Investigating children's performance on object- and picture-based vocabulary assessments in global contexts: Evidence from Kisumu, Kenya

Rebecca Zhu (ryz@stanford.edu)[1], Tabitha Nduku Kilonzo (tnduku@rti.org)[2],
Joab Ochieng Arieda (joab@swapkenya.org)[3], Arnav Verma (arnavv@stanford.edu)[1],
Judith E. Fan (jefan@stanford.edu)[1], Michael C. Frank (mcfrank@stanford.edu)[1]
[1]Department of Psychology, Stanford University, [2]International Education, RTI International Nairobi, [3]Safe Water and AIDS Project

Abstract

Assessments of early cognitive and linguistic abilities typically involve picture stimuli. As these assessments spread worldwide, researchers make an implicit assumption: that children across contexts understand pictures in the same way, at the same developmental timepoint. What if this assumption does not hold for some or all kinds of pictures? In the present research, a preregistered sample of 128 3- to 7-year- olds from Kisumu, Kenya participated in a Swahili vocabulary assessment. Using a within-subjects design, each participant completed vocabulary trials in four formats (i.e., objects, photographs, cartoons, black-and-white line drawings). Preregistered analyses showed that children performed equally accurately across object, photograph, and cartoon formats, but less accurately in the line drawing format. However, exploratory analyses suggested that a subset of line drawings drove this difference. These findings suggest that caution is necessary in the use of picture stimuli and that assessments involving line drawings may sometimes underestimate children's capacities.

Keywords: cognitive development, picture comprehension, cross-cultural research, measurement

Humans possess remarkable, and potentially species-unique, capacities to understand and use various kinds of visual media (e.g., pictures, videos, scale models, maps). In particular, previous research shows that high-income children growing up in Western contexts understand pictures early in development. For example, high-income U.S. toddlers know that pictures refer to actual objects in the world (Preissler & Carey, 2004) and appreciate the role of intention when interpreting pictures (Gelman & Ebeling, 1998). Moreover, children in these settings can learn novel concepts from picture books and apply these concepts to actual entities (Ganea, Ma, & DeLoache, 2011). But how do these abilities develop?

While many cognitive scientists argue that some kind of experience with pictures is necessary for the picture comprehension (DeLoache, Pierroutsakos, & Uttal, 2003; Zhu et al., 2025), it is unclear exactly what quality and quantity of picture experience facilitates this development. One possibility is that only minimal picture experience is necessary for the development of full comprehension, and thus that children across diverse environments understand pictures in similar ways, at similar developmental timepoints. Another possibility is that more extensive picture experience (e.g., being surrounded by picture books, television screens, billboards, posters) is necessary for the development of full comprehension, and thus that children living in environments without many picture books or other visual media understand pictures at a later timepoint than their counterparts living in environments with an abundance of pictures. Since children growing up in high-income, urban contexts typically all possess extensive picture experience, cognitive scientists must work with global populations to explore possible consistency or diversity in children's developing capacities to understand and learn from pictures.

Moreover, the most widely used learning materials (e.g. books, posters) and assessments of early cognitive and linguistic abilities use picture stimuli (Fernald, Prado, Kariger, & Raikes, 2017). As these materials and assessments spread across the world, researchers make the implicit assumption that children across contexts understand pictures in the same way, at the same developmental time point. What if this assumption does not hold? Differences in picture comprehension can change the efficacy of learning materials and the validity of assessments (Draper et al., 2022; Jukes et al., 2024). Thus, it is also important to investigate when and how children understand pictures across cultures and contexts to determine how to appropriately translate learning materials and assessments globally.

