Developmental Psychology

Investigating the Validity of Picture-Based Assessments Across Cultures and Contexts: Evidence From Young Children in Kenya and the United States

Rebecca Zhu, Tabitha Nduku Kilonzo, Jan M. Engelmann, and Alison Gopnik

Online First Publication, August 4, 2025. https://dx.doi.org/10.1037/dev0002050

AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

https://doi.org/10.1037/dev0002050

Investigating the Validity of Picture-Based Assessments Across Cultures and Contexts: Evidence From Young Children in Kenya and the United States

Rebecca Zhu[1], Tabitha Nduku Kilonzo[2], Jan M. Engelmann[3], and Alison Gopnik[3]

[1] Department of Psychology, Stanford University

[2] International Education, RTI International, Nairobi, Kenya

[3] Department of Psychology, University of California, Berkeley

Many childhood assessments rely on picture stimuli, but children in diverse early environments possess varying amounts of experience with pictures. Two preregistered experiments, conducted in 2022–2023, investigated whether picture assessments are valid across diverse contexts. Low-to-middle-income children (n = 192, 2–7 years, 85 females, all Black) in their first month of formal schooling in Mombasa County, Kenya, an early environment with relatively few pictures, performed more accurately on an object vocabulary task than a picture vocabulary task ($\beta = 0.07$, $p < .001$; Experiment 1). Middle-to-high-income children ($n = 96$, 2–3 years, 52 females, predominantly White and Asian) in the San Francisco Bay Area, an early environment with relatively more pictures, performed similarly on object and picture vocabulary tasks ($\beta = 0.02$, $p = .60$; Experiment 2). Consequently, these results tentatively suggest that assessments involving pictures may *underestimate* children's capacities in some contexts. To accurately measure developing capacities in children from diverse backgrounds, it is critical that assessment tools are appropriately adapted to environmental contexts.

Public Significance Statement

Researchers frequently use picture-based assessments across cultures and contexts. We found that preschoolers in a Kenyan context (Mombasa County) were more accurate on an object-based vocabulary assessment than a picture-based vocabulary assessment, whereas toddlers and preschoolers in an American context (San Francisco Bay Area) were equally accurate on an object-based vocabulary assessment and a picture-based vocabulary assessment. These findings tentatively suggest that picture-based assessments may sometimes underestimate children's capacities, particularly in some global contexts where children may have relatively little experience with pictures.

Keywords: cognitive development, pictures, vocabulary, assessments, global contexts

The most widely used assessments of early cognitive and socioemotional abilities—such as the Bayley Scales of Infant and Toddler Development, Denver Developmental Screening Test, Peabody Picture Vocabulary Test, and Wechsler Preschool and Primary Scale of Intelligence—were all initially developed in high-income Western contexts (Fernald et al., 2017). These assessments often involve picture stimuli that may be less frequent or entirely unfamiliar to children in other cultures and contexts (Callaghan et al., 2011, 2012; Walker et al., 2013). It is unknown how children growing up in diverse global contexts might interpret

Lana Karasik served as action editor.

Rebecca Zhu https://orcid.org/0000-0003-0957-5962

The study materials, anonymized data, and R code are available on the Open Science Framework (https://osf.io/xupy4/). The authors declare that there are no conflicts of interest.

This work was supported by a Natural Sciences and Engineering Research Council of Canada Post-Graduate Doctoral Fellowship (Grant 532517-2019) and a John L. Simpson Research Fellowship in International Studies to Rebecca Zhu, as well as the Center for Effective Global Action, the Jacobs Foundation, the Aga Khan Foundation, and the Schools2030 Program. The authors are grateful to Helen Pitchik and Lia Fernald for helpful input throughout the research process. They are also grateful to Diana Chanya, Florence Ndungu, Fred Ochieng, Habona Jillo, Hector Juma, Juma Mohamed, and Pauline Masese for assistance with data collection in Mombasa County, Kenya, and to Mina Chang, Brooke Quynh Ho, Tiara Mckinney, Niki Parker, Audrey Kaufman, Lily Zihui Zhu, and Sasha Zitter for assistance with data collection in Berkeley, United States. Moreover, the authors are thankful for the extremely talented Zoe Carwin for assistance with stimuli creation. Thanks also to the parents and children who made this research possible.

Rebecca Zhu played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, supervision, visualization, writing–original draft, and writing–review and editing. Tabitha Nduku Kilonzo played a supporting role in conceptualization, investigation, methodology, supervision, and writing–review and editing. Jan M. Engelmann played a supporting role in conceptualization, funding acquisition, methodology, and writing–review and editing. Alison Gopnik played a supporting role in conceptualization, methodology, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Rebecca Zhu, Department of Psychology, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305, United States. Email: ryz@stanford.edu

these kinds of infrequent or unfamiliar stimuli, and it is possible that having differing interpretations of these kinds of stimuli might consequently change the validity of assessment scores (Cromwell et al., 2014; Jukes et al., 2024; Ranjitkar et al., 2018).

Previous research with North American convenience samples shows that toddlers and preschoolers growing up in relatively high-income Western contexts possess a sophisticated understanding of pictures. U.S. toddlers understand that pictures are simultaneously objects—for example, a small, thin photo—as well as representations—for example, a depiction of a car (DeLoache et al., 1998; Preissler & Bloom, 2007). U.S. 2- and 3-year-olds are also sensitive to the artists' intentions when interpreting the content of a picture (Bloom & Markson, 1998; Preissler & Bloom, 2008). Moreover, 3- and 4-year-olds can form abstract inferences from sparse visual representations, for example, by labeling a big circle as an elephant and a small circle as a mouse (Bloom & Markson, 1998). Given this well-established body of literature demonstrating that relatively high-income Western children possess an early-emerging capacity to understand pictures, assessments involving pictures are likely to be valid in these high-income Western contexts, even for relatively young children. For example, the most recent version of the Peabody Picture Vocabulary Test (Dunn, 2019) is deemed appropriate for children ages 2.5 years and up.

It is unclear to what extent the development of picture comprehension relies on early experience with pictures. Some research suggests that recognizing the content of pictures may be an intuitive cognitive capacity: For example, 3-month-olds can perceive facial expressions on screens (Barrera & Maurer, 1981), suggesting that infants can already understand the content of visual representations. Additionally, psychologists Hochberg and Brooks (1962) did not expose their own child to visual representations until 19 months of age, but their child was still able to recognize and name objects in pictures upon his very first encounters. However, researchers have also argued that the development of picture comprehension depends in part on picture experience (Callaghan, 2020). In one study, Pirahã adults, who live a hunter–gatherer lifestyle in Brazil and have limited exposure to modern visual media, showed some difficulty recognizing the content of two-tone black-and-white images (Yoon et al., 2014). In an autobiographical account, a Maasai warrior originating from rural Tanzania detailed his initial confusion when first encountering black-and-white drawings and movies (Saitoti, 1986).

Consequently, children who have different understandings of pictures might respond differently to information conveyed through pictures. Just as it is inappropriate to assess the vocabulary of English-speaking American children with instructions in Swahili (i.e., a language they are likely unfamiliar with), it may also be inappropriate to assess the cognitive capacities of children growing up in picture-sparse environments with experimental paradigms involving pictures (i.e., a medium they are unfamiliar with). Differences in picture comprehension may drastically change the validity of assessments in global contexts (Zhu, Kilonzo, et al., 2025; Zuilkowski et al., 2016). Indeed, some early studies show that children growing up in picture-sparse contexts perform more accurately on tasks involving objects than on tasks involving pictures. Toddlers living in a rural village in the Kibaha-Pwani District of Tanzania's Coast Region—an early environment with few pictures—successfully learned a novel word for a novel object, but did not learn a novel word for a photograph of the novel object (Walker et al., 2013). Similarly, preschoolers living in picture-sparse rural environments in India (i.e., a village 70 km from Vijayawada, Andhra Pradesh) and Peru (i.e., a village in the rural Montaro Valley area of the Central Highlands) performed more accurately on a false belief task involving objects than on a false belief task involving black-and-white line drawings (Callaghan et al., 2012). However, due to the challenges of conducting cross-cultural research, these previous studies have relied on relatively small sample sizes: For example, Walker et al.'s (2013) studies with Tanzanian toddlers involved an average of 15 participants (ranging from 12 to 20 participants) in each age group, and Callaghan et al.'s (2012) research involved an average of 13.5 Peruvian child participants (ranging from 12 to 15 participants) and 14 Indian child participants (ranging from 11 to 17 participants) per age group. Thus, while these studies tentatively suggest that children who predominantly encounter pictures in school materials and assessment tools might perform less accurately on assessments that involve picture stimuli, more work with larger sample sizes is required to conceptually replicate and extend these findings. Moreover, while previous research has examined differences in children's performance on tasks involving object and photograph stimuli (Walker et al., 2013) and object and black-and-white line drawing stimuli (Callaghan et al., 2012), we investigate differences in children's performance on tasks involving object stimuli and another kind of picture stimuli frequently used in children's learning materials and assessments, namely colored cartoons.

