CONTENTS

1 Abstract
2 Introduction
3 Methods
4 Results
5 Discussion
ABSTRACT
OBJECTIVES

• Develop a protocol for measuring engagement in children presented with captioned and uncaptioned *Peep and the Big Wide World* content.

• Test the proposed methodology in a pilot study.

• Incorporate key lessons from the pilot study into a protocol for larger-scale studies that can be adapted to address a wide range of pediatric conditions, learning styles, and accommodations.
INTRODUCTION

Academic research and anecdotal classroom evidence suggest that closed captioning improves reading skills in all children, particularly those with language-based learning disorders and for whom English is a second language. Although some studies report that captions increase student enjoyment, our understanding of the effects of closed captioning on attention and learning is currently limited. ¹⁴
OPPORTUNITY

WGBH NCAM’s engagement by the U.S. Department of Education to ensure consistent accessibility access to the full suite of *Peep and the Big Wide World* educational materials offered an opportunity for research collaboration between NCAM and Mad*Pow to envision a methodology that would lay the foundation for a larger study addressing a compelling question for which little formal research has been done.

- Are there attention and/or learning effects among student populations who do not require closed captioning, but are exposed to it in educational environments?
HYPOTHESES

Pilot Hypothesis:

• Viewing captioned content may have attention effects for children who do not formally require closed captioning.

Future Hypothesis:

• Accommodations including captions, audio descriptions, and keyboard accessibility may increase engagement and learning in children for whom these accommodations were not originally intended.
METHODS

We recruited two typically-developing children and two children with ADHD, none of whom have a hearing impairment.

• Each participant watched a 9-minute video with captions, a different 9-minute video without captions, and played the same game with and without captions for 3 minutes.

• During each activity, we recorded affective signs of engagement and disengagement, response times to auditory probes, and gaze patterns and fixations using eye-tracking equipment.

• After each activity, we asked the participants questions to assess their self-reported enjoyment and comprehension.
RESULTS

• Eye-tracking, the secondary task technique, observation of affective signs, and self-report were feasible methods for measuring children’s engagement with captioned and uncaptioned educational programming.

• While not definitive due to the small sample size, preliminary data from the pilot study indicated that closed captioning may affect engagement.
INTRODUCTION
BENEFITS OF CLOSED CAPTIONING IN CHILDREN WITH NORMAL HEARING

A survey of published academic literature showed that closed captioning may have positive effects on student enjoyment and literacy.

• Studies have shown that children report increased engagement and enjoyment of captioned media over alternatives such as print and uncaptioned media. ¹⁴

• Bimodal presentation of sound and text enhanced word recognition and recognition memory, the ability to recognize previously encountered objects and concepts. ¹⁵

• Using captioned media can boost the literacy skills of all children, especially English language learners (ELLs) and children with language-based learning disorders. ¹⁴

• For ELLs, captioned media can help improve vocabulary acquisition, listening comprehension, and word recognition. ¹⁴

• Many children who struggle with reading avoid text altogether, and closed captioning gives them additional print exposure. ¹⁴
BENEFITS OF CLOSED CAPTIONING IN CHILDREN WITH SPECIAL NEEDS

Some studies suggest that closed captioning may have additional benefits for children with special needs.

- Evmenova showed that captioned videos improved comprehension in adolescents and young adults with intellectual disability (ID). 15

- Koskinen et al. found that children with learning disabilities (LD) showed significant improvements in word recognition after viewing captioned television with sound. 15

- Kirkland demonstrated that special education students benefitted from and preferred captioned media. 15
FRAMEWORK FOR MEASURING ENGAGEMENT

• Based on an internal literature review, we developed a tripartite framework for measuring engagement in children. 2,12

• Engagement has affective, cognitive, and behavioral components that can be assessed in various ways.
METHODS FOR MEASURING AFFECTIVE ENGAGEMENT

We used the following metrics to measure affective engagement:

• **Signs of engagement** – Observed participants’ facial expressions and body language, recording signs of engagement, which include smiling, laughing, concentrating, excitable bouncing, positive vocalizations, and leaning forward. 5, 6