Indeed, there is already some evidence that picture-based assessments may underestimate children's capacities, in some global contexts (Callaghan et al., 2011; Walker, Walker, & Ganea, 2013; Zhu, Kilonzo, Engelmann, & Gopnik, 2024). For example, preschoolers living in rural environments in India (i.e., a village 70 kilometers 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 when the task was presented using objects rather than black-and-white line drawings (Callaghan et al., 2011). Similarly, young toddlers living in a rural village in the Kibaha-Pwani District of Tanzania's Coast Region do not map a novel word to a novel photograph, but succeed in mapping a novel word onto a novel object (Walker et al., 2013). Moreover, low- to middle-income 2- to 7-year-olds in their first month of formal schooling living in Mombasa County, Kenya (i.e., a mix of urban and rural areas) performed significantly more accurately on an object-based vocabulary assessment than on a cartoon-based vocabulary assessment. In contrast, high-income urban and suburban 2- and 3-year-olds living in the San Francisco Bay Area, U.S. performed equally accurately on object-based and cartoon-based vocabulary assessments (Zhu et al., 2024). Furthermore, in a sample of toddlers living in urban Kisumu, Kenya, variation in toddlers' early picture experience is related to variation in toddlers' ability to learn from pictures (Zhu et al., 2025). Thus, this data suggests that assessments involving pictures may underestimate children's capacities in some global contexts.

However, more work is necessary to investigate the generalizability of these initial findings. For example, it is unclear if these initial findings, which were either conducted with young toddlers and preschoolers (Callaghan et al., 2011; Walker et al., 2013; Zhu et al., 2025) or children in their very first month of formal schooling (Zhu et al., 2024), might generalize to older children with more formal schooling experience, and thus likely also more picture experience. Determining whether assessments involving pictures are appropriate for children who have already completed some formal schooling will provide insight into whether researchers and policymakers need to adapt assessment tools for many children in global contexts, or perhaps only the youngest learners.

Moreover, previous cross-cultural work on children's picture comprehension contrasted object-based tasks with only one other kind of picture-based task (i.e., photographs, black-and-white line drawings) (Callaghan et al., 2011; Walker et al., 2013; Zhu et al., 2024). However, different kinds of picture stimuli may pose various advantages and disadvantages for young children's understanding of pictures (Pierroutsakos & DeLoache, 2003; Simcock & DeLoache, 2006). For example, color photographs or cartoons may be easier for children to understand and learn from, because they are more perceptually similar to objects than black-and-white line drawings (Ganea et al., 2011). However, black-and-white line drawings may be more culturally neutral than color photographs or clipart, as the former typically depict fewer contextually salient features (e.g., skin color). Thus, the present experiment contrast objects with three common picture formats (i.e., photographs, cartoons, black-and-white line drawings), to explore whether possible differences between object and picture stimuli hold for multiple kinds of picture formats.

Methods

Open Science Statement

Preregistration, materials, deidentified data, and analysis scripts are publicly available at `https://osf.io/dgz9w/`.

Participants

A preregistered sample of 128 Kisumu children (M = 5.39 years; range = 3.01-7.99 years; 58 girls, 70 boys; 92% attending daycare/school; M = 2.48 years in daycare/school, range = 0-7 years in daycare/school) participated in a Swahili vocabulary assessment. Researchers tested two additional children whose data were excluded due to fussiness (one child) or experimenter error (one child).

Kisumu is the third largest city in Kenya, and a relatively small metro region containing a population of approximately one million individuals. The majority of Kisumu households are classified as low or middle socioeconomic status, with substantial variation in income levels and access to basic household assets (Were et al., 2022). Thus, Kisumu children are likely to have more variation in their picture experiences, compared to U.S. convenience samples.

Children were recruited from a local database maintained by a local non-profit organization, the Safe Water and AIDS Project (SWAP), and participated in the experiment in a quiet room at the SWAP office. The experiment was approved by a U.S. university's Committee for the Protection of Human Subjects, as well as by the Kenya Medical Research Institute (KEMRI), the National Commission for Science, Technology, and Innovation (NACOSTI), and the local Kisumu County government. All parents of child participants provided informed consent.

