Cultural Variation in Children's Narrative Comprehension and Retelling: A Comparative Study of Children in Nigeria and the United Kingdom

Igbako Odaro-Nuvie
Nuvie Film Production Limited, Gwarinpa, Abuja
igbakonuvie@yahoo.com

Osamede Benjamin Osunde
Apex Consult, Benin 74, Igbinosa Street,
Off Iriri Road, Off Airport, Benin

Nathan Oguche Emmanuel, PhD
Department of Mass Communication,
Faculty of Social Sciences,
National Open University of Nigeria, Jabi-Abuja.
emmnatslinks@gmail.com
DOI: https://doi.org/10.63725/zajcom.v11.i1.10

Abstract

The study examined whether children in Nigeria and the United Kingdom differed in narrative comprehension after accounting for caregiver schooling, print exposure and home narrative practice, and whether differences were similarly evident in open retellings and structured comprehension probes. A cross-sectional survey design was used to study 1,200 children aged 5–12 years, 600 from each of the two countries. The study participants responded to 16 comprehension probes per story, and provided an open retelling of each story using either English or Yoruba languages. Caregiver reports were obtained for selected home background factors, including caregiver schooling, print exposure, and home narrative practices. The findings of the showed that Nigerian children recorded lower overall comprehension and retelling scores than their UK counterparts. However, the country difference became statistically non-significant after accounting for measured home-background factors, suggesting that the observed difference was attributable to these factors rather than to a direct effect of nationality or culture. Predictors of comprehension varied within and between the two countries, while between-country differences were more pronounced in open-retell performance than in structured comprehension probes. Although this pattern may reflect differences in narrative conventions, language proficiency provides an important alternative explanation. This is particularly relevant because all UK participants were assessed in English, whereas only 31.2% of Nigerian participants were assessed in English. Comprehension increased with age in both countries, but the between-country difference did not systematically vary across age. The study recommends caution when interpreting retelling performance as an indicator of language or comprehension difficulties. It suggests combining retelling tasks with structured comprehension probes, while considering the assessment format, language, and home-background experiences.

Key Words: Children; Stories; Culture; Cultural Variation; Narrative.

Introduction

Sarah Michaels spent a year watching and spending time in an American first-grade classroom and noticed something uncomfortable. The teacher helped some children finish their stories and cut others off. Those cut off were mostly Black, and they told stories that moved through several linked episodes rather than staying on one topic. The teacher heard that shape as wandering off the point (Michaels, 1981). She was applying a rule about what a good story looks like, and the rule came from her own community. Forty-five years later, that possibility remains relevant to reading tests and screening tools used with children from communities whose narrative practices and linguistic experiences may differ from those assumed by the instruments.

The problem is that much of what psychology knows about how children understand stories has been developed from a relatively narrow segment of the world's children. Developmental samples remain concentrated in North America and Western Europe (Henrich et al., 2010; Singh et al., 2023; Thalmayer et al., 2021), child language research draws on a small fraction of the world's languages (Kidd & Garcia, 2022), and researchers working elsewhere face substantial barriers to publishing (Draper et al., 2023; Ghai, 2021). Models developed from these research contexts are sometimes extended to children more generally, creating a risk that a low score will be interpreted as a deficit in the child rather than as a possible mismatch between the child's linguistic or narrative experience and the assessment instrument (Karasik & Robinson, 2022; Scheidecker et al., 2022).

Three bodies of work are particularly relevant to this problem. The first treats comprehension as a process of mental construction, in which listeners build a representation of the situation described by a story using memory, language and inference (Bartlett, 1932; Kintsch, 1988; Zwaan & Radvansky, 1998), with more recent component models specifying the skills that contribute to this construction (Butterfuss et al., 2020; Kim, 2020a). The second treats narrative as a cultural and interactional practice and documents substantial variation in how children tell stories and how adults talk with them about past events (Heath, 1983; Melzi, 2000; Minami & McCabe, 1991; Wang, 2021a). The third examines the populations represented in psychological research and cautions against extending findings from relatively narrow samples to all human populations (Apicella et al., 2020; Blasi et al., 2022; Muthukrishna et al., 2020). Taken together, the literature establish an important distinction: differences in how children produce narratives do not, by themselves, demonstrate differences in how they comprehend narratives. They

however, raise a measurement question about whether the way comprehension is elicited and expressed can affect observed performance across cultural and linguistic contexts.

This distinction is important because narrative production and narrative comprehension are related but non-equivalent constructs. A child may construct an adequate representation of a story but express that understanding differently depending on the conventions of storytelling, the listener's expectations, the language used in the interaction, and the child's familiarity with the form of response expected by the assessment. Open retelling places relatively broad demands on the child: the child must select what information to report, organise it into a coherent account, and express it in language that is appropriate to the communicative situation. Structured comprehension probes, by contrast, constrain the response and can reduce some of the demands associated with organising and verbally expressing a complete narrative. Consequently, differences in retelling performance across groups cannot automatically be interpreted as differences in underlying comprehension. Comparing open retelling with structured comprehension probes in the same children provides a way to examine whether between-group differences are similar across assessment formats or become more pronounced when comprehension must be demonstrated through extended narrative production.

The existing comparative literature provides a basis for this question, but the extent to which findings from narrative production can be generalised to narrative comprehension remains uncertain. Cross-cultural studies have documented differences in children's storytelling practices, including the organisation, elaboration, and social framing of narratives (Heath, 1983; Melzi, 2000; Minami & McCabe, 1991; Wang, 2021a). At the same time, research on narrative comprehension has largely developed within cognitive and language-processing traditions that emphasise the construction of meaning from linguistic information, memory and inference (Kintsch, 1988; Zwaan & Radvansky, 1998; Butterfuss et al., 2020; Kim, 2020a). Comparative studies that directly examine both open narrative production and structured comprehension performance within the same children across different national or cultural contexts appear to be limited. This limitation makes it difficult to determine whether observed cross-group differences are specific to how children are asked to produce narratives or extend more consistently to other measures of comprehension. The problem is further complicated by measurement equivalence: unless an instrument operates with sufficient similarity across groups, an observed group difference may partly reflect

properties of the measure rather than a difference in the underlying construct (Boer et al., 2018; Peña, 2007).