• **Signs of disengagement** – Observed participants’ facial expressions and body language, recording signs of disengagement, which include frowning, signs of boredom such as fiddling or playing with their ears, shrugging, negative vocalizations, sighing, yawning, and turning away from the screen. 5, 6

• **Reported fun** – Recorded the participants’ subjective measurement of fun using the Smileyometer, a modified Likert scale co-designed with and for children. 5, 12

• **Endurability** – Assessed the participants’ willingness to repeat the experience using a modified Again-Again table. 5, 12
METHODS FOR MEASURING COGNITIVE ENGAGEMENT

We used the following metrics to measure cognitive engagement:

- **Secondary task technique** – Measured participants’ response times to an auditory, visual, or tactile probe in order to gauge their attentional allocation to the primary activity. Longer response times suggest a higher level of attention to the primary task. 4, 12

- **Subjective perception of time** – Asked participants how long each activity lasted. Underestimation of the passage of time suggests higher engagement. 12, 7
METHODS FOR MEASURING BEHAVIORAL ENGAGEMENT

We used the following metrics to measure behavioral engagement:

- **Visual attention** – Measured the total amount of time spent looking at the screen with a stopwatch or eye-tracking equipment. ⁴

- **Off-task behavior** – Recorded any off-task behavior during the session. ²
STUDY GOALS

Our primary study goals were to:

• Determine whether the proposed research methods, namely eye-tracking, the secondary task technique, observation of affective signs, and self-report effectively measure engagement in children.

• Identify areas for improvement and refinement in the study protocol to aid future research efforts.
METHODS
RECRUITMENT
POTENTIAL STUDY POPULATIONS

The purpose of this pilot study was to develop a protocol that could be adapted for children with a variety of special needs and accessibility considerations including:

- Learning disabilities (LD)
- Autism spectrum disorder (ASD)
- Intellectual disability (ID)
- Attention deficit hyperactivity disorder (ADHD)
- English language learners (ELLs)

We attempted to recruit typically developing, ASD, ADHD, and ELL participants, but only tested the pilot protocol with typically developing children and children with ADHD due to recruitment limitations.
LEARNING DISABILITIES

For the purposes of this pilot study, we assumed the following about learning disabilities:

• Learning disabilities (LD) are a heterogeneous group of disorders characterized by a failure to adequately acquire, retrieve, and use information. 22

• LD can be caused by congenital or acquired abnormalities in brain structure and function and are often multifactorial in etiology. 22

• Captions have been shown to be beneficial in children with language-based learning disorders or learning disorders in reading. 14
For the purposes of this pilot study, we assumed the following about autism spectrum disorder:

- Autism spectrum disorder (ASD) is a neurodevelopmental disorder with deficits in social interaction and communication and restricted and repetitive activities, interests, and behaviors. 23
- Some children have associated intellectual impairment, and usually their verbal skills are weaker than their non-verbal skills. 23
- There is wide variability in language impairment, but generally the delay in receptive language is more pronounced than the delay in expressive language. 23
- Asperger syndrome is a subtype of ASD that is characterized by deficits in reciprocal social interaction, stereotyped patterns of behavior, and the absence of clinically significant delay in language or cognitive function. 23
- Rate of comorbid psychiatric diagnoses in children with Asperger syndrome and high-functioning autism is 74%. 23
INTELLECTUAL DISABILITY

For the purposes of this pilot study, we assumed the following about intellectual disability:

• Intellectual disability (ID) is a neurodevelopmental disorder characterized by deficits in intellectual and adaptive functioning that affects 1% of the population. 24

• ID was previously referred to as mental retardation (MR). 24

• These children have difficulty with learning, reasoning, problem-solving, abstract thinking, and judgment and their IQs are two standard deviations or more below the mean or less than 65 to 75. 24

• In children younger than 5 years old, failure to meet the expected developmental milestones in several areas of intellectual functioning is called global developmental delay (GDD). 24