The current research examined the validity of assessments involving pictures in two contexts, namely Mombasa County, Kenya, and the San Francisco Bay Area, United States. We assessed children's knowledge of nouns, number words, and color words, through a receptive vocabulary task involving either picture or object stimuli. Experiment 1 explored whether low-to-middle-income Mombasa County children in their first year of formal schooling (i.e., Preprimary 1)—who may generally possess less picture experience than their high-income Western counterparts—might perform more accurately on a receptive vocabulary task involving objects, compared to the same receptive vocabulary task involving pictures. Experiment 2 explored whether high-income children in the San Francisco Bay Area, United States—who possess substantially more picture experience—might perform equally accurately on a receptive vocabulary task involving objects and a receptive vocabulary task involving pictures. Overall, the current research aims to address the questions of whether picture assessments are valid in global contexts.

Experiment 1

In Experiment 1, we investigated whether low-to-middle-income children in their first month of formal schooling in Mombasa County, Kenya, might perform more accurately on an object vocabulary task than on a picture vocabulary task. If low-to-middle-income children in Mombasa County indeed perform more accurately on an object vocabulary task than on a picture vocabulary task, this might suggest that assessments involving pictures are not always valid, particularly for children growing up in low-to-middle-income global contexts who may possess relatively less experience with pictures. In contrast, if low-to-middle-income children in Mombasa County perform equally accurately on object and picture vocabulary tasks, this might

suggest that assessments involving pictures are valid across diverse early contexts, despite substantial variation in picture experience across these contexts.

Method

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow Journal Article Reporting Standards for quantitative research in psychology (Appelbaum et al., 2018). We preregistered the design and analyses for both studies at https://aspredicted.org/. All preregistrations, data, analysis code, and research materials are available on the Open Science Framework (https://osf.io/xupy4/). Data were analyzed using R (Version 4.2.0; R Core Team, 2021) and the packages *dplyr* (Version 1.1.4; Wickham et al., 2023), *ggplot2* (Version 3.5.1; Wickham, 2016), and *lsr* (Version 0.5.2; Navarro, 2015).

Participants

We preregistered a stopping rule of 192 participants. Consequently, we worked with 192 preschoolers ($M = 4.52$ years, $SD = .95$ years, range = 2.38–7.48 years, 85 females and 107 males) in 12 schools in five subcounties (i.e., Nyali, Mvita, Jomvu, Likoni, Kisauni) of Mombasa County, Kenya. These 12 schools participate in the Schools2030 Program, a diverse coalition of educators, school leaders, researchers, international organizations, and governments across 10 countries, working toward the goal of supporting and improving students, teachers, and education systems worldwide (Schools2030, 2023). Within the sample of 192 preschoolers, there was substantial variation in both maternal education levels ($M = 8.10$ years of formal schooling, $SD = 5.07$ years, range = 0–24 years) and amount of picture experience, as measured by the eight-item picture experience questionnaire ($M = 47.60\%$, $SD = 28.08\%$, range = 0%–100%).

In a between-subjects design, half of the sample participated in the object condition (i.e., a vocabulary task involving objects) and the other half of the sample participated in the picture condition (i.e., a vocabulary task involving pictures). Researchers tested four additional children, whose data were excluded due to experimenter error. Children participated in a quiet spare room (e.g., the library) in their school. All preschoolers spoke Swahili as their native language and were tested by local research assistants fluent in Swahili. The experiment was approved by a U.S. university's Committee for the Protection of Human Subjects, as well as multiple relevant Kenyan government authorities, including the Kenya Medical Research Institute; the National Commission for Science, Technology, and Innovation; and the local Mombasa County Ministry of Education. All parents of child participants provided informed consent. Children received small gifts (i.e., pencil and face mask) for participating. Since all children were tested in a school setting, local researchers suggested that pencils would be particularly appropriate and useful gifts. All data were collected in May 2022.

Stimuli and Procedure

The present research methods (i.e., direct child assessment and caregiver questionnaire) were developed in close collaboration between U.S. and Kenyan researchers. Local research assistants presented children with a Swahili vocabulary task involving three blocks comprised of three different kinds of words: nouns (i.e., toothbrush/mswaki, crayon/crayoni, hat/kofia, fork/uma), number words (i.e., one/moja, two/mbili, three/tatu, four/nne), and color words (i.e., red/nyekundu, blue/samawati, yellow/manjano, green/kijani). These three kinds of words were chosen to create variability in trial difficulty: Anecdotal conversations with local Mombasa research assistants, many of whom were also parents to young children themselves, suggest that low-to-middle-income children living in Mombasa County typically learn common nouns fairly early in development, number words slightly later in development, and color words only when they enter the formal school system. Swahili color words should be especially difficult because schools initially teach children color words in English rather than Swahili.

Each block had four trials, leading to a total of 12 trials overall. The order of the blocks, as well as the order of the trials within each block, was randomized across participants. On each trial, the local experimenter laid out an array of four items in a quadrant (see Figures 1–3) and prompted the participant to select one of the four items (e.g., "Nionyeshe mswaki," meaning "Show me a toothbrush" in Swahili). Within each block, the placement of the target item was counterbalanced across trials. For a full list of target and distractor items, see Table 1.

Caregiver Questionnaire

To gain more information on individual children's early environments and experiences with pictures, each child's caregiver completed a brief questionnaire through a 10- to 15-min in-person or phone interview with a local researcher. Caregivers provided additional information regarding multiple aspects of the children's environment through eight picture experience questions (e.g., "Have you or anyone else in the household looked at picture books with your child in the past three days?"; Does your house have pictures/posters/paintings/calendars on the wall?), four child-directed speech questions (e.g., "Do you ever encourage the child to talk while you listen?"; "Did you get time to hold your child and talk with affection yesterday?"), and four learning material questions (e.g., "Are materials that teach colors, sizes and shapes available for the child to play with?"; "Is a real or toy musical instrument available to the child?"). These questions were adapted from the Home Observation Measurement of the Environment Inventory (Bradley & Caldwell, 1984), a measure frequently used in public health research to assess children's home environments (e.g., Luoto et al., 2021). Caregivers also provided information about their socioeconomic status, measured through total years of maternal education (Antonoplis, 2023). Caregivers also provided additional information about their household materials and assets, which were not entered into the analyses. The full questionnaire can be found in Appendix A.

Results and Discussion

Demographics

There was a significant amount of variation (i.e., in terms of age, maternal education, and picture experience) across the entire sample of preschoolers. However, there were no significant differences in the children's demographics between the two experimental conditions. Specifically, there were no age differences between children in the object condition ($M = 4.54$ years, $SD = 1.00$ years, range = 2.65–7.48

Figure 1

Stimuli for a Noun Trial (i.e., "Show Me a Hat"; "Nionyeshe Kofia") in the Picture Condition (Left) and the Object Condition (Right)

Note. See the online article for the color version of this figure.

years) and children in the picture condition ($M = 4.50$ years, $SD = .91$ years, range = 2.38–6.95 years), $t(189) = .28$, $p = .78$. Similarly, there were no differences in maternal education levels between children in the object condition ($M = 8.13$ years of formal schooling, $SD = 5.63$ years, range = 0–24 years) and children in the picture condition ($M = 8.08$ years of formal schooling, $SD = 4.48$ years, range = 0–17 years), $t(189) = .07$, $p = .94$. Finally, there were no differences in levels of picture experience (i.e., as measured by the eight-item picture experience questionnaire) between children in the object condition ($M = 47.79\%$, $SD = 28.85\%$, range = 0%–100%) and children in the picture condition ($M = 47.40\%$, $SD = 27.44\%$, range = 0%–100%), $t(190) = .43$, $p = .67$. One caregiver in the object condition did not report their child's age, and another caregiver in the picture condition did not report maternal education levels.

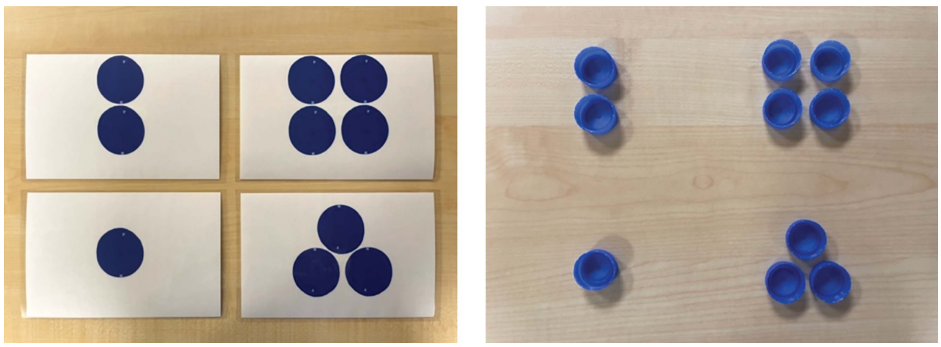
Accuracy

In a preregistered analysis, we found that children in the object condition had a higher proportion of correct responses compared to children in the picture condition ($\beta = 0.07$, $SE = .02$, $p < .001$). In additional exploratory analyses, we investigated whether children consistently had higher proportions of correct responses in the object condition compared to the picture condition, across multiple kinds of words. Children had a higher proportion of correct responses in the object condition than the picture condition for trials involving nouns ($\beta = 0.11$, $SE = .03$, $p < .001$) and number words ($\beta = 0.09$, $SE = .04$, $p = .02$). However, children's performance on trials involving color words did not differ between the object and picture conditions ($\beta = 0.01$, $SE = .02$, $p = .66$). All significant results remained significant at $p < .05$ after correcting for multiple comparisons (Benjamini & Hochberg, 1995). Consequently, children had higher proportions of correct responses in the object condition than the picture condition for the kinds of words they were familiar with (i.e., nouns and number words), but not for the kinds of words they were unfamiliar with (i.e., color words; see Figure 4).