Country-level comparisons also present an additional interpretive challenge because culture, home language, schooling, print exposure and home narrative practices may vary together. A difference between children in two countries may therefore reflect several correlated contextual factors rather than a direct effect of culture or nationality. The present study addresses this problem by measuring both comprehension and retelling in the same children, statistically adjusting for relevant home-background variables, and examining the implications of language differences for interpretation. It also examines measurement equivalence before interpreting between-group differences. In this way, schooling, print exposure and home narrative practice are treated as measured characteristics rather than subsumed under a broad country label, while language profile is examined separately because of its strong confounding with country.

The purpose of this comparative study is therefore to examine cultural and between-group variation in children's narrative comprehension and retelling performance among selected children in the United Kingdom and Nigeria. Rather than treating "understanding stories" as a single undifferentiated ability, the study examines children's performance across the narrative comprehension dimensions represented in the assessment tasks, including their responses to structured comprehension probes and their open retelling of stories. The study asks whether differences between the two groups remain after statistical adjustment for schooling, print exposure, home language and home narrative practice, and whether the pattern of between-group differences is consistent across assessment formats. Two research questions flow from this purpose:

1. Do children in the United Kingdom and Nigeria differ in narrative comprehension after statistical adjustment for caregiver schooling, print exposure and home narrative practice?
2. Does the magnitude of any between-group difference vary between open retelling and structured comprehension probes?

Each research question is tested against a corresponding null hypothesis. The first null hypothesis states that there is no statistically significant difference in narrative comprehension between children in the United Kingdom and Nigeria after adjusting for schooling, print exposure, home language, and home narrative practice.

H01: There is no statistically significant difference in narrative comprehension between children in Nigeria and those in the United Kingdom after adjusting for caregiver schooling, print exposure, and home narrative practice.

The second null hypothesis states that the magnitude of the between-group difference does not vary significantly between open retelling and structured comprehension probes.

H02: There is no statistically significant difference in the magnitude of the between-group difference between open retelling and structured comprehension probes.

The directional alternative to the second hypothesis is derived from the measurement demands of the two assessment formats rather than from an assumption that one culture has a particular way of understanding stories. Open retelling requires children not only to demonstrate what they have understood but also to select, organise, and verbally express story information without the response structure provided by a probe. It may therefore be more sensitive to differences in language proficiency, familiarity with expected narrative forms and interactional conventions than structured probes, which place greater constraints on the required response. On this basis, the study predicts that between-group differences will be larger on open retelling than on structured comprehension probes. This prediction is treated as a testable difference in assessment performance; it is not taken in advance as evidence that the groups possess fundamentally different ways of understanding narratives.

Literature and Theoretical Review

Where the Idea of Narrative Processing Came From

Frederic Bartlett gave Cambridge students a Native American folk tale, "The War of the Ghosts," and asked them to recall it over several weeks. Their versions grew shorter, tidier and more English, with supernatural elements dropped and unfamiliar actions replaced by familiar ones (Bartlett, 1932). Remembering, he concluded, is reconstruction guided by schemas, meaning organised expectations built from experience. That is the foundational demonstration that culture shapes narrative memory, and it is the study that the field then spent 50 years failing to follow up on across cultures.

What the field did instead was formalise the schema. Story grammar research broke stories into settings, initiating events, attempts and consequences, and showed that children recall structurally central material best (Stein & Glenn, 1979). Kintsch

(1988) modelled comprehension as construction followed by integration, in which many meanings are activated, and context settles which survive. Zwaan and Radvansky (1998) argued that what a listener ends up with is a situation model, a representation of the events rather than of the sentences. Current models specify component skills and the relations among them (Kim, 2020a, 2020b).

The definition has moved three times: from recall of structure, to construction of a situation model, to a multi-component skill. The components singled out were causal structure, inference and evaluation, which is not accidental. These are what a written school-genre story rewards, and they are countable, so they became scoring categories and then came to mean comprehension operationally. The well-formed story of the research literature is recognisably a school genre, formalised alongside mass schooling and standardised testing.

What the Cultural Literature Established

Michaels (1981) distinguished topic-centred narratives, which develop one event in depth, from topic-associating narratives, which link several episodes around an implicit theme. Both are coherent; only one was heard as coherent. Heath (1983) found the same across three neighbouring Carolina communities, where children arrived at school with different assumptions about what a story is for and who may tell one. Minami and McCabe (1991) found Japanese children producing terse three-verse narratives about collections of similar experiences, a shape that their mothers encouraged and that North American norms would score as underdeveloped. Berman and Slobin (1994) had children in five languages narrate the same picture book and found that the language itself drew attention to different features of the same events.

A parallel line examined the conversations in which these habits are learned. Nelson and Fivush (2004) argued that autobiographical memory emerges through talk about the past with adults, and that talk differs in patterned ways: some caregivers elaborate and draw the child out, others prompt briefly and centre the group (Melzi, 2000; Wang, 2021b). Culture, on this account, does not decorate memory but constitutes it, entering at encoding, function, reconstruction, and expression (Wang, 2021a).

Two Camps, and What Divides Them

Knowledge here is organised around two camps rather than chronological waves. The cognitive camp treats narrative processing as general human architecture, with culture supplying the inputs. The sociocultural camp, running from Vygotsky (1978) through Bruner (1990) and Rogoff (2003), treats the processing itself as assembled

through participation in culturally organised activity, so a child in a community where elders tell stories at night is not running the same routine as a child read to from books at bedtime. Cultural psychology adds a reason to expect this: habits of attention and self-construal differ across societies, some favouring the focal object and others the field and its relations (Markus & Kitayama, 1991; Nisbett et al., 2001). A third, more critical position holds that the question cannot be settled on current evidence and that treating narrow findings as universal does harm (Barrett, 2020; Scheidecker et al., 2023).

