tests. *Archives of Assessment Psychology, 8*(1), 23–35.