In additional exploratory analyses, we also tested children's performance against chance levels. Since there were four possible answers on each trial, chance performance was 25%. Overall, children performed above chance levels in both the object condition, $M = 48.52\%$, $SE = 1.41\%$, $t(95) = 16.66$, $p < .001$, and the picture condition, $M = 41.49\%$, $SE = 1.48\%$, $t(95) = 11.18$, $p < .001$. Furthermore, we examined children's performance against chance levels for each kind of word. On the noun trials, children performed above chance levels in both the object condition, $M = 68.75\%$,

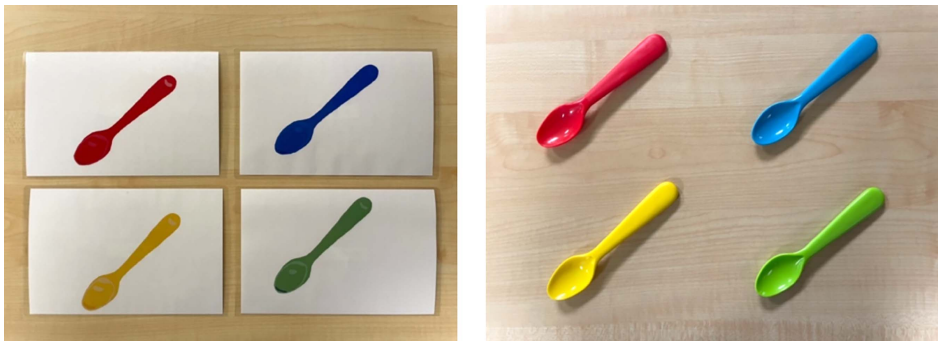
Figure 2

Stimuli for a Number Trial (i.e., "Show Me Three"; "Nionyeshe Tatu") in the Picture Condition (Left) and the Object Condition (Right)



Note. See the online article for the color version of this figure.

Figure 3
Stimuli for a Color Trial (i.e., "Show Me Blue"; "Nionyeshe Samawati") in the Picture Condition (Left) and the Object Condition (Right)



Note. See the online article for the color version of this figure.

$SE = 1.99\%$, $t(95) = 21.94$, $p < .001$, and the picture condition, $M = 57.55\%$, $SE = 2.29\%$, $t(95) = 14.24$, $p < .001$. Similarly, on the number trials, children performed above chance levels in both the object condition, $M = 50.26\%$, $SE = 2.88\%$, $t(95) = 8.77$, $p < .001$, and the picture condition, $M = 41.41\%$, $SE = 2.51\%$, $t(95) = 6.55$, $p < .001$. However, on the color trials, children performed at chance levels in both the object condition, $M = 26.56\%$, $SE = 1.61\%$, $t(95) = .97$, $p = .33$, and the picture condition, $M = 25.52\%$, $SE = 1.74\%$, $t(95) = .30$, $p = .76$. All significant results remained significant at $p < .05$ after correcting for multiple comparisons (Benjamini & Hochberg, 1995).

Side-Biasing

In additional exploratory analyses, we investigated the frequency of children's side-biasing behavior across the two conditions. Side-biasing behavior (i.e., choosing the same quadrant on all four trials within a block) typically occurs when young children find a task to be especially confusing or difficult. Since the placement of the target item was counterbalanced across trials within each block, side-biasing behavior did not skew children's overall accuracy on a task (e.g., a child who consistently picked the bottom-right quadrant would not be more or less accurate than 25% chance levels). However, side-biasing behavior may provide additional insight into which kinds of trials children found especially difficult.

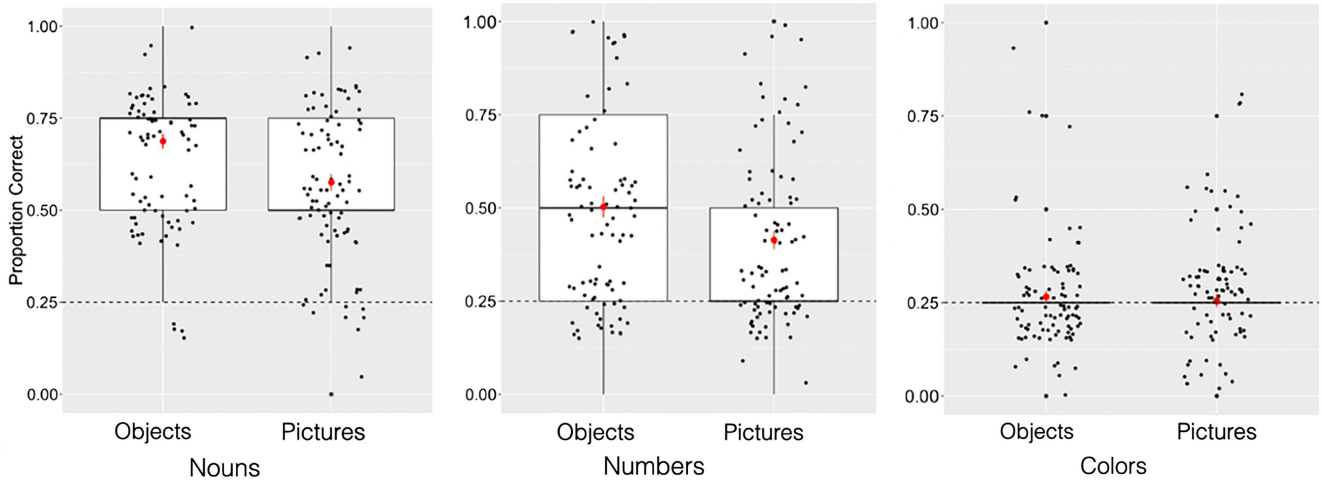
We found that children were significantly less likely to demonstrate side-biasing behavior in the object condition than the picture condition ($\beta = 0.11$, $SE = .06$, $p = .04$), suggesting that children found the object condition to be overall less difficult than the picture condition. Specifically, children in the object condition side-biased on 21% of the blocks, whereas children in the picture condition side-biased on 28% of the blocks. In line with other results, this overall side-biasing difference between conditions was driven only by the trials involving words that children had some familiarity with. In other words, children were less likely to side-bias in the object condition than the picture condition on the noun block, $\chi^2(1, N = 192) = 6.17$, $p = .01$, and the number block, $\chi^2(1, N = 192) = 6.63$, $p = .01$, but they side-biased equally frequently between the object condition and the picture condition in the color block, $\chi^2(1, N = 192) = 0.19$, $p = .66$. All significant results remained significant after correcting for multiple comparisons (Benjamini & Hochberg, 1995). On the noun block, preschoolers in the object condition side-biased on 1% of the blocks, whereas preschoolers in the picture condition side-biased on 10% of the blocks. On the number block, preschoolers in the object condition side-biased on 15% of the blocks, whereas preschoolers in the picture condition side-biased on 31% of the blocks. In contrast, on

Table 1
Full List of Stimuli

Prompt ("Show me …")	Target item	Distractor Item No. 1	Distractor Item No. 2	Distractor Item No. 3
A toothbrush	Toothbrush	Pencil	Scissors	Watch
A hat	Hat	Handkerchief	Shirt	Book
A crayon	Crayon	Pencil sharpener	Chalk	String
A fork	Fork	Knife	Ball	Mirror
One	One pebble	Two pebbles	Three pebbles	Four pebbles
Two	Two straws	One straw	Three straws	Four straws
Three	Three bottle caps	One bottle cap	Two bottle caps	Four bottle caps
Four	Four erasers	One eraser	Two erasers	Three erasers
Red	Red bowl	Blue bowl	Yellow bowl	Green bowl
Blue	Blue spoon	Red spoon	Yellow spoon	Green spoon
Yellow	Yellow cup	Blue cup	Red cup	Green cup
Green	Green plate	Blue plate	Yellow plate	Red plate

Figure 4

Data From Experiment 1: Proportion of Trials in Which Low-to-Middle-Income Mombasa County Preschoolers Chose the Correct Item, for Each Word Type



Note. Boxplot lower and upper hinges correspond to the 25th and 75th percentiles. See the online article for the color version of this figure.

the color block, preschoolers in the object condition side-biased on 48% of the blocks, whereas preschoolers in the picture condition side-biased on 44% of the blocks.