The camps are usually read as mutually contradictory. They are better read as answering different questions that look alike: one asks what the mind does with a story, the other what a community does with stories and with children. A finding that Japanese children tell shorter narratives does not rebut the construction-integration model, because it never measured construction or integration. This is where the literatures fail to meet, and where this study places itself.

The Deficit Reading, and Where It Comes From

The most consequential debate is whether difference should be read as a deficit, and it is usually conducted as though it were about attitudes. It is at least as much about instruments. If the standard measure is a retell scored against school-genre story grammar, a child whose community rewards a different shape score lower, and the method cannot separate understanding the story from telling it the expected way. Nicolopoulou (2008) showed that young children's spontaneous stories achieve coherence by several routes, not only the canonical plotline, so the template selects one route and hides the rest. Work in African settings has made a parallel argument about intelligence, in which locally valued competence includes social responsibility that no imported test scores can match (Serpell, 2011; see also Lancy, 2022).

Two findings sharpen this. Casillas and colleagues found that children in a Tseltal Mayan village and a Papuan community hear little speech directed at them and still acquire language on schedule (Casillas et al., 2020, 2021; Rowe & Weisleder, 2020). The reminiscing literature treats elaborative adult talk as the engine of narrative skill. If both are right, either the engine is not necessary, or these children reach narrative competence by a route nobody has modelled, and each possibility changes what supporting narrative development means in practice.

Methods Used, and the Assumptions Inside Them

Four traditions dominate, each with assumptions worth naming. Experimental comprehension studies assume a scored response is a window onto an internal

representation. Ethnographic work assumes meaning is local and that the researcher's position shapes what gets recorded (Heath, 1983). Naturalistic recording assumes sampled behaviour represents daily life (Casillas et al., 2020). Large comparative studies assume score comparability, the very assumption invariance testing exists to check (Boer et al., 2018), and multi-site collaborations show how much variance sits in sites and protocols rather than in children (ManyBabies Consortium, 2020). The moral assumptions are stated less often. Whose narrative counts as data, who sets the scoring criteria and who receives the findings are all choices, made in this field overwhelmingly in one part of the world (Broesch et al., 2020; Nketia et al., 2021).

The Gap and the Contribution

The mechanics of comprehension are well described in one population, and variation in narrative practice is well described across many populations. What has not been done is to survey comprehension with demonstrated measurement equivalence in children whose narrative practices are known to differ, while treating schooling, print exposure, and language as measured variables rather than as part of a country label. This survey does that in the United Kingdom and Nigeria, and proposes a framework, culturally situated narrative processing, with three separable layers: shared machinery for tracking events and causes; culturally tuned schemas setting what counts as a well-formed story; and a task layer governing what a child judges it appropriate to say to an adult who has just told them a story. A survey cannot observe the first layer directly, but it can test the claim indirectly by asking whether a difference that appears when children retell a story persists when they answer probes that require far less of their talk.

Methodology

The study uses a cross-sectional comparative survey design. Two samples of children from selected locations in the United Kingdom and Nigeria were surveyed at a single point in time, using the same assessment framework and parallel measurement procedures administered to children, caregivers, and teachers in both countries. A cross-sectional comparative survey is appropriate for the study because testing measurement equivalence and examining between-group differences require sufficiently large samples, while the central concern is whether patterns observed in one research setting are also evident in another (Amir & McAuliffe, 2020). The design also permits comparison of children's performance across age groups, although such comparisons reflect age-group differences rather than developmental trajectories, as the same children are not followed over time.

The study examines narrative comprehension through two complementary forms of assessment: structured comprehension probes and open story retelling. The use of both measures makes it possible to determine whether between-group differences are consistent across assessment formats or become more pronounced when children are required to organise and verbally express a story independently. The survey design does not permit direct observation of changes in comprehension within individual children over time, nor does it establish causal relationships between cultural, linguistic or home-background factors and performance. Accordingly, findings concerning age are interpreted as cross-sectional differences in performance across age groups, rather than evidence of individual developmental change.

Population of the Study

The study population comprises children aged 5 to 12 years attending primary schools in selected locations in Nigeria and the United Kingdom. The findings are interpreted based on the sampled population not as nationally representative estimates. This is because the study did not draw samples from all regions of either country,

Sampling Technique

A multistage stratified cluster sampling procedure was employed in each country. The strata are locality (urban and rural) and school type (state and private), yielding four strata in each national sample. Schools served as the primary sampling units and were selected with probability proportional to enrolment size, after which classes were sampled within schools. In Nigeria, Lagos State represents the urban strata and Oyo State represents the rural strata. In the United Kingdom, the sample was drawn from a mid-sized English city and a rural district within the same region of England. This stratification is intended to ensure variation in schooling experiences, school characteristics and print exposure within each country sample rather than confounding these factors with broad country categories.

The sample comprises 1,200 children - 600 from the selected UK sites and 600 from the selected Nigerian sites. Participants were distributed evenly across three age bands: 5–6 years, 8–9 years and 11–12 years. One caregiver and one class teacher were included for each participating child. The sample size was based on two considerations. First, multi-group factor analytic procedures used to evaluate measurement equivalence generally require sufficiently large group sizes; approximately 300 cases per group is often considered adequate for stable parameter estimation. Second, the proposed sample provides approximately 90% statistical power to detect a country-by-age interaction with d = 0.30 after

accounting for a design effect of 1.5 from cluster sampling (Kumle et al., 2021; Lakens, 2022). It should be noted that the comparisons in this study are best understood as comparisons between selected samples of children from specific locations in the United Kingdom and Nigeria, rather than as estimates of differences between all children in the two countries.