Stimuli & Procedure

We designed a Swahili vocabulary task involving four blocks, each presented in a different format (e.g., objects, photographs, cartoons, black-and-white line drawings). Each block consisted of four trials, leading to a total of sixteen trials. The order of the blocks, as well as the order of the trials within each block, was semi-randomized across participants.

Target words were selected from Swahili Communicative Development Inventory (CDI) production data from two communities in Coastal Kenya (Alcock et al., 2015). We used CDI production data, rather than comprehension data, because the former had more items and less noise in the data. We found approximately 76 viable words (i.e., typically artifact nouns) in the Swahili CDI. We split these words into four quartiles by age of acquisition. Each block included one target word from each of the four quartiles, ensuring that the four blocks were relatively similar in difficulty.

In addition to the target word, each array consisted of one near distractor and two random distractors. Near and random distractors were all artifacts that Kisumu children were relatively familiar with. Many distractors were items on the Swahili CDI (Alcock et al., 2015) or used in previous research with Kenyan preschoolers (Zhu et al., 2024). Furthermore, local SWAP employees again confirmed that all artifacts were familiar to Kisumu children. Similarity scores were calculated through correlations from the THINGS database (Hebart et al., 2019). Near distractors were conceptually similar to, and had high correlations with, target items (average correlation = 0.82, range = 0.66-0.95). Random distractors were less conceptually similar to, and had lower correlations with, target items (average correlation = 0.36, range = 0.17-0.53).

Using a within-subjects design, each participant completed all sixteen vocabulary trials. Across participants, each target word was presented in all four formats an equal number of items. On each trial, a local experimenter presented children with either a laminated piece of paper depicting four objects, or a tray with four objects (see Fig. 1) 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.

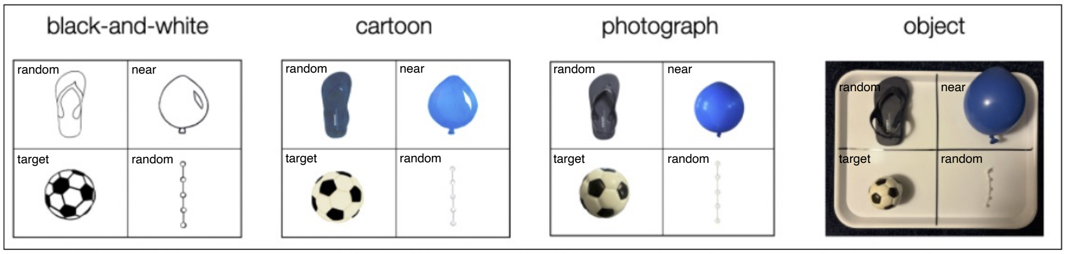

Figure 1: Stimuli for a noun trial (Show me the ball / Nionyeshe mpira) across the four conditions.

	Target Item (Swahili)	Near Distractor	Random Distractor	Random Distractor
Set 1	Comb (Kichana)	Brush	Watch	Bracelet
	Shoe (Kiatu)	Sock	Medicine	Telephone
	Chalk (Chaki)	Crayon	String	Handkerchief
	Fork (Uma)	Spoon	Straw	Necklace
Set 2	Cup (Kikombe)	Bottle	Soap	Boots
	Shirt (Shati)	Hat	Eraser	Bucket
	Scissors (Makasi)	Knife	Bag	Bib
	Button (Kifungo)	Stone	Stick	Shovel
Set 3	Book (Kitabu)	Paper	Jeans	Coat
	Bowl (Bakuli)	Plate	Pants	Scarf
	Box (Sanduku)	Basket	Bottle Cap	Key
	Shorts (Kaptula)	Dress	Clock	Remote Control
Set 4	Ball (Mpira)	Balloon	Earring	Sandal
	Toothbrush (Mswaki)	Toothpaste	Mirror	Bead
	Sweater (Sweta)	Skirt	Glasses	Toilet Paper
	Pencil (Penseli)	Pencil Sharpener	Towel	Zipper

Table 1: Full list of stimuli.