Picture Experience

In two preregistered secondary analyses, we investigated whether low-to-middle-income Mombasa preschoolers with relatively more picture experience performed more accurately on the picture task than low-to-middle-income Mombasa preschoolers with relatively less picture experience. In the first analysis, we operationalized picture experience as the total aggregated score from our eight-item picture experience questionnaire. We divided children into more experience (M = 69.66% on picture experience questionnaire) and less experience (M = 22.05% on picture experience questionnaire) groups using a median split across the entire sample along this aggregated picture experience variable, leading to a total of 50 more experience (M = 70.00% on picture experience questionnaire) and 46 less experience (M = 22.83% on picture experience questionnaire) children in the object condition and 53 more experience (M = 69.34% on picture experience questionnaire) and 43 less experience (M = 21.22% on picture experience questionnaire) children in the picture condition. We conducted a linear regression with accuracy as the dependent variable, and condition (i.e., object, picture), picture experience (i.e., more experience, less experience), and the interaction between condition and picture experience as between-subjects independent variables. Moreover, we controlled for child gender, maternal education, child-directed speech, and variety of learning materials in the home. There was a significant main effect of condition, such that children performed more accurately in the object condition than the picture condition ($\beta = 0.71$, $SE = .35$, $p = .04$). However, there was no effect of picture experience ($\beta = 0.30$, $SE = .37$, $p = .40$) or interaction between condition and picture experience ($\beta = 0.22$, $SE = .48$, $p = .65$). There was also an effect of child gender ($\beta = 0.64$, $SE = .24$, $p = .009$), such that girls' responses (M = 47.65%, SE = 2.00%) were significantly more accurate than boys' responses (M = 42.91%, SE = 1.60%). There was also a small but significant effect of maternal education ($\beta = -0.07$, $SE = .03$, $p = .005$), such that higher maternal education levels were related to lower child accuracy on the experimental task. All other variables were not significant (see Appendix Table B1 for full results).

In the second analysis, we conducted a similar linear regression using a slightly different picture experience variable, specifically looking at picture books. Since previous research suggests that looking at picture books, rather than the mere presence of picture books, may relate to children's ability to learn from pictures (Zhu, Pitchik, et al., 2025), we conducted another analysis operationalizing picture experience as whether children had recently looked at picture books with an adult or older sibling (i.e., Have you or anyone else in the household looked at a picture book with your child in the last 3 days? Yes or no?). Once again, we divided children into "more experience" and "less experience" groups across the entire sample along the looking at picture books item, leading to a total of 35 more experience and 61 less experience children in the object condition and 38 more experience and 58 less experience children in the picture condition. The second preregistered analysis involving the looking at picture books variable yielded the same results as the first preregistered analysis involving an aggregated picture experience variable. Once again, there was a significant main effect of condition, such that children performed more accurately in the object condition than the picture condition ($\beta = 0.94$, $SE = .30$, $p = .002$). However, there was no effect of looking at picture books ($\beta = 0.10$, $SE = .37$, $p = .78$) or interaction between condition and looking at picture books ($\beta = 0.29$, $SE = .49$, $p = .56$). There was also an effect of child gender ($\beta = 0.63$, $SE = .24$, $p = .01$), such that girls' responses were significantly more accurate than boys' responses, and an effect of maternal education ($\beta = -0.07$, $SE = .03$, $p = .006$), such that higher maternal education levels were related to lower child accuracy on the experimental task. All other variables were not significant (see Appendix Table B2 for full results).

Furthermore, we conducted similar exploratory analyses, but with picture experience as a continuous variable. First, we conducted a linear regression with accuracy as the dependent variable and condition (i.e., object, picture), picture experience (i.e., as a continuous variable out of eight total questionnaire items), and the interaction between condition and picture experience as between-subjects independent variables. Again, we controlled for child gender, maternal education, child-directed speech, and variety of learning materials in the home. Results from the exploratory analysis matched results from our preregistered analyses. There was a significant main effect of condition, such that children performed more accurately in the object condition than the picture condition ($\beta = 1.03$, $SE = .47$, $p = .03$). However, there was no effect of picture experience ($\beta = 0.59$, $SE = .69$, $p = .39$) or interaction between condition and picture experience ($\beta = 0.44$, $SE = .85$, $p = .60$). There was also an effect of child gender ($\beta = 0.61$, $SE = .24$, $p = .01$), such that girls' responses were significantly more accurate than boys' responses, and a small but significant effect of maternal education ($\beta = -0.08$, $SE = .03$, $p = .003$), such that higher maternal education levels were related to lower child accuracy on the experimental task. All other variables were not significant. Similarly, we ran another linear regression with accuracy as the dependent variable and condition (i.e., object, picture), picture experience (i.e., as a continuous variable out of eight total questionnaire items), and the interaction between condition and picture experience as between-subjects independent variables, but without any additional control variables. There was a significant main effect of condition, such that children performed more accurately in the object condition than the picture condition ($\beta = 1.03$, $SE = .49$, $p = .03$). However, there was no effect of picture experience ($\beta = -0.06$, $SE = .64$, $p = .92$) or interaction between condition and picture experience ($\beta = 0.40$, $SE = .88$, $p = .65$; see Appendix Tables B3 and B4 for full results and Appendix Table B5 for the correlation matrix of all predictor variables).

Moreover, we conducted a simpler exploratory analysis comparing the task accuracy of children with more or less picture experience on the picture vocabulary task, without accounting for control variables. Using a median split across the entire sample along the aggregated eight-item picture experience variable, we find no difference in performance in the picture condition between children with more picture experience ($M = 40.57\%$, $SE = 5.57\%$) and children with less picture experience ($M = 42.64\%$, $SE = 6.50\%$), $t(94) = .70$, $p = .49$. Similarly, we once again find no difference in performance in the picture condition between children who had looked at picture books recently ($M = 41.67\%$, $SE = 6.76\%$) and children who had not ($M = 41.38\%$, $SE = 5.43\%$), $t(94) = .09$, $p = .92$.

The present results show that Mombasa County preschoolers in their first month of formal schooling performed more accurately on a vocabulary assessment involving objects than on the same vocabulary assessment involving pictures. Moreover, these preschoolers also side-biased less frequently on the vocabulary assessment involving objects than on the same vocabulary assessment involving pictures. Both overall effects (i.e., greater accuracy on object than picture assessments and less side-biasing on object than picture assessments) were driven only by their performance on trials involving noun and number words—specifically, words that Mombasa County preschoolers already have some familiarity with—and not by their performance on trials involving color words—specifically, words that Mombasa County preschoolers have very little, if any, familiarity with. However, our individual differences analyses found no evidence of a relation between picture experience and performance on the picture-based vocabulary assessment. Overall, these results suggest that picture assessments may underestimate children's performance, at least in some early childhood development contexts—in this case, with young Mombasa County children with minimal formal schooling.

Experiment 2

Experiment 1 found that low-to-middle-income Mombasa County children in their first month of formal schooling perform more accurately on an object vocabulary task than on a picture vocabulary task. This result tentatively suggests that assessments involving picture stimuli may not be appropriate for children growing up in diverse early environments with less pictures.

However, while variation in picture experience may be related to variation in performance on a picture task, an alternative explanation is that children around the world might show a universal "object advantage." In other words, even high-income U.S. children with an early-emerging understanding of pictures (Bloom & Markson, 1998; DeLoache et al., 1998; Ganea et al., 2009, 2011; Gelman & Ebeling, 1998; Preissler & Bloom, 2007, 2008; Preissler & Carey, 2004) might perform more accurately on tasks involving objects than on the same tasks involving pictures. For example, children may universally find objects to be more interesting than pictures, and a heightened motivation to complete tasks involving objects may account for the accuracy differences between performance on picture and object assessments. Indeed, some previous research shows that Canadian preschoolers, and even U.S. adults, might perform differently on executive function tasks involving objects than pictures (Beaucage et al., 2020; Gomez et al., 2018).

Consequently, Experiment 2 investigated whether young children in a middle-to-high-income urban setting (i.e., San Francisco Bay Area) might also demonstrate an "object advantage"—that is, if high-income, urban, American children will also perform more accurately on the object vocabulary task than on the picture vocabulary task. If children in both Mombasa County and the San Francisco Bay Area perform more accurately on the object vocabulary task than the picture vocabulary task, then assessments involving pictures might underestimate children's early capacities across early environmental contexts, independent of children's degree of picture experience. However, if children in the San Francisco Bay Area perform equally accurately on the object and picture vocabulary tasks, then assessments involving pictures might only underestimate the early capacities of children growing up in low-to-middle-income global contexts, who possess relatively less experience with pictures.

Method

Participants

We preregistered a stopping rule of 96 participants. Consequently, we worked with 96 preschoolers ($M = 2.88$ years, $SD = .61$ years, range = 2.01–3.95 years, 52 females and 44 males). Children were recruited from either a participant database or preschool, both of which reflected local convenience samples. To avoid a ceiling effect in the U.S. sample, we focused on the youngest part of the age range (i.e., the Mombasa sample involved 2- to 7-year-olds; the U.S. sample involved 2- and 3-year-olds).

U.S. participants were primarily White (39 participants), mixed White-Asian (20 participants), and Asian (17 participants), with the remainder of the participants identifying as mixed White-Black (five participants), mixed White-Latina (three participants), Black (three participants), Latina (two participants), Middle Eastern (one participant), mixed White-Middle Eastern (one participant), mixed Latina-Middle Eastern (one participant), and mixed Asian-Black-Latina-White (one participant). Two participants chose not to provide an ethnicity, and another participant identified their ethnicity as "Other." Researchers tested six additional children, whose data were excluded due to refusal to complete the experiment (four), experimenter error (one), or parental interference (one).

Children participated in a quiet spare room in their preschool or in a quiet lab space on a university campus. All participants spoke and heard predominantly English (i.e., more than 50% of the time) and were tested in English. The experiment was approved by a U.S. university's Committee for the Protection of Human Subjects. All parents of child participants provided informed consent. Children received small gifts (e.g., stickers) for participating. All data were collected between September 2022 and July 2023.