• In children without hearing impairment, language development is often an indicator of future intelligence. 24
ATTENTION DEFICIT HYPERACTIVITY DISORDER

For the purposes of this pilot study, we assumed the following about attention deficit hyperactivity disorder:

• Attention deficit hyperactivity disorder (ADHD) is a psychiatric condition characterized by hyperactivity, impulsivity, and/or inattention. 25

• Children with the predominantly hyperactive subtype are unable to sit still or inhibit their behavior. These symptoms usually manifest by age 4 and peak at ages 7 to 8. 25

• Children with the predominantly inattentive subtype have difficulty focusing their attention and have slower cognitive processing, often appearing to be daydreaming or “off task.” Inattentive symptoms usually become apparent by ages 8-9. 25
ENGLISH LANGUAGE LEARNERS

For the purposes of this pilot study, we assumed the following about English language learners:

• Many English language learners (ELLs) are currently or have been previously enrolled in an English language instruction educational program. ¹⁷

• Immigrant students are defined as children who have been enrolled in a school in the U.S. for less than or equal to 3 academic years. ¹⁷
**RECRUITMENT GOALS**

When recruiting study participants, we sought out:

Two typically-developing participants:

- Ideally one typically-developing participant and one ELL participant
- Mix of genders
- Mix of ages between 6 and 8 years old
- At or above a first-grade reading level

Two children with neurodevelopmental or psychiatric disorders:

- Ideally one participant with ADHD (predominantly inattentive subtype) and one participant with high-functioning ASD
- Mix of genders
- Mix of ages between 6 and 8 years old
- At or above a first-grade reading level
SCREENING FOR TYPICAL DEVELOPMENT

We screened for typically-developing children who do not formally require closed captioning to access content by assessing whether or not prospective participants met the appropriate developmental milestones for 6 to 8 year olds across several key domains including:

• Speech/language
• Social
• Behavioral/emotional
• Adaptive functioning

Screener questions were informed by an internal literature review and an interview with a developmental pediatrician.
SAMPLE QUESTIONS

Speech/language:

• How often does your child speak in full sentences and ask complex questions that begin with “who, what, where, when, and why?” (expressive language)

• How often does your child seem to understand what you’re saying? (receptive language)

Social:

• How often is your child able to easily approach other children and engage in back-and-forth conversation with them?

Behavioral/emotional:

• How often does your child have difficulty maintaining attention to what needs to be done or listening when spoken to directly? (attention)

• How often does your child fidget with his or her hands or feet and squirm when seated? (hyperactivity)

Adaptive functioning:

• How often is your child independent in his or her daily activities (i.e., able to brush his or her teeth and perform basic household chores, etc.)?
SCREENING FOR ASD

When screening for children with ASD, we wanted to eliminate participants with co-morbid conditions or significant intellectual impairment to avoid confounding factors, but given the wide variability of language dysfunction in ASD and the known benefits of closed captioning in children with language-based learning disorders, we selected for children with mild language impairment.

Sample questions for delayed receptive language:

• Does your child respond when you call his or her name? For example, does he or she look up, talk or babble, or stop what he or she is doing?

• Does your child understand when you tell him or her to do something? For example, if you don’t point, can your child understand simple commands such as “put the book on the chair” or “bring me the blanket?”
SCREENING FOR ADHD

When screening for children with ADHD, we wanted to select for children with predominantly inattentive symptoms and eliminate those with predominantly hyperactive symptoms in order to maximize our chances of collecting usable data. Screener questions were based on the Vanderbilt ADHD Diagnostic Rating Scale, a parent/teacher rating scale for determining ADHD subtype that comprises 9 questions for inattention and 9 questions for hyperactivity. A score of 6/9 in either domain is positive for ADHD. 16

Sample question for inattention:

• How often does your child follow through when given directions and finish activities?

Sample question for hyperactivity:

• How often does your child have difficulty playing quietly?
SCREENING FOR ENGLISH LANGUAGE LEARNERS

When screening for ELLs, we wanted to select for children whose primary language is Spanish and who are currently enrolled in an ELL program.