Results

Confirmatory Analyses

We fit generalized linear mixed-effects models to children's scores to assess the factors that influenced their response accuracy (Bates, 2010). A first preregistered mixed-effects model was fit to the response scores (accurate/inaccurate), including fixed effects for format (object/photograph/cartoon/black-and-white line drawing). We initially planned to include random slopes for the effect of condition on each child, and condition on each item. However, the model with this random effects structure failed to converge; per lab standard operating procedures, the reported model used the maximal random effects structure that did converge, which included random intercepts for each child and item. The model yielded a main effect of format, such that children were more accurate in the object format than the black-and-white line drawing format ($\hat{\beta} = -0.44$, 95% CI $[-0.78, -0.10]$, $z = -2.55$, $p = .011$). There was no difference between children's accuracy in the object format and the photograph format ($\hat{\beta} = 0.00$, 95% CI $[-0.34, 0.34]$, $z = -0.01$, $p = .994$), or in the object format and the cartoon format ($\hat{\beta} = -0.11$, 95% CI $[-0.45, 0.23]$, $z = -0.65$, $p = .517$).

We also ran a second preregistered mixed-effects model including child age as an additional variable. This second mixed-effects model was fit to the response scores (accurate/inaccurate), including fixed effects for format (object/photograph/cartoon/black-and-white line drawing) and age (in years), and the interactions between format and age. We initially planned to include random slopes for the effect of condition on each child, and the three-way interaction between condition, age, and item. However, the model with this random effects structure again failed to converge, and the reported model used the maximal random effects structure that did converge, which included random intercepts for each child and item. The model yielded a main effect of age, such that older children were more accurate ($\hat{\beta} = 0.58$, 95% CI $[0.39, 0.77]$, $z = 6.03$, $p < .001$), and a main effect of format, such that children were more accurate in the object format than the black-and-white line drawing format ($\hat{\beta} = -0.48$, 95% CI $[-0.82, -0.14]$, $z = -2.74$, $p = .006$). There was a significant interaction between the black-and-white line drawing format and age, such that the accuracy difference between the object and black-and-white line drawing formats was greater for older than younger children ($\hat{\beta} = -0.31$, 95% CI $[-0.56, -0.06]$, $z = -2.47$, $p = .014$). All other main effects and interactions were not

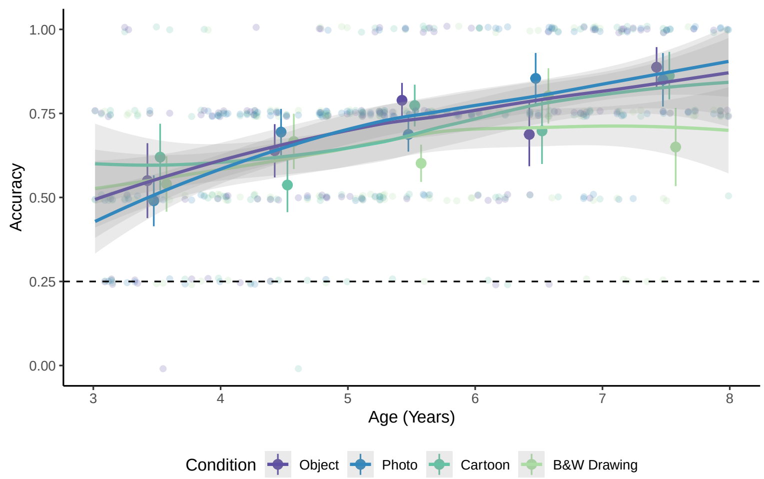

Figure 2: Proportion of accurate responses across conditions and age.

significant (all p's > .20).