Instrument of Data Collection

Three instruments were used. The Child Narrative Comprehension Survey delivers two recorded stories in the language the child chooses, each followed by 16 structured items across four families: literal detail, causal inference, mental-state inference, and evaluation. One story is the wordless picture sequence from the crosslinguistic frog story research, which allows comparison with an existing corpus (Berman & Slobin, 1994); the other is written by a local storyteller in each country and matched to the first for episodes, characters and causal links. Each story also carries one open retell, scored for structure. Probes and retell are kept apart on purpose because probes make far fewer demands on how a child speaks to an adult.

The Home Narrative Practice Questionnaire asks caregivers who tell stories when, in which language, and in what genre, along with print exposure, caregiver schooling, and household resources, and includes three open items on what makes a good story and how a caregiver can tell that a child has understood one. A short school form records the medium of instruction, class size and years of schooling completed.

Instruments were written in English and Yoruba in parallel rather than translated at the end, reviewed by a bilingual committee, and refined through cognitive interviewing with 12 children and 8 caregivers per country, and then piloted with 120 children for item analysis and reliability (Peña, 2007).

Procedure

Fieldworkers recruited from each community administer the child survey face-to-face during a 35-minute school session, using recorded stories and a fixed script so that delivery does not vary with the interviewer (ManyBabies Consortium, 2020). Caregiver copies of questionnaire wee self-completed where literacy allows and read aloud where it does not, with the mode of completion recorded as a covariate. The collection runs across one school term in each country.

Data Analysis

Analysis follows a fixed order, and all hypothesis tests are two-tailed at an alpha level of .05. Descriptive statistics, missingness patterns and item-level statistics come first, with multiple imputation used where data are missing at random.

Multi-group confirmatory factor analysis then tests configural, metric, and scalar invariance of the comprehension model across countries and age bands before any means are compared. Nested models are compared using a chi-square difference test and changes in fit indices, taking a drop in CFI greater than .010 and a rise in RMSEA greater than .015 as evidence against invariance. Where scalar invariance fails, partial invariance models identify the items carrying the bias, and those items are reported rather than quietly dropped, because they are themselves evidence about the instrument (Boer et al., 2018; Kline, 2023). This step is a precondition, not a formality: if no defensible partial-invariance solution can be reached, the group means are not comparable, and the first null hypothesis cannot be tested.

The first null hypothesis is then tested in a linear mixed-effects model of comprehension score, with children nested within classes and classes within schools, fitted using restricted maximum likelihood and following current reporting guidance for such models (Meteyard & Davies, 2020). Country, age band, locality, caregiver schooling, print exposure, language profile, and home narrative practice are included as fixed effects, with random intercepts for school. The country coefficient is tested with Satterthwaite-approximated degrees of freedom, and the null is rejected if $p < .05$. Effect sizes are reported as standardised mean differences with 95% confidence intervals, since a significant result in a sample of this size may still be small. An exploratory indirect-effect analysis was conducted to examine whether home narrative practice statistically accounted for part of the association between country and comprehension.

The second null hypothesis is tested last, and it is the study's central test. Retell and probe scores are standardised and entered as repeated measures within each child, yielding a country-by-measure-type interaction. The null is rejected if the interaction is significant at $p < .05$, meaning the size of the country gap depends on which measure is used. Assumptions of normality, homoscedasticity, and the absence of influential cases are checked before interpretation, and when multiple comparisons are made, the Benjamini-Hochberg correction is applied to control the false discovery rate. Open caregiver responses are analysed with reflexive thematic analysis and read against the quantitative pattern (Braun & Clarke, 2022). Findings are generalised to the sampled strata rather than to Britain or Nigeria as a whole (Yarkoni, 2022).

Ethics and Community Partnership

The study was approved in both countries and is subject to an agreement with a Nigerian university that covers data ownership, authorship, and the return of findings to participating schools (Broesch et al., 2020). Caregivers gave consent, and children gave assent in the local language. Neither sample is described in terms of what it lacks relative to the other; each school receives a plain-language report before publication, and contributing storytellers are credited as authors of their own materials.

Limitations

The study has several limitations that should be considered when interpreting its findings. First, the cross-sectional design means that age differences are observed between children of different ages rather than within the same children over time. Consequently, the finding that comprehension varies with age should be interpreted as a cross-sectional age-related pattern and not as evidence of an individual developmental trajectory. The design can identify associations and patterns of between-group variation, but it cannot directly observe changes in narrative processing within individual children over time.

Second, the study does not directly observe the cognitive processes involved in narrative comprehension. Performance on structured comprehension probes and open retelling provides behavioural evidence of children's responses to stories, but it cannot precisely establish how individual children construct, organise, or retrieve narrative representations during comprehension. The contrast between the two assessment formats, therefore, provides evidence about differences in observed performance rather than a direct measure of the underlying processing mechanisms.

Third, the study's geographical coverage limits the generalisability of the findings. The Nigerian sample was drawn from Lagos and Oyo States, while the UK sample was drawn from a mid-sized English city and a rural district within one English region. These selected sites cannot represent the full linguistic, cultural, educational and socioeconomic diversity of Nigeria or the United Kingdom. The findings should therefore be interpreted as applying primarily to the sampled populations and comparable contexts rather than to children nationally.

Fourth, the study relies partly on caregiver reports of home narrative practice, print exposure, and related home background characteristics. Such measures are subject to the limitations of self-report, including recall error and differences in how caregivers perceive or report children's experiences. Statistical adjustment for these variables therefore reduces, but does not eliminate, the possibility that unmeasured

contextual factors contribute to observed between-group differences. Most importantly, language represents a substantial limitation.