Similar to Experiment 1, Experiment 2 used a preregistered between-subjects design, in which half of the sample participated in the object condition (i.e., a vocabulary task involving objects) and the other half of the sample participated in the picture condition (i.e., a vocabulary task involving pictures). We assumed that the overwhelming majority of high-income children in the San Francisco Bay Area would possess near-ceiling levels of picture experience. For example, in a brief survey of 32 children from a local Bay Area museum, parents reported that they had an average of 109.85 picture books in the home (range = 20–1,000 books, SD = 181.55 books) and that they looked at picture books with their child an average of 6.69 days per week (range = 2.00–7.00 days, SD = 1.12 days). We assume that the current sample is similar in demographics (e.g., predominantly White and Asian, high income) to the children surveyed at the museum. Consequently, since we did not plan to conduct individual differences analyses with this sample of high-income children in the San Francisco Bay Area, we did not collect additional measures of early environment (e.g., maternal education and picture experience). Moreover, since we preregistered only analyses comparing experimental conditions at the group level, Experiment 2 used a smaller sample size than Experiment 1.

Stimuli and Procedure

The stimuli and experimental procedure of Experiment 2 were identical to the experimental procedure of Experiment 1, except that the vocabulary task was conducted in English rather than in Swahili.

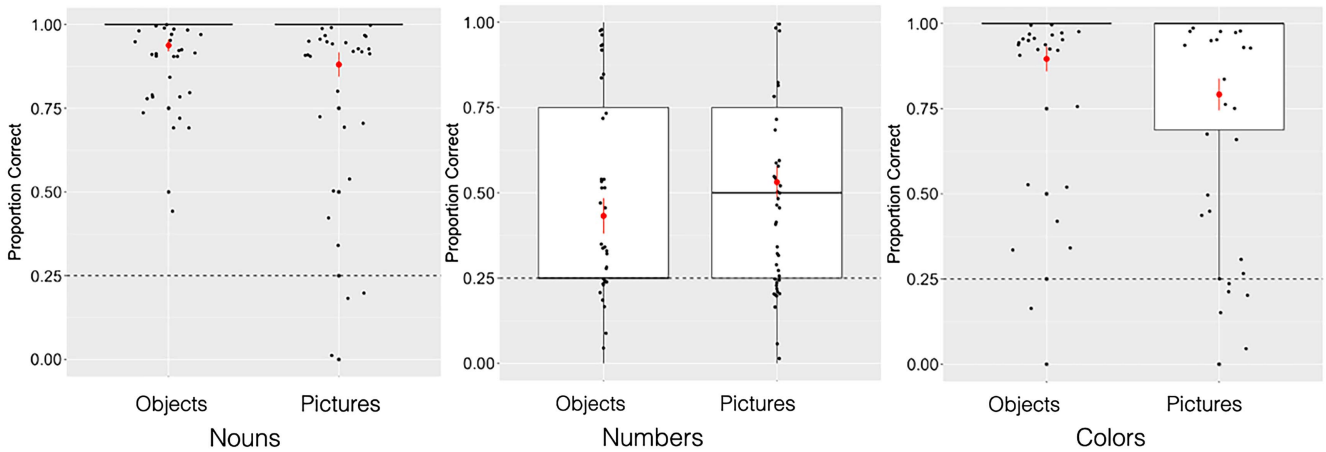
Results and Discussion

Accuracy

In a preregistered analysis, we found that there was no accuracy difference between children in the object condition and children in the picture condition ($\beta = 0.02$, $SE = .04$, $p = .60$). In additional exploratory analyses, we investigated whether children consistently had higher proportions of correct responses in the object condition compared to the picture condition, across multiple kinds of words. We found that there was no accuracy difference between children in the object condition and children in the picture condition, on trials involving nouns ($\beta = 0.06$, $SE = .04$, $p = .16$), trials involving number words ($\beta = -0.10$, $SE = .07$, $p = .15$), and trials involving color words ($\beta = 0.10$, $SE = .06$, $p = .08$). Consequently, middle-to-high-income San Francisco Bay Area children overall performed equally accurately in the object and picture conditions. Moreover, there was no overall trend across blocks, such that middle-to-high-income San Francisco Bay Area children did not consistently perform more accurately on the object condition over the picture condition. Rather, middle-to-high-income San Francisco Bay Area children trended toward performing more accurately on the object condition over the picture condition on the noun and color word trials, and other times trended toward performing more accurately on the picture condition over the object condition on the number word trials (see Figure 5).

Figure 5

Data From Experiment 2: Proportion of Trials in Which Middle-to-High-Income San Francisco Bay Area Preschoolers Chose the Correct Item, for Each Word Type



Note. Boxplot lower and upper hinges correspond to the 25th and 75th percentiles. See the online article for the color version of this figure.

In additional exploratory analyses, we also tested children's performance against chance levels. Since there were four possible answers on each trial, chance performance was 25%. Overall, children performed above chance levels in both the object condition, $M = 75.52\%$, $SE = 2.46\%$, $t(47) = 20.53$, $p < .001$, and the picture condition, $M = 73.26\%$, $SE = 2.52\%$, $t(47) = 13.55$, $p < .001$. Furthermore, we examined children's performance against chance levels for each kind of word. On the noun trials, children performed above chance levels in both the object condition, $M = 93.75\%$, $SE = 1.75\%$, $t(47) = 39.38$, $p < .001$, and the picture condition, $M = 88.02\%$, $SE = 3.65\%$, $t(47) = 17.29$, $p < .001$. On the number trials, children also performed above chance levels in both the object condition, $M = 43.23\%$, $SE = 5.20\%$, $t(47) = 3.51$, $p = .001$, and the picture condition, $M = 53.13\%$, $SE = 4.50\%$, $t(47) = 6.24$, $p < .001$. On the color trials, children also performed at chance levels in both the object condition, $M = 89.58\%$, $SE = 3.63\%$, $t(47) = 17.77$, $p < .001$, and the picture condition, $M = 79.17\%$, $SE = 4.67\%$, $t(47) = 11.60$, $p < .001$. All significant results remained significant at $p < .05$ after correcting for multiple comparisons (Benjamini & Hochberg, 1995). Thus, middle-to-high-income San Francisco Bay Area preschoolers performed significantly above chance on trials involving nouns, number words, and color words, in both the object and picture conditions.

Side-Biasing

In additional exploratory analyses, we investigated the frequency of children's side-biasing behavior (i.e., choosing the same quadrant on all four trials within a block) across the two conditions.

We found that children were less likely to demonstrate side-biasing behavior in the object condition than the picture condition ($\beta = 0.07$, $SE = .03$, $p = .045$), though this result was not significant after correcting for all four side-biasing comparisons (Benjamini & Hochberg, 1995). Children in the object condition side-biased on 4% of the blocks, whereas children in the picture condition side-biased on 11% of the blocks. There was no difference in the frequency of children's side-biasing behaviors in the object and picture conditions on the noun blocks, $\chi^2(1, N = 96) = 0$, $p = 1.00$, and number blocks, $\chi^2(1, N = 96) = 0.32$, $p = .57$. There was a difference in the frequency of children's side-biasing behaviors in the object and picture conditions on the color blocks, $\chi^2(1, N = 96) = 4.44$, $p = .04$, but this difference was not significant after correcting for all four side-biasing comparisons (Benjamini & Hochberg, 1995). On the noun block, preschoolers in the object condition side-biased on 0% of the blocks, whereas preschoolers in the picture condition side-biased on 2% of the blocks. On the number block, preschoolers in the object condition side-biased on 13% of the blocks, whereas preschoolers in the picture condition side-biased on 19% of the blocks. In contrast, on the color block, preschoolers in the object condition side-biased on 0% of the blocks, whereas preschoolers in the picture condition side-biased on 13% of the blocks.

General Discussion

The present research investigates whether early childhood cognitive assessments involving picture stimuli are valid in global contexts. In Experiment 1, we presented low-to-middle-income children in their first month of formal schooling in Mombasa County, Kenya, with a vocabulary task based on the classic Peabody Picture Vocabulary Task, an assessment tool that is widely used by researchers in psychology, public health, and development economics in global settings (e.g., Akseer et al., 2022; Heo et al., 2020). We find that while children performed at above chance levels on both the object vocabulary task and the picture vocabulary task, children also performed significantly more accurately on the object vocabulary task than on the picture vocabulary task. In Experiment 2, we presented middle-to-high-income children in the San Francisco Bay Area with the same vocabulary task and found no difference in performance between the object vocabulary task and the picture vocabulary task in this middle-to-high-income U.S. sample.