Sample questions:

• Does your child generally understand directions and questions in Spanish?

• Does your child generally respond in Spanish?
PILOT SCREENER

Please refer to the screener for additional information.
STUDY DESIGN
LAB SETUP
KEY CONSIDERATIONS FOR LAB SETUP

We outfitted our usability lab to be more child-friendly, and the pilot suggested some guidelines for future or follow-on lab setup:

• **Use standard office lighting** because some children with ASD are hypersensitive to visual stimuli and testing in a darkened room heightens the relative brightness of the screen. Additionally, darkness can increase pupil dilatation, which can impede tracking with certain systems.  

• **Strike a balance** between a child-friendly environment and relatively sparse décor to make participants comfortable, but minimize the risk of directing the child’s attention away from the screen.  

• **Allow participants’ parents to remain** with them throughout the testing session and give them ample time to become acclimated to the lab environment.  

• **Use a child-size, non-rotating, height-adjustable chair** to ensure comfortable, ergonomic seating and minimize participant movement. The chair seat should not rotate; kids wiggle.
PROTOCOL OVERVIEW

We invited children to our usability lab where a professional moderator facilitated the study activities.

• Each participant watched a 9-minute video with captions, a different 9-minute video without captions, and played the same game with and without captions for 3 minutes.

• While participants engaged in these activities (watching and playing), we asked them to engage in a secondary task (described on the following page).

• We gave participants a rest break between activities, the second of which was a snack break where participants were given parent-approved foods.

• During each activity, we recorded affective signs of engagement and disengagement, response times to auditory probes, and eye-tracking data to measure visual attention.

• After each activity, we interviewed participants to assess self-reported enjoyment and comprehension.
**TASK ORDER**

<table>
<thead>
<tr>
<th>Participant #</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Break</th>
<th>Task 3</th>
<th>Task 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Marble Mover video with captions</td>
<td>Which Fish game</td>
<td>Snack</td>
<td>Which Fish game with captions</td>
<td>Snow Daze video</td>
</tr>
<tr>
<td>P2</td>
<td>Marble Mover video</td>
<td>Which Fish game with captions</td>
<td>Snack</td>
<td>Which Fish game</td>
<td>Snow Daze video with captions</td>
</tr>
<tr>
<td>P3</td>
<td>Snow Daze video</td>
<td>Which Fish game with captions</td>
<td>Snack</td>
<td>Which Fish game</td>
<td>Marble Mover video with captions</td>
</tr>
<tr>
<td>P4</td>
<td>Snow Daze video with captions</td>
<td>Which Fish game</td>
<td>Snack</td>
<td>Which Fish game with captions</td>
<td>Marble Mover video</td>
</tr>
</tbody>
</table>

* Participants were randomized to different task orders, all of which alternated between captioned and uncaptioned media to minimize the potential bias of viewing order and content-based viewing preferences.
SECONDARY TASK TECHNIQUE OVERVIEW

Psychological research indicates that the secondary task technique is a useful and well-documented method for measuring attention.

• The secondary task technique entails instructing participants to perform a primary task (i.e., attending to media content) while simultaneously completing a secondary task (i.e., pressing a button in response to an auditory, visual, or tactile reaction time probe). 4, 20

• In theory, the more attention a participant devotes to the primary task, the less attention they have available for the secondary task. This increase in mental capacity allocated to the primary task is reflected in slower response times to the secondary task (i.e., the longer their response time, the more focused they were on the primary stimulus). 4

• For the purposes of this study, the primary stimuli were *Peep and the Big Wide World* videos and online game and the secondary stimulus was an audio chime. The secondary task was to silence the auditory probe by pressing the space bar on a nearby keyboard.
REACTION TIME MEASUREMENT TOOL

Provision of the secondary stimuli with timed capture of the participants’ responses was accomplished through the creation of a unique reaction time measurement program by one of Mad*Pow’s in-house developers.

- This tool was automatically synchronized with the start of the primary stimuli (videos and online game) and played a chime at key intervals during the viewing/play experience. The timing was designed to appear random to individual participants, but was consistent across participants.