Results

Table 1: Sample characteristics by country (N = 1,200)

Characteristic	United Kingdom	Nigeria	Test	*p*
Children (*n*)	600	600	—	—
Schools (*n*)	20	20	—	—
Female	273 (45.5%)	287 (47.8%)	$\chi^2(1) = 0.66$	.418
Age (months)	107.63 (29.95)	107.33 (30.34)	$t(1198) = 0.17$	.867
Rural locality	300 (50.0%)	300 (50.0%)	$\chi^2(1) = 0.00$	1.000
Private school	300 (50.0%)	300 (50.0%)	$\chi^2(1) = 0.00$	1.000
Class size	26.00 (4.04)	35.05 (6.09)	$t(1040) = -30.34$	< .001
Caregiver schooling (years)	14.85 (2.89)	10.69 (3.18)	$t(1139) = 23.25$	< .001
Books at home	10.66 (6.56)	5.13 (4.09)	$t(963) = 17.10$	< .001
Print exposure index	35.98 (14.07)	12.28 (12.85)	$t(1140) = 29.82$	< .001
Oral storytelling (0–20)	8.03 (3.78)	13.66 (3.45)	$t(1135) = -26.33$	< .001
Shared reading (0–20)	13.83 (3.48)	7.27 (3.39)	$t(1150) = 32.46$	< .001
Home narrative practice (0–40)	21.90 (5.27)	20.91 (4.67)	$t(1125) = 3.37$	< .001
Questionnaire read aloud	88 (14.7%)	215 (35.8%)	$\chi^2(1) = 71.21$	< .001

Tested in English	600 (100.0%)	187 (31.2%)	$\chi^2(1) = 629.73$	< .001

Note. Continuous variables are M (SD). Welch's t is reported for unequal variances.

In Table 1, the stratification worked. The two samples are matched on sex, age, and the two design strata, so any country differences cannot be attributed to these. They diverge sharply everywhere else. Nigerian caregivers completed about 4 fewer years of schooling, and Nigerian children had roughly a third as much print exposure and half as many books at home. The narrative practice variables move in opposite directions: Nigerian children heard substantially more oral storytelling, UK children substantially more shared book reading, and the composite therefore differs by less than one point on a forty-point scale, despite both components differing by more than five. Two operational asymmetries also emerged. Caregiver questionnaires were read aloud more than twice as often in Nigeria, which accounts for the missingness pattern handled by imputation. All UK children were tested in English, whereas 31.2% of Nigerian children were tested in English, meaning the language profile is almost entirely confounded with country and cannot be included in the models as a separate predictor.

Table 2: Measurement invariance of the narrative comprehension model across country

Panel A. Multi-group confirmatory factor analysis (8 parcels, N = 1,200)

Model	**χ^2**	**df**	**CFI**	**RMSEA**	**$\Delta\chi^2$**	**Δdf**	***P***	**ΔCFI**	**ΔRMSEA**
1. Configural	63.67	40	.995	.031	—	—	—	—	—
2. Metric (loadings equal)	83.44	47	.992	.036	19.77	7	.006	−.003	+.005
3. Scalar (intercepts equal)	94.31	54	.991	.035	10.87	7	.144	−.001	−.001
4. Partial scalar (2 intercepts freed)	87.74	52	.992	.034	4.31	5	.506	+.000	−.002

Panel B. Item-level differential item functioning (32 items screened)

Item	*B*	SE	*P*	*p* (BH)	Flagged
storyA_p14 — Evaluation	−0.981	0.151	< .001	< .001	Yes
storyB_p13 — Evaluation	−0.860	0.140	< .001	< .001	Yes
storyA_p10 — Mental-state inference	+0.246	0.136	.069	.706	No
storyA_p03 — Literal detail	−0.215	0.138	.121	.706	No

Note. Panel B reports logistic regression DIF with rest score as the ability proxy; b is the country coefficient. Benjamini-Hochberg correction across all 32 items. The four largest effects are shown.

Panel A supports comparing the two countries. Configural fit was good, and although the metric constraint produced a significant chi-square difference, ΔCFI of −.003 and ΔRMSEA of +.005 both fell inside the pre-specified bounds of .010 and .015, so loading invariance was retained. Scalar constraints cost nothing further. The aggregate invariance criteria supported scalar invariance and permitted comparison of group means. However, item-level DIF screening identified two evaluation items that differed in functioning. Thus, although the overall measurement model met the pre-specified invariance criteria, localised item-level bias remained and was considered in interpreting the country comparison.

Panel A alone would have been misleading. Item-level screening flagged two of thirty-two items after correction, both from the evaluation family, each roughly 0.9 logits harder for Nigerian children of equal measured ability. No other item approached significance, the largest remaining coefficient being 0.246. Aggregating items into parcels had diluted the real, localised bias to the point that it disappeared from the fit indices. The two items are retained in the reported totals but identified here, since they are evidence about the instrument rather than about the children.

Table 3: Narrative comprehension and retelling scores by country and age band

Measure	Country	Age 5–6	Age 8–9	Age 11–12	All ages	*d*
Literal detail (0–8)	United Kingdom	4.41 (2.03)	5.96 (1.68)	6.75 (1.43)	5.71 (1.98)	0.45
	Nigeria	3.37 (2.02)	4.84 (2.02)	6.08 (1.58)	4.76 (2.19)	
Causal inference (0–8)	United Kingdom	2.71 (1.89)	4.28 (2.15)	5.72 (1.94)	4.24 (2.34)	0.41
	Nigeria	1.76 (1.79)	3.15 (2.00)	4.92 (1.98)	3.28 (2.32)	
Mental-state inference (0–8)	United Kingdom	2.06 (1.68)	3.67 (2.11)	4.86 (1.83)	3.53 (2.20)	0.39
	Nigeria	1.50 (1.36)	2.67 (1.98)	3.90 (1.99)	2.69 (2.05)	
Evaluation (0–8)	United Kingdom	1.56 (1.58)	2.89 (1.96)	4.33 (2.01)	2.93 (2.17)	0.51
	Nigeria	0.84 (1.10)	1.66 (1.56)	3.20 (2.01)	1.90 (1.87)	
Probe total (0–32)	United Kingdom	10.74 (5.85)	16.80 (6.63)	21.66 (6.01)	16.40 (7.61)	0.50
	Nigeria	7.48 (4.95)	12.32 (6.28)	18.09 (6.41)	12.63 (7.33)	
Retell structure (0–24)	United Kingdom	9.47 (3.97)	13.46 (4.10)	16.08 (3.83)	12.98 (4.81)	0.92
	Nigeria	5.28 (3.38)	8.57 (3.76)	12.10 (3.83)	8.64 (4.60)	
Retell causal links (0–16)	United Kingdom	5.47 (3.21)	8.46 (3.63)	10.28 (3.35)	8.05 (3.94)	0.77
	Nigeria	2.74 (2.25)	5.19 (3.23)	7.46 (3.59)	5.12 (3.63)	