Exploratory Analyses

Item effects Response accuracy was lowest in the black-and-white condition on only 6 out of 16 items, with extremely low response accuracy in the black-and-white condition on 3 out of 16 items (i.e., "book", "button" and "fork"; see Fig. 3). Thus, in order to explore the model using the maximal random effects structure accounting for more item-level variation, which did not converge in the preregistered regressions, we also conducted exploratory Bayesian analyses. We analyzed the data with the `brms` R package (Bürkner, 2017) using the default flat priors. Each model underwent a warm- up period of 1000 iterations, followed by four sampling chains with 2000 iterations each.

A first exploratory Bayesian mixed effects model was fit to the response scores (accurate/inaccurate), including fixed effects for format (object/photograph/cartoon/black-and-white line drawing), and random slopes for the effect of condition on each child, and condition on each item. We used the `brms` default prior structure, flat priors on fixed effects and t distributions on random effects. This model instantiated the preregistered maximal random effects structure that failed to converge. The Bayesian model converged, Gelman-Rubin statistic = 1.00. There was no difference in performance between object and photograph formats, emmeans: -0.08, 95% credible interval [-0.67, 0.5], object and cartoon formats. emmeans: 0.02, 95% credible interval [-0.75, 0.75], or object and black-and-white line drawing formats, emmeans: 0.41, 95% credible interval [-0.25, 1.15].

We also ran a second exploratory Bayesian mixed-effects model including child age as an additional variable, again paralleling our earlier analysis but with the full random effects structure. This second mixed-effects model was fit to the response scores (accurate/inaccurate), including fixed effects for format (object/photograph/cartoon/black-and-white line drawing) and age (in years), and random slopes for the effect of condition on each child, and the three-way interaction between condition, age, and item. This Bayesian model also converged, Gelman-Rubin statistic = 1.00. Once again, there was no difference in performance between object and photograph formats, emmeans: -0.15, 95% credible interval [-0.78, 0.44], object and cartoon formats, emmeans: 0.05, 95% credible interval [-0.7, 0.74], or object and black-and-white line drawing formats, emmeans: 0.63, 95% credible interval [-0.07, 1.31].

We were curious whether the difference between these Bayesian models and the earlier preregistered frequentist analyses was due to their differing random effect structures. To investigate this question, we ran Bayesian analyses with the same random effects structure as the convergent preregistered models. Thus, we fit parallel exploratory Bayesian mixed effects models for condition both with and without age interactions, in both cases omitting random condition by item slopes. Paralleling the preregistered analyses, there was a difference between object and black-and-white line drawing formats, emmeans: 0.45, 95% credible interval [0.08, 0.8] in the condition model, and the same effect emerged in the model with age interactions emmeans: 0.48, 95% credible interval [0.15, 0.84]. These models provide support for the idea that changes to the random effect structure were responsible for changes in the significance of the black-and-white line drawing effect between the preregistered and exploratory models.

Error analysis A further signal of children's understanding across conditions comes from the distribution of their errors. If they had more understanding of the task, they might be more likely to choose near distractors over random (far) distractors. To explore this hypothesis, we examined the distribution of children's errors across ages and conditions. Figure 4a shows the full distribution of children's responses, while Figure 4b shows the same distribution renormalized without target choices. Especially for the youngest ages, there was some evidence of greater near target choices in the object con-

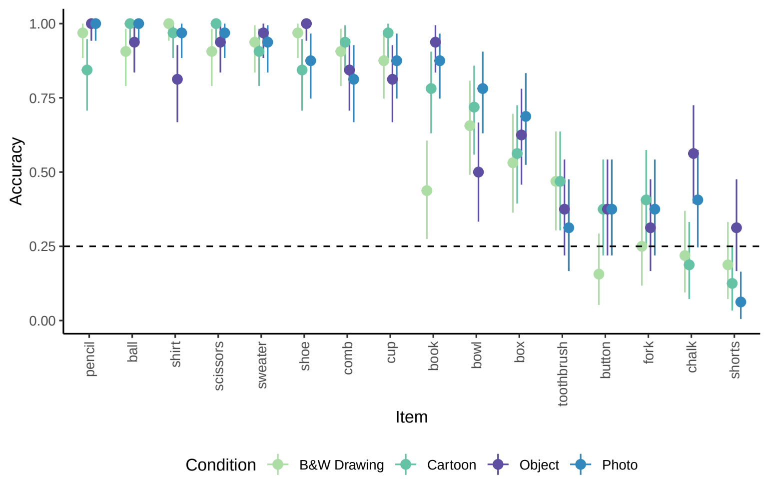

Figure 3: Response accuracy by item and condition. Error bars are binomial 95% confidence intervals.

dition than in the other conditions.