Furthermore, additional analyses showed that, in Experiment 1, the accuracy difference between picture and object assessments was driven by performance on trials involving the kinds of words that low-to-middle-income Mombasa County children are typically familiar with (i.e., noun and number words, but not color words). Specifically, on the noun and number word trials, low-to-middle-income Mombasa County children performed more accurately on the object assessment than the picture assessment. In contrast, on the color word trials, low-to-middle-income Mombasa County children performed at chance levels on both the object and picture assessments. Low-to-middle-income Mombasa County children's side-biasing behavior also showed a similar pattern, such that these children tended to side-bias less when presented with object assessments than picture assessments, and this side-biasing difference between picture and object assessments was driven by trials involving the kinds of words that young children in this context are typically familiar with. Specifically, on the noun and number word blocks, low-to-middle-income Mombasa County children side-biased less frequently on the object assessment than on the picture assessment. In contrast, on the color word block, low-to-middle-income Mombasa County children side-biased equally frequently on the object and picture assessments. In Experiment 2, middle-to-high-income San Francisco Bay Area children performed equally accurately on object and picture assessments across the three kinds of word trials (i.e., nouns, number words, and color words). They trended toward performing more accurately on the object assessment than the picture assessment on the noun and color word trials, but trended toward performing more accurately on the picture assessment than the object assessment on the number word trials. Moreover, while middle-to-high-income San Francisco Bay Area children, like their Mombasa County counterparts, seemed to side-bias less frequently on the object assessment than the picture assessment, these results were null after correcting for multiple comparisons. In general, middle-to-high-income San Francisco Bay Area children also demonstrated less side-biasing behavior than low-to-middle-income Mombasa County children (i.e., side-biasing on 8% of all blocks and 25% of all blocks, respectively).

Taken together, these results tentatively suggest that assessments involving pictures may *underestimate* children's cognitive capacities in some early environmental contexts, potentially in contexts with fewer pictures and visual representations (e.g., televisions, smartphones, billboards). Accurate assessment of child development is crucial to understanding the determinants of developmental outcomes in diverse populations and to evaluating and refining policies and programs designed to improve these outcomes. Thus, future research should replicate and extend these findings across various early childhood environments, in order to examine the extent

to which pictures underestimate children's cognitive capacities across cultures and contexts.

The present experiments also possess limitations. The primary goal of this set of experiments was to compare children's performance between assessments involving object and picture stimuli. Consequently, we caution that the current data sets are not necessarily a valid measure of children's vocabulary sizes. For example, in Experiment 1, we likely underestimate low-to-middle-income Mombasa County children's knowledge of color words. In Experiment 1, we made the methodological decision to conduct the entire assessment in children's first language, namely Swahili. However, anecdotal conversations with local researchers suggest that low-to-middle-income children living in Mombasa County typically learn color words only when they enter the formal school system and that schools initially teach children color words in English rather than Swahili. Thus, while there is a substantial difference in overall accuracy scores between children in Mombasa County and the San Francisco Bay Area, such that children in Mombasa County generally performed less accurately than children in the San Francisco Bay Area (though see the number trials in the object condition, in which children in Mombasa County trend toward more accurate performance than children in the San Francisco Bay Area), these overall accuracy scores may be related to variables other than vocabulary size (e.g., the language the assessment was conducted in). Consequently, we note that these data are not useful for comparisons of *overall accuracy* across contexts (Frank et al., 2024). Instead, these data sets are useful for comparisons of *differences between picture and object assessment accuracy* across contexts.

We caution against making strong claims about the validity of picture-based assessments in global contexts, since another limitation is that the secondary individual differences analyses in Experiment 1 did not find a significant association between looking at pictures and performance on the picture vocabulary task. Consequently, the current research cannot demonstrate an explicit significant association between children's picture experience and their performance on picture-based assessments. An alternative explanation for the present pattern of results is that low-to-middle-income children in Mombasa County show a performance difference between object- and picture-based assessments due to stimulus specificity effects, for example, because they learn different kinds of words from objects and pictures. It is possible that low-to-middle-income Mombasa County children typically learn the words involved in the present vocabulary assessment (i.e., artifact names, numbers) from objects rather than pictures, but typically learn other kinds of words (e.g., safari animal names) from pictures rather than objects. More research investigating how children growing up in this context learn from both objects and pictures is necessary to rule out this alternative possibility. Thus, while we tentatively speculate that an "object advantage"—that is, increased accuracy on an object vocabulary task relative to a picture vocabulary task—exists for low-to-middle-income Mombasa County preschoolers but not for middle-to-high-income San Francisco Bay Area preschoolers because the latter group has more picture experience than the former group, further research is required to demonstrate that picture experience (as opposed to another kind of contextual difference) is indeed the critical variable involved. Nonetheless, while the present research does not provide a definitive mechanistic explanation for the observed differences, the results still suggest that assessment tools used in global contexts should be chosen with caution.

The child vocabulary and caregiver questionnaire should also be improved upon in future research. For example, the present child vocabulary assessment only allows for binary scoring—that is, whether children selected the correct items or not. However, many other vocabulary assessments (e.g., the Peabody Picture Vocabulary Test) present target items (e.g., a toothbrush), near conceptual distractors (e.g., toothpaste), and random distractors (e.g., a skirt), which allow for more detailed analyses of children's error patterns. For example, if children err by selecting near conceptual distractors, then they may still have some degree of understanding of the task at hand. However, if children err by selecting random distractors, then they may have a substantial amount of difficulty with either the target word, the stimuli format, or both. Thus, future work should use assessments that allow for these kinds of detailed error analyses, thus leading to a more nuanced understanding of children's cognitive capacities.

Moreover, future work should consider using more extensive caregiver questionnaires, in order to capture more kinds of picture experience that the present research may have missed. For example, an extensive picture experience questionnaire may not only measure children's exposure to pictures (e.g., being in the presence of picture books), but also more comprehensively assess their engagement with pictures (e.g., opportunities to map pictures to real-world objects or have explicit conversations about pictures as symbols with the caregiver). Furthermore, the present caregiver questionnaire may have a Western middle-class bias, as it focuses primarily on picture books, a kind of visual media that middle-to-high-income Western children have extensive experience with. Children growing up in diverse global contexts may possess less experience with picture books, but possibly more experience with other kinds of visual media (e.g., murals, videos, scale models, billboards, magazines, newspapers). Additionally, the present research did not measure picture experience in the U.S. sample, limiting cross-cultural comparisons and potentially inadvertently positioning Western culture as the norm, without the need for further exploration or justification (Draper et al., 2023). Thus, future work investigating the relation between picture experience and picture comprehension should measure children's exposure to, and engagement with, many diverse kinds of visual media, in all samples involved in the research.

Another especially important direction for future research is to pinpoint a *mechanistic explanation* for why some children may perform more accurately on the object task than the picture task. What *kind* of early experience with pictures might help children perform more accurately on tasks involving picture stimuli? Some researchers have posited that the development of picture comprehension is a fundamentally social endeavor (Callaghan, 2020), perhaps requiring direct interactions with parents and teachers and explicit conversations about picture as symbolic representations. In contrast, it is also possible that merely passive experience with visual representations—for example, simply seeing posters on the wall or watching television at home—might be necessary and sufficient for the acquisition of picture comprehension and more accurate performance on tasks involving pictures. Consequently, more empirical research investigating what quality and quantity of picture experience is sufficient for picture comprehension, as well as consequent performance on picture-based assessments, is necessary.

Thus, another related direction for future research is to investigate whether the current findings *generalize across populations*. While the low-to-middle-income preschoolers in our Mombasa County

sample showed significant amounts of variation in picture experience, Mombasa is the second largest city in Kenya and still an urban environment. Previous research involving western Kenyan toddlers in urban Kisumu and rural Rachuonyo suggests that there may be variation in children's early environments, and possibly their consequent development, across urban and rural settings within a single country (Zhu, Pitchik, et al., 2025). Thus, future research should explore the generalizability of this current paradigm in more rural contexts. For example, do children in more rural areas of Coastal Kenya (e.g., Lamu or Kilifi), who may possess even less picture experience than their Mombasa County counterparts, also perform more accurately on an object task than a picture task? Might children in these more rural areas of Coastal Kenya show an even larger difference in accuracy between object tasks and picture tasks? Investigating the generalizability of the current findings across cultures and contexts might provide stronger support that picture-based assessments indeed sometimes underestimate children's capacities, as well as additional insight regarding the mechanisms driving overall difference between performance on object tasks and picture tasks.

Finally, a final direction for future research is to investigate whether the current findings *generalize across tasks*. Researchers use picture stimuli not only in vocabulary tasks but also in a wide variety of other cognitive and socioemotional assessments, such as experimental tasks investigating relational reasoning (Christie & Gentner, 2010; Hochmann et al., 2017) and emotion recognition (Dillon et al., 2017; Gao & Maurer, 2009). Empirical data showing that children perform more accurately in object tasks than picture tasks on a wide variety of cognitive and social assessments would provide converging evidence that the kind of stimuli (i.e., pictures vs. objects) rather than the kind of content (e.g., linguistic, cognitive, or socioemotional reasoning) is the critical factor underlying differences between children's performance on picture and object tasks. Indeed, some research provides evidence that the kind of stimuli matters: Toddlers in rural coastal Tanzania are better able to learn a new word for an object than for a picture (Walker et al., 2013), and preschoolers in rural Peru and rural India perform similarly to their Canadian counterparts on a false belief task involving objects but less accurately on the same false belief task involving pictures (Callaghan et al., 2012). Overall, more research with diverse global child samples is needed in order to investigate the extent to which assessments involving picture stimuli may be valid or invalid across varying early childhood contexts.