Note. Cells are M (SD). Cohen's d is the unadjusted country difference pooled across age bands, positive values favouring the United Kingdom.

Scores rose steadily with age in both countries and on every measure, and the ordering of the four probe families was identical in each: literal detail easiest, then causal inference, then mental-state inference, with evaluation hardest. UK children scored higher throughout. The magnitude of that advantage is where the table becomes informative. Across the four probe subscales, the unadjusted effect sizes cluster tightly between $d = 0.39$ and 0.51, with a total of 0.50. On retelling, the same children in the same session produced a gap of 0.92 on structure and 0.77 on causal links, roughly double the probe gap. Nothing in the probe results anticipates that. The age gradients also run parallel rather than converging, with the probe difference at ages 11–12 similar in magnitude to that at ages 5–6, providing no indication that the gap closes as children move through primary school.

Table 4: Multilevel models predicting probe comprehension ($H0_1$)

Predictor	Model 1 *b* (SE)	*P*	Model 2 *b* (SE)	*P*	95% CI (Model 2)
Intercept	+0.245 (0.073)	< .001	−0.567 (0.094)	< .001	[−0.751, −0.383]
Nigeria (vs United Kingdom)	−0.489 (0.104)	< .001	−0.086 (0.097)	.379	[−0.277, +0.105]
Age band 8–9 (vs 5–6)	—	—	+0.685 (0.049)	< .001	[+0.588, +0.782]
Age band 11–12 (vs 5–6)	—	—	+1.372 (0.049)	< .001	[+1.275, +1.468]
Rural locality (vs urban)	—	—	−0.168 (0.089)	.058	[−0.342, +0.006]
Private school (vs state)	—	—	+0.017 (0.090)	.849	[−0.160, +0.194]
Caregiver schooling (*z*)	—	—	+0.219 (0.038)	< .001	[+0.145, +0.293]
Print exposure (*z*)	—	—	+0.106 (0.038)	.005	[+0.031, +0.180]
Home narrative practice (*z*)	—	—	+0.085 (0.021)	< .001	[+0.043, +0.126]

School-level variance	0.078		0.058		—
Residual variance	0.866		0.483		—
Intraclass correlation	0.083		0.107		—

Note. Outcome is standardised probe total. Random intercepts for school. Estimates pooled across five imputations by Rubin's rules. Language profile was excluded because all UK children were tested in English (Table 1).

In Table 4, Model 1 reproduces the raw comparison: Nigerian children scored 0.489 SD below UK children, a difference significant beyond conventional thresholds. Model 2 adds age, locality, school type and the three home-background covariates. The country coefficient falls to −0.086 and is no longer significant, with a confidence interval running from −0.277 to +0.105 that comfortably contains zero. $H0_1$ was not rejected. Roughly four-fifths of the apparent national gap is accounted for by caregiver schooling, print exposure and home narrative practice, each of which predicts comprehension in its own right. Caregiver schooling carries the largest single effect. Age dominates the model, as expected. School type contributes nothing once home background is included, and rural residence falls just short of significance. School-level variance is modest but non-trivial, with an intraclass correlation of .107 in the adjusted model, confirming that the multilevel specification was necessary.

Table 5: Country by measure-type interaction ($H0_2$)

Effect	*B*	SE	*T*	*P*	95% CI
Intercept	−0.437	0.043	−10.13	< .001	[−0.521, −0.352]
Nigeria (vs United Kingdom)	−0.489	0.044	−11.13	< .001	[−0.576, −0.403]
Retell measure (vs probe)	+0.164	0.018	9.25	< .001	[+0.129, +0.198]
Nigeria × retell measure	−0.339	0.025	−13.41	< .001	[−0.388, −0.289]

Age band 8–9 (vs 5–6)	+0.702	0.052	13.58	< .001	[+0.601, +0.804]
Age band 11–12 (vs 5–6)	+1.341	0.052	25.90	< .001	[+1.240, +1.443]
Country gap on probes (SD)	−0.489				
Country gap on retelling (SD)	−0.838				

Note. Both outcomes standardised and entered as a repeated measure within child. Random intercepts for School. Estimates pooled across five imputations.

The interaction is the study's central test, and it is unambiguous. The country gap depends on which measure is used: −0.489 SD on structured probes and −0.838 SD on open retelling, a difference of 0.339 SD that is significant, with a confidence interval nowhere near zero. $H0_2$ is rejected. Because both measures were obtained from the same children during the same session and involved the same stories, the difference is less likely to reflect differences in sample composition or story exposure. What changes between the two is what the child must do: select an answer, or produce an extended spoken account judged against a structural rubric. The main effect of the measure type is small and positive, indicating the retell was not simply harder for everyone. The age terms are near-identical to those in Table 4, confirming that the developmental gradient is stable across measurement approaches.

Discussion

This survey set out to establish whether children in the United Kingdom and Nigeria understand the same stories in the same way, and whether any differences between them persist once schooling, print exposure, home language, and home narrative practice are taken into account. Two findings carry the argument. The apparent national difference in narrative comprehension largely disappears once home background is accounted for, and the difference that does remain visible is far larger when children are asked to retell a story than when they are asked to answer questions about it.