We fit a frequentist mixed effects model to these data as in our preregistered analyses, predicting the choice of a far (as opposed to a near) distractor on error trials as a function of age and condition. This model showed main effects for the photo, cartoon, and black-and-white line drawing conditions as compared to the object condition ($\hat{\beta} = 1.02$, 95% CI $[0.40, 1.64]$, $z = 3.22, p = .001$, $\hat{\beta} = 1.06$, 95% CI $[0.45, 1.67]$, $z = 3.39$, $p < .001$, and $\hat{\beta} = 1.98$, 95% CI $[1.31, 2.66]$, $z = 5.76$, $p < .001$ respectively), suggesting significantly less choice of the near distractor overall compared with the object condition. In other words, errors were less random and more influenced by similarity in the object condition overall. There was also a significant interaction of age and black-and-white line drawing condition ($\hat{\beta} = -0.68$, 95% CI $[-1.14, -0.22]$, $z = -2.91$, $p = .004$). The lower accuracy on line drawing trials for older children was mirrored by a concomitant rise in near distractor choices.

Discussion

The present experiment investigated whether young children in urban Kisumu, Kenya performed equally accurately on vocabulary assessments involving objects and three kinds of pictures (i.e., photographs, cartoons, black-and-white line drawings). Preregistered analyses suggest that young children in this context perform equally accurately on vocabulary assessments involving objects, photographs, and cartoons, but less accurately on assessments involving black-and-white line drawings. However, further exploratory Bayesian analyses with a full random effects model did not find significant differences in children's response accuracy between vocabulary assessments involving objects, photographs, cartoons, and black-and-white line drawings. These exploratory analyses may suggest that the accuracy difference between object-based and black-and-white line drawing-based assessments is not consistent across items, but rather driven by specific items. Consequently, the present experiment tentatively suggests that assessments involving black-and-white line drawings may underestimate the capacities of Kisumu children already attending daycare or formal schooling, whereas cartoon- and photograph-based assessments may be valid assessments in this population.

Exploratory analyses examining children's error patterns allow for another view of the present data, beyond binary correct/incorrect scoring. Specifically, if children's errors consist primarily of near distractors, they may still have some degree of understanding of the task at hand. Preliminary error analyses suggested that children tended to make more near distractor errors than random distractor errors when presented with objects, potentially suggesting a greater degree of understanding in the object condition than the picture conditions. However, this exploratory result was modest in size and should be interpreted with caution.

More data and more item diversity are also required in order to conclude that the specific accuracy difference between object-based and black-and-white line drawing-based assessments is generalizable. It is possible that this difference only appears for some kinds of black-and-white line drawings (e.g., particularly ambiguous depictions), or that the line drawings used in the present experiment were particularly difficult for an idiosyncratic reason. Thus, future research should attempt to replicate this effect with a new sample of items.