- The tones were silenced when the participants pressed the space bar and the tool captured their response times, the intervals between tone generation and depression of the space bar, for each secondary stimulus. The response times were recorded in milliseconds on a separate server for later analysis.
SECONDARY TASK TECHNIQUE PROTOCOL

We adapted the methodology used by Anderson *et al.* in their research on attentional allocation in preschoolers watching Sesame Street episodes. 4

- Participants were presented with 15 auditory probes per video segment and 6 auditory probes per game at 10-60 second intervals.

- The moderator provided instruction to the participant regarding the primary and secondary tasks. Participants were allowed to ask questions at any time and each segment contained a practice probe that was not scored.

- Participants were not prompted regarding the secondary stimuli (audio tones) once the primary stimulus (video or game) was activated.

- We measured the response times in milliseconds for optimal precision.
### Video Probe Timings

<table>
<thead>
<tr>
<th>Probe</th>
<th>Time</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>0:02</td>
<td>32 s</td>
</tr>
<tr>
<td>1</td>
<td>0:34</td>
<td>55 s</td>
</tr>
<tr>
<td>2</td>
<td>1:29</td>
<td>45 s</td>
</tr>
<tr>
<td>3</td>
<td>2:14</td>
<td>17 s</td>
</tr>
<tr>
<td>4</td>
<td>2:31</td>
<td>50 s</td>
</tr>
<tr>
<td>5</td>
<td>3:21</td>
<td>35 s</td>
</tr>
<tr>
<td>6</td>
<td>3:56</td>
<td>15 s</td>
</tr>
<tr>
<td>7</td>
<td>4:11</td>
<td>25 s</td>
</tr>
<tr>
<td>8</td>
<td>4:36</td>
<td>48 s</td>
</tr>
<tr>
<td>9</td>
<td>5:24</td>
<td>12 s</td>
</tr>
<tr>
<td>10</td>
<td>5:36</td>
<td>58 s</td>
</tr>
<tr>
<td>11</td>
<td>6:34</td>
<td>28 s</td>
</tr>
<tr>
<td>12</td>
<td>7:02</td>
<td>19 s</td>
</tr>
<tr>
<td>13</td>
<td>7:21</td>
<td>50 s</td>
</tr>
<tr>
<td>14</td>
<td>8:11</td>
<td>15 s</td>
</tr>
<tr>
<td>15</td>
<td>8:26</td>
<td></td>
</tr>
</tbody>
</table>

### Game Probe Timings

<table>
<thead>
<tr>
<th>Probe</th>
<th>Time</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>0:02</td>
<td>18 s</td>
</tr>
<tr>
<td>1</td>
<td>0:20</td>
<td>32 s</td>
</tr>
<tr>
<td>2</td>
<td>0:52</td>
<td>43 s</td>
</tr>
<tr>
<td>3</td>
<td>1:35</td>
<td>22 s</td>
</tr>
<tr>
<td>4</td>
<td>1:57</td>
<td>15 s</td>
</tr>
<tr>
<td>5</td>
<td>2:12</td>
<td>45 s</td>
</tr>
<tr>
<td>6</td>
<td>2:57</td>
<td></td>
</tr>
</tbody>
</table>
SAMPLE PARTICIPANT INSTRUCTIONS

The following is an example of the task prompts our moderator gave to participants before they watched the videos or played the online game. The italicized text represents instructions for the moderator, and was not shared with the participants.

**Moderator verbal prompt:** “I’m going to show you [Marble Mover or Snow Daze] on the screen now and I want you to watch and enjoy the show. While you’re watching the show, I’m going to have you play a game. Once in a while, you’re going to hear a chime, a noise that sounds like a bell, and it’ll be your job to turn off the chime by pressing the space bar. I’d like you to press the space bar as fast as you can after you hear the chime. We’re going to practice doing that now. Put your finger on the space bar and listen for the first chime.”