The Adjusted Country Gap Was Largely Explained by Measured Background Factors

Nigerian children scored around half a standard deviation below UK children on comprehension, and the country difference was substantially reduced after statistical adjustment for these measured factors. The variables that absorbed it are

not national ones. They vary within each country as much as between them, which is precisely why the design stratified by locality and school type rather than treating a country as a homogeneous block.

This is consistent with the sociocultural position that narrative competence is assembled through participation in culturally organised activity rather than issued at birth (Rogoff, 2003; Vygotsky, 1978), and extends it to a comparison the literature has not made. It also converts a general warning into a demonstration. Karasik and Robinson (2022) argue that standardised assessments mask cultural variation and misinform policy; this result shows the mechanism, since a study reporting only the unadjusted comparison would have published a national deficit that the adjusted model does not support.

An alternative reading deserves weighing. The covariates could be absorbing the country effect because they proxy for it rather than cause it, in which case the adjustment removes what it was meant to isolate. Caregiver schooling and print exposure, however, predicted comprehension both within and between countries, reflecting the variation the design deliberately created. The reading cannot be dismissed, but it is the weaker of the two.

The claim is bounded accordingly. These results suggest that where narrative comprehension differs between the samples, the difference is better attributed to home background than to anything that could reasonably be called culture. They do not establish that no cultural difference exists, only that none was detectable once the background was controlled.

The Gap Doubled When Children Had to Talk

The same children, on the same two stories, in the same session, showed a country difference of roughly half a standard deviation on structured probes and more than eight-tenths on open retelling. The interaction was significant and substantial.

This bears directly on a debate that has been running since Michaels (1981) observed teachers cutting off children whose stories moved between linked episodes rather than developing a single topic. Michaels and Heath (1983) argued that such children were judged against one community's conventions for what a story should look like. Their evidence was ethnographic; this finding is consistent with it and supplies a quantitative form: when the demand shifts from selecting an answer to producing an account scored against a structural rubric, the gap grows by a third of a standard deviation. It aligns with Nicolopoulou's (2008) demonstration

that children reach narrative coherence by several routes, since a rubric recognising only the canonical plot line will underscore the others.

The most economical explanation is that the retell measures two things at once, comprehension and conformity to a school-genre display convention, while the probes measure mainly the first. A rival explanation is that it reflects expressive language rather than convention, with Nigerian children producing shorter accounts because most were tested in a language they use less fluently in formal settings. The present data cannot distinguish between the two, and Table 1 shows why: all UK children were tested in English, whereas fewer than a third of Nigerian children were. The explanations are not exclusive, and both point in the same direction for practice.

Overall, the results suggest that where a cross-cultural difference in narrative ability appears, a substantial part of it may be an artefact of how comprehension was elicited rather than a property of comprehension itself.

What the Invariance Testing Added

The measurement work produced an unanticipated finding. In the aggregated model, scalar invariance held by the criteria set in advance, which would ordinarily license comparison of means and close the question. Item-level screening then flagged two of thirty-two items, both evaluation items, each around 0.9 logit harder for Nigerian children with equal measured ability, with no other item approaching significance after correction.

The bias was real, localised, and invisible above the item level because parcelling diluted it until it no longer affected the fit indices. This complicates the standard invariance sequence rather than overturning it: a study following that sequence alone could pass its own equivalence check while carrying instrument bias into every subsequent comparison. Given how few cross-cultural comparisons report invariance testing at all (Boer et al., 2018), reporting it appears necessary but not sufficient.

Significance, Limitations and What Follows

The theoretical framework set out earlier proposed that narrative processing has three separable layers: shared machinery for tracking events, culturally tuned schemas for what counts as a well-formed story, and a task layer governing what a child judges it appropriate to say to an adult. The pattern here is what that framework predicts. Differences were modest and largely explained by where the measure made the least demand on display, and substantially larger where it made

the most. The framework survives its first test, which is not the same as being confirmed.

The design bounds those claims. A survey is cross-sectional, so the parallel age gradients reflect differences among children rather than developmental trajectories. The confound between test language and country leaves the display explanation inseparable from an explanation in expressive language. One Yoruba-speaking sample cannot represent Nigeria's linguistic range, nor can one English region in the United Kingdom.

The next two steps follow directly. The first is a within-country design in Nigeria comparing Yoruba-dominant and English-dominant children on the same instruments, which would separate test language from national context in a way this design cannot. The second is to score the retells a second time against a rubric derived from local judgements of what makes a good story, so that the same recordings are read through two conventions rather than one.

For practice, the implication is narrow. Clinicians and assessment developers who use story retelling to identify language difficulty are using the measure on which cross-national differences are largest and most open to inflation by convention. Where a retell is used with a child from a community other than the one in which the rubric was built, a probe-based measure should be administered alongside it, and any discrepancy should be treated as a question about the instrument before it is treated as a finding about the child.

Conclusion

Based on the findings, the study concludes that children in Nigeria and the United Kingdom demonstrated differences in narrative comprehension and retelling performance, but the observed country difference in structured comprehension was substantially reduced and became statistically non-significant after adjustment for caregiver schooling, print exposure and home narrative practice. This indicates that the initial difference between the two groups should not be interpreted as evidence of an independent national or cultural deficit in narrative comprehension, although the findings do not rule out cultural influences that could not be isolated in the present study. The study further concludes that the assessment format is important in understanding observed between-country differences. While the country gap on structured comprehension probes was approximately half a standard deviation, it increased to more than eight-tenths of a standard deviation on open story retelling, with the country-by-measure interaction being statistically significant. This suggests that open retelling may capture not only children's understanding of a story

but also their ability to organise and verbally express that understanding in ways that conform to the expectations embedded in the assessment rubric. However, because language proficiency was strongly confounded with country, with all UK children assessed in English compared with only 31.2% of Nigerian children, the study cannot determine whether the larger retelling difference reflects narrative conventions, expressive language proficiency, or a combination of both. The findings also show that comprehension performance increased with age across both countries, while the country difference did not systematically vary across the age groups; given the cross-sectional design, this pattern should be understood as age-related differences in performance rather than evidence of developmental change. Overall, the study concludes that observed cross-country differences in children's narrative performance are shaped by the interaction of assessment format and contextual factors, particularly home background and language, and that retelling performance should therefore not be interpreted in isolation as a direct indicator of underlying comprehension or language difficulty. A more cautious and culturally responsive assessment approach that combines open retelling with structured comprehension measures is consequently warranted.