Prior studies showed that Mombasa preschoolers, as well as toddlers and preschoolers in other global contexts, perform more accurately on object- than photograph- and cartoon-based assessments (Walker et al., 2013; Zhu et al., 2024). What variables might account for the differences between our present findings and these prior studies? One possibility is that formal schooling may drive the discrepancy between our present findings and previous findings. Many of the children in the present studyalready attend daycares and schools, and these additional learning contexts may provide children with

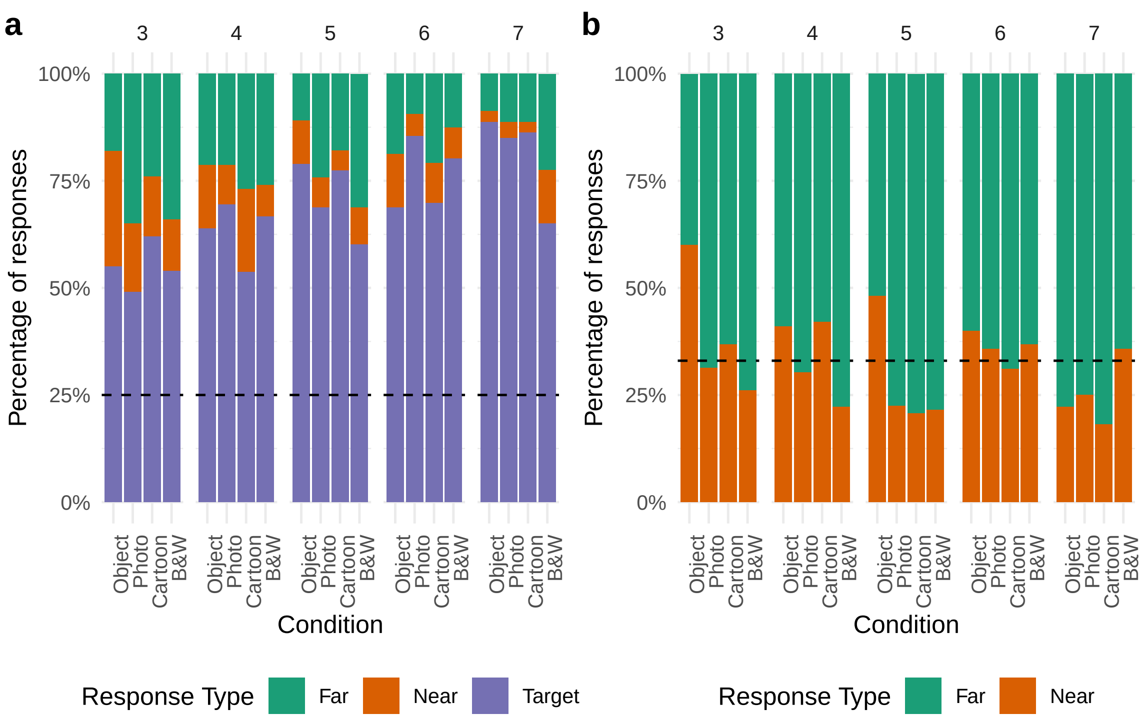

Figure 4: Error analysis showing the distribution of target and distractor choices by condition and age. (a) Distribution including target choices. (b) Distribution of errors, renormalized without target choices.

more opportunities to access picture books and other kinds of visual media. Another possibility is that the children in the present study are simply older: we worked with 3- to 7-year-olds, (average age of approximately 5.5 years), whereas other studies worked with 2-year-olds (Walker et al., 2013) and 2- to 7-year-olds, with an average age of approximately 4.5 years (Zhu et al., 2024). A final possibility is that there is something distinct about Kisumu, such that work from other Kenyan contexts (i.e., Mombasa County) does not generalize to there. Indeed, Kisumu is substantially more urban than some, though not all, previously researched global contexts, such as rural India and Peru (Callaghan et al., 2011) and rural Tanzania (Walker et al., 2013). In general, we caution that not all global contexts are the same, and do not readily generalize to each other. Instead, it is more fruitful to articulate specific variables which may vary across and within contexts, and thus drive variation in behavior (Bohn, Fong, Pope-Caldwell, Stengelin, & Haun, 2024).

Moreover, it is unclear whether U.S. children might perform similarly to their Kisumu counterparts across object- and picture-based conditions. Indeed, while previous research with toddlers and preschoolers in Berkeley, California suggest that high-income U.S. children perform equally accurately on vocabulary assessments involving objects and cartoons (Zhu et al., 2024), it is unclear whether U.S. children in general show no performance difference across other kinds of picture formats, especially across all ages. Thus, more research on picture comprehension, and possible differences in performance across picture- and object-based tasks, is required not only in understudied global contexts, but also Western convenience samples.