**Moderator procedural instruction:** Present the first auditory probe and praise the participant for responding quickly or encourage them to respond if they hesitate. Repeat this sequence a few times and then present the probes at different intervals while casually chatting with the participant. Continue to present auditory probes until the participant is responding consistently.
EYE-TRACKING OVERVIEW

Eye-tracking data provided objective measures of participants’ visual attention and gaze patterns.

• There are two types of eye movements: fixations and saccades. Fixations are points of regard lasting for at least 100 milliseconds and saccades are rapid eye movements between points of regard. People process information during fixations and do not process information during saccades. ¹³

• Children with developmental disabilities may have more missing gaze data than their neurotypical counterparts because they are often less attentive and more prone to blinking, but we can control for these confounding factors by calculating the time spent looking at captions as a proportion of the total gaze time onscreen. ⁹
EYE-TRACKING PROTOCOL AND METHODOLOGY

We used the eye-tracking equipment described below during the pilot study in the following ways:

- We used the EyeTech VT3 mini, which has an operating distance of 40-75 cm, a sampling rate of 60 Hz, and a head movement range of 31.5 x 22.5 cm.

- We attached the eye tracker to a desktop computer through its magnetic mount and positioned the participant at arm’s length from the screen in a child-size chair.

- We used a 5-point calibration sequence, which is brief enough to retain attention and accurate enough to ensure usable gaze data.  

- Participants were positioned so that their eyes were in the middle of the display window to maximize data capture in the event of straightening, slouching, or swaying during testing.

- We intermittently adjusted the eye tracker angle and chair height and placement during testing to attempt to accommodate changes in participant position.
KEY CONSIDERATIONS FOR EYE-TRACKING EQUIPMENT

Our experiences during the pilot study uncovered equipment-related limitations and suggested guidelines for selecting eye-tracking systems to better support the study protocol in later use.

• Eye-tracking systems should be unobtrusive, account for head motion, and not actively restrict participant movement. Kids wiggle. 9

• A table-top eye tracker or model integrated within the display monitor may be preferable because children may resist chin-rest or head-mounted systems that restrict their head movement or require them to wear equipment. 9

• Choose an appropriate sampling rate. Most eye-tracking systems have a minimum sampling rate of 50 Hz, which is generally appropriate for assessing children’s gaze patterns as they scan static images or watch dynamic content. The sampling rate should be high enough to address the research question, but low enough to accommodate head movement. Higher sampling rates often compromise freedom of head movement. 9
POST-ACTIVITY ENGAGEMENT QUESTIONS

Reported fun:

- What did you think of this video? (Smileyometer)

![Smileyometer]

Endurability:

- Would you like to watch it again? (modified Again-Again table)

Subjective perception of time:

- How long were you watching the video?
POST-ACTIVITY COMPREHENSION QUESTIONS

• What object does Peep find in the beginning of the show?
• Who helps Peep move this object to the top of the hill?
• What does Quack want Peep and Chirp to help him build?
• What do Peep, Chirp, and Beaver Boy discover how to do?
ADDITIONAL CONSIDERATIONS

- The auditory probe in the secondary task technique could be replaced with a visual probe in larger-scale studies. The auditory probe was continuous; the sound persisted until the participant turned it off by pressing the space bar. A visual probe would likely be momentary; it would last for a predetermined length of time (e.g., five seconds) and then disappear.

- The visual probe could be a child-friendly image, such as a popular cartoon character or animal, that is embedded in the video and appears intermittently on the screen in no discernable pattern (e.g., every 10-60 seconds), but is large enough that the participant sees it every time to ensure that data is collected for each probe.

- Additionally, this study protocol could be modified to test a wide range of accommodations beyond closed captioning, including audio description and tactile enhancement.
Please refer to the moderator’s guide for additional information regarding the study protocol.
RESULTS
METHODOLOGICAL PILOT STUDY RESULTS

The piloted methodology appears to support capture of the following data points for affective engagement:

- Tally of participants’ signs of engagement (e.g., smiling, laughing, concentrating, excitable bouncing, positive vocalizations, and leaning forward), signs of disengagement (e.g., frowning, signs of boredom such as fiddling or playing with their ears, shrugging, negative vocalizations, sighing, yawning, and turning away from the screen), and responses to the enjoyment and endurability questions to obtain an overall sense of their affective engagement and contextualize their eye-tracking and reaction time data.
METHODOLOGICAL PILOT STUDY RESULTS

The piloted methodology appears to support capture of the following data points for the secondary task technique and eye-tracking:

Secondary task technique:

- Individual response times (measured in milliseconds) to each auditory probe and average response times for segments.