The validity of early childhood assessments across cultures and contexts has serious implications for both basic cognitive science and applied research in education, public health, and development economics. For example, if current experimental paradigms involving picture stimuli are inappropriate in some global contexts, it is possible that researchers may consequently underestimate the developmental timepoints at which diverse groups of children acquire certain cognitive and socioemotional capacities. Moreover, understanding the mechanisms that facilitate or impede children's performance on assessments involving pictures can help researchers design more appropriate and effective interventions for children growing up in low-resource settings. A study conducted in rural Western Kenya found that children who received a dialogic reading parenting intervention (i.e., in which parents encourage children to actively formulate and articulate their ideas over discussions of picture books) had significant improvements in their expressive vocabulary scores, such that children were able to accurately name more pictures (Knauer et al., 2020). More research is required to understand what mechanism is responsible for this effect: Are children who participate in the dialogic reading intervention performing more accurately on the expressive vocabulary assessment because of an increase in their vocabulary size or a change in their picture comprehension abilities? Thus, accurate assessment tools are critical to early childhood development research across a wide variety of disciplines.

In summary, the present research shows that low-to-middle-income Mombasa County preschoolers in their first year of formal schooling perform more accurately on an object vocabulary task than on a picture vocabulary task, tentatively suggesting that assessments involving picture stimuli may underestimate children's cognitive capacities in some early childhood contexts. In contrast, children from a North American convenience sample did not show any performance differences between the object and picture vocabulary tasks. Though future research should investigate the generalizability of the current findings, our research makes an initial contribution to a better understanding of possible changes required to adapt assessment tools to global contexts (Draper et al., 2023; Jukes et al., 2024). Accurate assessment of child development is crucial not only to conducting valid research in basic cognitive science but also to evaluating early childhood policies and health interventions designed to improve the lives of children around the world.

References

Akseer, N., Tasic, H., Nnachebe Onah, M., Wigle, J., Rajakumar, R., Sanchez-Hernandez, D., Akuoku, J., Black, R. E., Horta, B. L., Nwuneli, N., Shine, R., Wazny, K., Japra, N., Shekar, M., & Hoddinott, J. (2022). Economic costs of childhood stunting to the private sector in low- and middle-income countries. *eClinicalMedicine*, *45*, Article 101320. https://doi.org/10.1016/j.eclinm.2022.101320

Antonoplis, S. (2023). Studying socioeconomic status: Conceptual problems and an alternative path forward. *Perspectives on Psychological Science*, *18*(2), 275–292. https://doi.org/10.1177/17456916221093615

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board Task Force Report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Barrera, M. E., & Maurer, D. (1981). The perception of facial expressions by the three-month-old. *Child Development*, *52*(1), 203–206. https://doi.org/10.2307/1129231

Beaucage, N., Skolney, J., Hewes, J., & Vongpaisal, T. (2020). Multisensory stimuli enhance 3-year-old children's executive function: A three-dimensional object version of the standard dimensional change card sort. *Journal of Experimental Child Psychology*, *189*, Article 104694. https://doi.org/10.1016/j.jecp.2019.104694

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

Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science*, *9*(3), 200–204. https://doi.org/10.1111/1467-9280.00038

Bradley, R. H., & Caldwell, B. M. (1984). The HOME inventory and family demographics. *Developmental Psychology*, *20*(2), 315–320. https://doi.org/10.1037/0012-1649.20.2.315

Callaghan, T. (2020). The origins and development of a symbolic mind: The case of pictoral symbols. *Interchange*, *51*(1), 53–64. https://doi.org/10.1007/s10780-020-09396-z

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., & Tomasello, M. (2011). Early social cognition in three cultural contexts. *Monographs of the Society for Research in Child Development*, *76*(2), vii–142. https://doi.org/10.1111/j.1540-5834.2011.00603.x

Callaghan, T., Rochat, P., & Corbit, J. (2012). Young children's knowledge of the representational function of pictorial symbols: Development across the preschool years in three cultures. *Journal of Cognition and Development*, *13*(3), 320–353. https://doi.org/10.1080/15248372.2011.587853

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373. https://doi.org/10.1080/15248371003700015

Cromwell, E. A., Dube, Q., Cole, S. R., Chirambo, C., Dow, A. E., Heyderman, R. S., & Van Rie, A. (2014). Validity of US norms for the Bayley Scales of Infant Development-III in Malawian children. *European Journal of Paediatric Neurology*, *18*(2), 223–230. https://doi.org/10.1016/j.ejpn.2013.11.011

DeLoache, J. S., Pierroutsakos, S. L., Uttal, D. H., Rosengren, K. S., & Gottlieb, A. (1998). Grasping the nature of pictures. *Psychological Science*, *9*(3), 205–210. https://doi.org/10.1111/1467-9280.00039

Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., & Duflo, E. (2017). Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics. *Science*, *357*(6346), 47–55. https://doi.org/10.1126/science.aal4724

Draper, C. E., Barnett, L. M., Cook, C. J., Cuartas, J. A., Howard, S. J., McCoy, D. C., Merkley, R., Molano, A., Maldonado-Carreño, G., Obradović, J., Scerif, G., Valentini, N. C., Venetsanou, F., & Yousafzai, A. K. (2023). Publishing child development research from around the world: An unfair playing field resulting in most of the world's child population under-represented in research. *Infant and Child Development*, *32*(6), Article e2375. https://doi.org/10.1002/icd.2375

Dunn, D. M. (2019). *Peabody Picture Vocabulary Test* (5th ed.). National Computer Systems Pearson.

Fernald, L. C. H., Prado, E., Kariger, P., & Raikes, A. (2017). *A toolkit for measuring early childhood development*. The World Bank.

Frank, M. C., Baumgartner, H., Braginsky, M., Kachergis, G., Lightbody, A., Sparks, R. Z., Zhu, R., Carlson, S. M., Graham, S., Lipina, S. J., Newcombe, N. S., Odgers, C. L., Pianta, R. C., Siegler, R. S., Snowling, M., Yoshikawa, Dodge, K. A., & Cubillo, A. (2024). *Learning Variability Network Exchange (LEVANTE): A global framework for measuring children's learning variability through collaborative data sharing*. PsyArXiv. https://doi.org/10.31234/osf.io/namx2

Ganea, P. A., Allen, M. L., Butler, L., Carey, S., & DeLoache, J. S. (2009). Toddlers' referential understanding of pictures. *Journal of Experimental Child Psychology*, *104*(3), 283–295. https://doi.org/10.1016/j.jecp.2009.05.008

Ganea, P. A., Ma, L., & Deloache, J. S. (2011). Young children's learning and transfer of biological information from picture books to real animals. *Child Development*, *82*(5), 1421–1433. https://doi.org/10.1111/j.1467-8624.2011.01612.x

Gao, X., & Maurer, D. (2009). Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions. *Journal of Experimental Child Psychology*, *102*(4), 503–521. https://doi.org/10.1016/j.jecp.2008.11.002

Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition*, *66*(2), B35–B47. https://doi.org/10.1016/S0010-0277(98)00022-5

Gomez, M. A., Skiba, R. M., & Snow, J. C. (2018). Graspable objects grab attention more than images do. *Psychological Science*, *29*(2), 206–218. https://doi.org/10.1177/0956797617730599

Heo, J., Krishna, A., Perkins, J. M., Lee, H. Y., Lee, J. K., Subramanian, S. V., & Oh, J. (2020). Community determinates of physical growth and cognitive development among Indian children in early childhood: A multivariate multilevel analysis. *International Journal of Environmental Research and Public Health*, *17*(1), Article 182. https://doi.org/10.3390/ijerph17010182

Hochberg, J., & Brooks, V. (1962). Pictorial recognition as an unlearned ability: A study of one child's performance. *The American Journal of Psychology*, *75*(4), 624–628. https://doi.org/10.2307/1420286

Hochmann, J. R., Tuerk, A. S., Sanborn, S., Zhu, R., Long, R., Dempster, M., & Carey, S. (2017). Children's representation of abstract relations in relational/array match-to-sample tasks. *Cognitive Psychology*, *99*, 17–43. https://doi.org/10.1016/j.cogpsych.2017.11.001

Jukes, M. C. H., Ahmed, I., Baker, S., Draper, C. E., Howard, S. J., McCoy, D. C., Obradović, J., & Wolf, S. (2024). Principles for adapting assessments of executive function across cultural contexts. *Brain Sciences*, *14*(4), Article 318. https://doi.org/10.3390/brainsci14040318

Knauer, H. A., Jakiela, P., Ozier, O., Aboud, F. E., & Fernald, L. C. H. (2020). Enhancing young children's language acquisition through parent–child book-sharing: A randomized trial in rural Kenya. *Early Childhood Research Quarterly*, *50*, 179–190. https://doi.org/10.1016/j.ecresq.2019.01.002

Luoto, J. E., Lopez Garcia, I., Aboud, F. E., Singla, D. R., Zhu, R., Otieno, R., & Alu, E. (2021). An implementation evaluation of group-based parenting to promote early childhood development in rural Kenya. *Frontiers in Public Health*, *9*, Article 653106. https://doi.org/10.3389/fpubh.2021.653106

Navarro, D. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners*. University of Adelaide. https://learningstatisticswithr.com/lsr-0.6.pdf

Preissler, M. A., & Bloom, P. (2007). Two-year-olds appreciate the dual nature of pictures. *Psychological Science*, *18*(1), 1–2. https://doi.org/10.1111/j.1467-9280.2007.01837.x

Preissler, M. A., & Bloom, P. (2008). Two-year-olds use artist intention to understand drawings. *Cognition*, *106*(1), 512–518. https://doi.org/10.1016/j.cognition.2007.02.002

Preissler, M. A., & Carey, S. (2004). Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development*, *5*(2), 185–212. https://doi.org/10.1207/s15327647jcd0502_2

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Ranjitkar, S., Kvestad, I., Strand, T. A., Ulak, M., Shrestha, M., Chandyo, R. K., Shrestha, L., & Hysing, M. (2018). Acceptability and Reliability of the Bayley Scales of Infant and Toddler Development-III among children in Bhaktapur, Nepal. *Frontiers in Psychology*, *9*, Article 1265. https://doi.org/10.3389/fpsyg.2018.01265

Saitoti, T. O. (1986). *The worlds of a Maasai warrior: An autobiography*. University of California Press.