Overall, the present findings contribute to an exploration of assessments used in global contexts, particularly contexts where children may possess highly variable amounts of picture experience. However, more future research is required to determine in exactly which contexts picture-based assessment tools are valid or invalid. Ultimately, this research will contribute to the development of more inclusive and effective curricula for children worldwide, as well as more appropriate evaluation of policies and programs to improve early child outcomes and educational trajectories.

References

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative Development Inventory (CDI) adaptation for two Kenyan languages. *Journal of Child Language*, *42*(4), 763–785. http://doi.org/10.1017/S0305000914000403

Bates, D. M. (2010). lme4: Mixed-effects modeling with r. Springer.

Bohn, M., Fong, F. T. K., Pope-Caldwell, S., Stengelin, R., & Haun, D. B. M. (2024). Understanding cultural variation in cognition one child at a time. *Nature Reviews Psychol-*

ogy, *3*(10), 641–643. http://doi.org/10.1038/s44159-024-00351-8

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. http://doi.org/10.18637/jss.v080.i01

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–viii, 1–142. http://doi.org/10.1111/j.1540-5834.2011.00603.x

DeLoache, J. S., Pierroutsakos, S. L., & Uttal, D. H. (2003). The origins of pictorial competence. *Current Directions in Psychological Science*, *12*(4), 114–118. http://doi.org/10.1111/1467-8721.01244

Draper, C. E., Barnett, L. M., Cook, C. J., Cuartas, J. A., Howard, S. J., McCoy, D. C., … Yousafzai, A. K. (2022). 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*, No Pagination Specified–No Pagination Specified. http://doi.org/10.1002/icd.2375

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

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. http://doi.org/10.1111/j.1467-8624.2011.01612.x

Gelman, S. A., & Ebeling, K. S. (1998). Shape and representational status in children's early naming. *Cognition*, *66*(2), B35–B47. http://doi.org/10.1016/s0010-0277(98)00022-5

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Wicklin, C. V., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, *14*(10), e0223792. http://doi.org/10.1371/journal.pone.0223792

Jukes, M. C. H., Ahmed, I., Baker, S., Draper, C. E., Howard, S. J., McCoy, D. C., … Wolf, S. (2024). Principles for Adapting Assessments of Executive Function across Cultural Contexts. *Brain Sciences*, *14*(4), 318. http://doi.org/10.3390/brainsci14040318

Pierroutsakos, S. L., & DeLoache, J. S. (2003). Infants' Manual Exploration of Pictorial Objects Varying in Realism. *Infancy*, *4*(1), 141–156. http://doi.org/10.1207/S15327078IN0401_7

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. http://doi.org/10.1207/s15327647jcd0502_2

Simcock, G., & DeLoache, J. (2006). Get the picture? The effects of iconicity on toddlers' reenactment from picture books. *Developmental Psychology*, *42*(6), 1352–1357. http://doi.org/10.1037/0012-1649.42.6.1352

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. http://doi.org/10.1037/a0029483

Were, V., Foley, L., Turner-Moss, E., Mogo, E., Wadende, P., Musuva, R., & Obonyo, C. (2022). Comparison of household socioeconomic status classification methods and effects on risk estimation: Lessons from a natural experimental study, Kisumu, Western Kenya. *International Journal for Equity in Health*, *21*(1), 47. http://doi.org/10.1186/s12939-022-01652-1

Zhu, R., Kilonzo, T. N., Engelmann, J., & Gopnik, A. (2024). Investigating the validity of assessments involving picture stimuli across cultures and contexts: Evidence from young children in Kenya and the U.S. *PsyArXiv*.

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), e13579. http://doi.org/10.1111/desc.13579