Eye-tracking:

- Time spent onscreen for all segments (video, video with captions, game, game with captions), time spent on captions for captioned segments (video with captions and game with captions), fixation count for screen for all segments, fixation count for captions for captioned segments, and time spent on captions/time spent onscreen (% screen time spent looking at captions) for captioned segments.
SAMPLE REACTION TIME DATA

The chart below shows one participant’s reaction times during the secondary task. The chart data reflect the response times between the participant hearing (or noticing) the secondary task audio prompt and pressing the space bar to silence it. Our working hypothesis was that reaction times may be indicative of engagement.

The pilot study’s use of a bespoke coded tool for playing an audio prompt and recording participants’ response times indicates that capture of this data is possible.

<table>
<thead>
<tr>
<th>Tasks Recorded</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1500925757614-Marble_Mover[cc]</td>
<td></td>
<td>987ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500926181422-WhichFish[cc]</td>
<td></td>
<td>1162ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500927418299-WhichFish[cc]</td>
<td></td>
<td>1487ms</td>
<td>1380ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15009988325-Marble_Mover[cc]</td>
<td></td>
<td>15392ms</td>
<td>1628ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500999996881-WhichFish[cc]</td>
<td></td>
<td>1120ms</td>
<td>1174ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150091271265-SnowDaze[cc]</td>
<td></td>
<td>1186ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150091327337-SnowDaze</td>
<td></td>
<td>1322ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150091426328-SnowDaze[cc]</td>
<td></td>
<td>1456ms</td>
<td>1658ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15009211320-Marble_Mover[cc]</td>
<td></td>
<td>1596ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150099378143-WhichFish[cc]</td>
<td></td>
<td>1560ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500996076104-WhichFish</td>
<td></td>
<td>1668ms</td>
<td>2043ms</td>
<td>4823ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500997102029-Marble_Mover[cc]</td>
<td></td>
<td>21ms</td>
<td>3451ms</td>
<td>55ms</td>
<td>20578ms</td>
<td>2103ms</td>
<td>4065ms</td>
<td>1677ms</td>
<td>10026ms</td>
<td>214ms</td>
<td>4106ms</td>
<td>575ms</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SAMPLE EYE-TRACKING GAZE DATA

In the sample eye-tracking screenshots below, the yellow lines indicate participants’ high-velocity eye movements (saccades), while the yellow circles indicate points at which they looked for at least 100 milliseconds (fixations). The higher the number within the circle, the more time the participant spent looking at that area of the screen.
SAMPLE EYE-TRACKING HEAT MAP

Eye-tracking data analysis provided “heat maps” to show, in aggregate, what participants looked at onscreen the most.

Greater concentrations of color indicate areas of focus, with red indicating where they spent the most time looking.
SAMPLE EYE-TRACKING AREAS OF INTEREST

In the sample eye-tracking screenshots below, study moderators were able to define "areas of interest," for which the eye-tracking software provided separate data points. For this study, we created an area of interest that encompassed the onscreen captions to help assess when and for how long participants looked at the captions.
LESSONS LEARNED
SECONDARY TASK TECHNIQUE

Some participants pressed the space bar repeatedly in between auditory probes or depressed it continuously for portions of the study, thereby disrupting the reaction time measurement program. Potential refinements for the secondary task technique apparatus include:

• Asking the participants to complete a more complex action instead of simply pressing the space bar (e.g., placing a coin or token into a bucket in response to each probe).

• Explicitly instructing the participants to only press the space bar in response to the probe.

• Increasing the interval between auditory probes to allow the participants adequate time to respond. We found that intervals of 10-15 seconds were too short and recommend intervals of 25-30 seconds or longer.