Schools2030. (2023). *Schools2030 Annual Report 2023: Catalyzing teacher-driven holistic learning innovations to achieve SDG4 by 2030*. Aga Khan Foundation. https://schools2030.org/download/schools2030-annual-report-2023/

Walker, C. M., Walker, L. B., & Ganea, P. A. (2013). The role of symbol-based experience in early learning and transfer from pictures: Evidence from Tanzania. *Developmental Psychology*, *49*(7), 1315–1324. https://doi.org/10.1037/a0029483

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation* (R package Version 1.1.4) [Computer software], https://dplyr.tidyverse.org

Yoon, J. M. D., Witthoft, N., Winawer, J., Frank, M. C., Everett, D. L., & Gibson, E. (2014). Cultural differences in perceptual reorganization in US and Pirahã adults. *PLOS ONE*, *9*(11), Article e110225. https://doi.org/10.1371/journal.pone.0110225

Zhu, R., Kilonzo, T. N., Zhu, L. Z., Fan, J. E., & Frank, M. C. (2025). *Cross-contextual variability in children's early understanding of visual media*. PsyArXiv. https://doi.org/10.31234/osf.io/79xh5_v1

Zhu, R., Pitchik, H. O., Kilonzo, T. N., Engelmann, J., Fernald, L. C., & Gopnik, A. (2025). The development of picture comprehension across early environments: Evidence from urban and rural toddlers in Western Kenya. *Developmental Science*, *28*(1), Article e13579. https://doi.org/10.1111/desc.13579

Zuilkowski, S. S., McCoy, D. C., Serpell, R., Matafwali, B., & Fink, G. (2016). Dimensionality and the development of cognitive assessments for children in Sub-Saharan Africa. *Journal of Cross-Cultural Psychology*, *47*(3), 341–354. https://doi.org/10.1177/0022022115624155

Appendix A

Caregiver Questionnaire

Home Environment—Picture Experience

1. Have you (or any other adult in the household) looked at a picture book with your child in the last 3 days?
   a. If yes, who did this?
   b. If yes, how many times was this done?
2. Has the child looked at a picture book on their own in the last 3 days?
   a. If yes, how many times was this done?
3. Have you (or any other adult in the household) looked at any materials with your child that have pictures in them in the last 3 days?
   a. If yes, who did this?
   b. If yes, how many times was this done?
4. Has the child looked at any other materials with pictures in them on their own in the last 3 days?
   a. If yes, how many times was this done?
5. Does your house have pictures/posters/paintings/calendars on the wall?
   a. If yes, what types of things do you have on the walls?
   b. If yes, how many do you have?
6. Has the child looked at any screens (e.g., TV, tablet, phone) in the last 3 days?
   a. If yes, name the item with the screen.
   b. How many times was this done?
7. Do you have any books for adults to read in your home?
   a. If yes, how many?
8. Does the child have children's books (i.e., books specifically for the child, not textbooks)?
   a. If yes, how many?

Home Environment—Child-Directed Speech

9. Do you teach the child simple verbal manners, such as saying please, thank you, and I'm sorry?
10. Do you ever encourage the child to talk while you listen?
11. Does anyone sing or play music for the child to listen or dance to?
12. Did you get time to hold your child and talk with affection yesterday?

Home Environment—Learning Materials

13. Are two or more materials that teach colors, sizes, and shapes available to the child to play with?
14. Are two or more playthings requiring fine motor movements available to the child (e.g., drawing with a pencil or stick, making patterns with buttons or caps)?
15. Are two or more play materials or games that help teach numbers available to the child, in which children and others count or use numbers (e.g., counting sticks, caps)?
16. Is a real or toy musical instrument available to the child (e.g., drum, singing, dancing)?

Home Environment—Assets

17. Does your household own any of the following?
    a. Mattress—foam (store bought)
    b. Cell phone—not smartphone

(*Appendices continue*)

c. Radio
d. Bicycle
e. Sofa
f. Chair
g. Lantern
h. Table
i. Electricity or solar supply
j. Motorcycle
k. Smartphone
l. Television
m. Computer

Home Environment—Household Materials

18. Do you own a homestead?
19. Do you own land for production?
20. How many rooms are there in your house?
21. What is the material of the household walls? (concrete brick, clay brick, or mud)
22. What is the material of the household roof? (corrugated iron, thatch)
23. What is the material of the household floor? (tiled, cemented, unfinished dirt)

Demographics

24. Who is responding? (mother, father, grandmother, other)
25. Who is the primary caregiver of the child?
26. How many years of school has the primary caregiver completed? (starting from Preprimary 1 as Year 1)
27. How many years of school has the mother completed? (starting from Preprimary 1 as Year 1)

Appendix B

Regression and Correlation Tables

Table B1
Regression Outputs—Picture Experience (Median Split)

Variable	Coefficient	*SE*	*p*
Intercept	6.66	0.62	<.001***
Condition	−0.71	0.35	.04*
Picture experience	0.31	0.37	.40
Child gender	−0.64	0.24	.009**
Child-directed speech	−0.02	0.17	.93
Home learning materials	−0.01	0.11	.94
Maternal education level	−0.07	0.03	.005**
Condition × Picture Experience	−0.22	0.48	.65

Note. A linear regression with accuracy as the dependent variable and condition (i.e., object, picture), picture experience (i.e., more experience, less experience), and the interaction between condition and picture experience as between-subjects independent variables, controlling for child gender, maternal education, child-directed speech, and home learning materials. *SE* = standard error.
* $p < .05$. ** $p < .01$. *** $p < .001$.

(*Appendices continue*)

Table B2

Regression Outputs—Looking at Pictures (Median Split)

Variable	Coefficient	*SE*	*p*
Intercept	6.73	0.61	<.001***
Condition	−0.94	0.30	.002**
Picture experience	−0.10	0.37	.78
Child gender	−0.63	0.24	.009**
Child-directed speech	0.01	0.17	.96
Home learning materials	0.01	0.10	.90
Maternal education level	−0.07	0.03	.006**
Condition × Picture Experience	0.29	0.49	.56

Note. A linear regression with accuracy as the dependent variable and condition (i.e., object, picture), looking at picture books (i.e., more experience, less experience), and the interaction between condition and picture experience as between-subjects independent variables, controlling for child gender, maternal education, child-directed speech, and home learning materials. *SE* = standard error.
** $p < .01$. *** $p < .001$.

Table B3

Regression Outputs—Picture Experience (Continuous Variable)

Variable	Coefficient	*SE*	*p*
Intercept	6.80	0.62	<.001***
Condition	−1.03	0.47	.03*
Picture experience	0.59	0.69	.39
Child gender	−0.61	0.24	.01*
Child-directed speech	−0.06	0.18	.72
Home learning materials	−0.06	0.10	.59
Maternal education level	−0.08	0.03	.003**
Condition × Picture Experience	0.44	0.85	.60

Note. A linear regression with accuracy as the dependent variable and condition (i.e., object, picture), picture experience (i.e., as a continuous variable out of eight questionnaire items), and the interaction between condition and picture experience as between-subjects independent variables, controlling for child gender, maternal education, child-directed speech, and home learning materials. *SE* = standard error.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table B4

Regression Outputs—Picture Experience (Continuous Variable), Without Control Variables

Variable	Coefficient	*SE*	*p*
Intercept	5.85	0.35	<.001***
Condition	−1.03	0.49	.03*
Picture experience	−0.06	0.64	.92
Condition × Picture Experience	0.40	0.88	.65

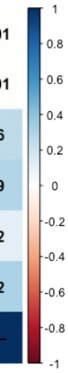
Note. A linear regression with accuracy as the dependent variable and condition (i.e., object, picture), picture experience (i.e., as a continuous variable out of eight questionnaire items), and the interaction between condition and picture experience as between-subjects independent variables, without any control variables. *SE* = standard error.
* $p < .05$. *** $p < .001$.

(*Appendices continue*)

Table B5

Correlations Among All Predictor Variables

Variable	1	2	3	4	5	6	7
1. Child Gender	—	−.03	−.03	−.03	0	−.08	−.01
2. Condition	−.03	—	−.03	−.03	−.06	−.01	−.01
3. Looking at Pictures	−.03	−.03	—	.53	.29	.39	.26
4. Total Picture Experience	−.03	−.03	.53	—	.3	.55	.29
5. Child Directed Speech	0	−.06	.29	.3	—	.21	.12
6. Home Learning Materials	−.08	−.01	.39	.55	.21	—	.32
7. Maternal Education	−.01	−.01	.26	.29	.12	.32	—



Note. See the online article for the color version of this table.

Received July 21, 2024
Revision received February 24, 2025
Accepted June 22, 2025 ■