• Using a dedicated accessibility switch shaped like a traditional, round button instead of the space bar.

• Using a visual probe (e.g., cartoon character that appears onscreen) instead of an audio tone.
ACCOUNTING FOR MOVEMENT

Some participants, particularly those with ADHD, squirmed, fidgeted, and changed positions frequently during the activity. Potential workarounds include:

- Using a non-mobile, non-rotating, ergonomic, height-adjustable, child-size chair that provides support and constrains movement (e.g., Stokke Tripp Trapp High Chair).

- Consider using the Tobii X-2 30 eye tracker to accommodate a greater range of head movement.
RECRUITMENT

When recruiting study participants using this protocol:

• Attempt to achieve an even mix of genders within each study population to avoid gender-based confounding variables (e.g., boys with ADHD are more likely to exhibit hyperactive symptoms than girls with ADHD).

• Peep content is designed for 3 to 5 year olds, but because we were testing closed captioning, we recruited older children between the ages of 6 and 8 years old who were able to read. Some participants grew bored with the content, in part because it was not age-appropriate. In larger-scale studies, consider testing content intended for older, school-age children.
CONCLUSION
**METHODOLOGICAL SUCCESS**

The tested methodology is promising and a larger-scale study addressing attention and learning is possible and may result in significant contributions to the existing academic literature on the benefits of closed captioning in children for whom it is not formally required.

This protocol and all related materials can be used, with necessary adjustments, to study a variety of learning styles and accommodations beyond closed captioning.
APPENDIX

METHODS SUMMARY TABLE
REFERENCES
<table>
<thead>
<tr>
<th>Method</th>
<th>Metrics</th>
<th>Capture Mechanism</th>
<th>Pass/Fail</th>
<th>Implications at Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye-tracking</td>
<td>Total time spent onscreen, total time spent on captions, fixation count for screen, fixation count for captions, and % screen time spent looking at captions</td>
<td>Automated eye-tracking</td>
<td>Pass = eye-tracking data captured</td>
<td>More time spent looking at the screen and/or captions suggests greater visual attention and higher behavioral engagement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fail = no eye-tracking data captured (e.g., too much head movement)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>More time spent looking at the screen and/or captions suggests greater visual attention and higher behavioral engagement</td>
<td></td>
</tr>
<tr>
<td>Secondary task technique</td>
<td>Average response times to auditory probes</td>
<td>Reaction time measurement tool</td>
<td>Pass = individual response times recorded</td>
<td>Longer average response times suggests higher cognitive engagement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fail = no response times recorded (e.g., participant never pushed the button)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Partial fail = moderator judgment that absence of response means participant is paying rapt attention</td>
<td></td>
</tr>
<tr>
<td>Observation of affective signs</td>
<td>Tally of signs of engagement (e.g., smiling, laughing, concentrating, excitable bouncing, positive vocalizations, and leaning forward)</td>
<td>Recorded video footage and moderator notes</td>
<td>Pass = observation of positive or negative affective signs</td>
<td>More signs of engagement than disengagement suggests higher affective engagement</td>
</tr>
<tr>
<td></td>
<td>Tally of signs of disengagement (e.g., frowning, signs of boredom such as fiddling or playing with their ears, shrugging, negative vocalizations, sighing, yawning, and turning away from the screen)</td>
<td></td>
<td>Fail = observation of neutral affective signs</td>
<td></td>
</tr>
<tr>
<td>Self-report</td>
<td>Responses to enjoyment (Smileyometer) and endurability (Again-Again) questions</td>
<td>Interview questions</td>
<td>Pass = questions answered</td>
<td>Positive responses to enjoyment and endurability questions suggests higher affective engagement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fail = questions not answered</td>
<td></td>
</tr>
<tr>
<td>Comprehension assessment</td>
<td>Responses to video-specific content questions</td>
<td>Interview questions</td>
<td>Pass = questions answered</td>
<td>Correct responses to comprehension questions suggests increased learning</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fail = questions not answered</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


REFERENCES