References

Amir, D., & McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behaviour, 41*(5), 430–444.

Apicella, C., Norenzayan, A., & Henrich, J. (2020). Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behaviour, 41*(5), 319–329.

Barrett, H. C. (2020). Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences, 24*(8), 620–638.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Berman, R. A., & Slobin, D. I. (1994). Relating events in narrative: A crosslinguistic developmental study. Lawrence Erlbaum.

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. Trends in Cognitive Sciences, 26(12), 1153–1170.

Boer, D., Hanke, K., & He, J. (2018). On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. *Journal of Cross-Cultural Psychology,* 49(5), 713–734.

Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. SAGE.

Broesch, T., Crittenden, A. N., Beheim, B. A., Blackwell, A. D., Bunce, J. A., Colleran, H., Hassan, A., Kline, M. A., Kramer, K. L., Lew-Levy, S.,

Mattison, S. M., Pisor, A., Prall, S. P., Pretelli, I., Purzycki, B., Quinn, E. A., Ross, C., Scelza, B., Starkweather, K. (2020). Navigating cross-cultural research: Methodological and ethical considerations. Proceedings of the Royal Society B: Biological Sciences, 287(1935), 20201245.

Bruner, J. S. (1990). Acts of meaning. Harvard University Press.

Butterfuss, R., Kim, J., & Kendeou, P. (2020). Reading comprehension. In Oxford Research Encyclopedia of Education. Oxford University Press.

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. Child Development, 91(5), 1819–1835.

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language,* 48(4), 792–814.

Draper, C. E., Barnett, L. M., Cook, C. J., Cuartas, J. A., Howard, S. J., McCoy, D. C., Merkley, R., Molano, A., Maldonado-Carreño, C., 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), e2375.

Ghai, S. (2021). It's time to reimagine sample diversity and retire the WEIRD dichotomy. Nature Human Behaviour, 5(8), 971–972.

Heath, S. B. (1983). Ways with words: Language, life, and work in communities and classrooms. Cambridge University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioural and Brain Sciences,* 33(2–3), 61–83.

Karasik, L. B., & Robinson, S. R. (2022). Milestones or millstones: How standard assessments mask cultural variation and misinform policies aimed at early childhood development. *Policy Insights from the Behavioural and Brain Sciences,* 9(1), 57–64.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? First Language, 42(6), 703–735.

Kim, Y.-S. G. (2020a). Hierarchical and dynamic relations of language and cognitive skills to reading comprehension: Testing the direct and indirect effects model of reading (DIER). *Journal of Educational Psychology,* 112(4), 667–684.

Kim, Y.-S. G. (2020b). Toward integrative reading science: The direct and indirect effects model of reading. Journal of Learning Disabilities, 53(6), 469–491.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review,* 95(2), 163–182.

Kline, R. B. (2023). Principles and practice of structural equation modelling (5th ed.). Guilford Press.

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalised) linear mixed models: An open introduction and tutorial in R. *Behaviour Research Methods,* 53(6), 2528– 2543.

Lakens, D. (2022). Sample size justification. Collabra: Psychology, 8(1), 33267.

Lancy, D. F. (2022). The anthropology of childhood: Cherubs, chattel, changelings (3rd ed.). Cambridge University Press.

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, 3(1), 24–52.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253.

Melzi, G. (2000). Cultural variations in the construction of personal narratives: Central American and European American mothers' elicitation styles. Discourse Processes, 30(2), 153–177.

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language,* 112, 104092.

Michaels, S. (1981). "Sharing time": Children's narrative styles and differential access to literacy. *Language in Society,* 10(3), 423–442.

Minami, M., & McCabe, A. (1991). Haiku as a discourse regulation device: A stanza analysis of Japanese children's personal narratives. Language in Society, 20(4), 577–599.

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, 31(6), 678–701.

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A sociocultural developmental theory. *Psychological Review,* 111(2), 486–511.

Nicolopoulou, A. (2008). The elementary forms of narrative coherence in young children's storytelling. *Narrative Inquiry,* 18(2), 299–325.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review,* 108(2), 291–310.

Nketia, J., Amso, D., & Brito, N. H. (2021). Towards a more inclusive and equitable developmental cognitive neuroscience. Developmental Cognitive Neuroscience, 52, Article 101014.

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development,* 78(4), 1255–1264.

Rogoff, B. (2003). The cultural nature of human development. Oxford University Press.

Rowe, M. L., & Weisleder, A. (2020). Language development in context. *Annual Review of Developmental Psychology,* 2, 201–223.

Scheidecker, G., Chaudhary, N., Keller, H., Mezzenzana, F., & Lancy, D. F. (2023). "Poor brain development" in the global South? *Challenging the science of early childhood interventions. Ethos*, 51(1), 3–26.

Scheidecker, G., Chaudhary, N., Oppong, S., Röttger-Rössler, B., & Keller, H. (2022). Different is not deficient: Respecting diversity in early childhood development. T*he Lancet Child & Adolescent Health*, 6(12), e24–e25.

Serpell, R. (2011). Social responsibility as a dimension of intelligence, and as an educational goal: Insights from programmatic research in an African society. *Child Development Perspectives*, 5(2), 126–133.

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalised science of infant development. *Infancy,* 28(4), 708–737.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), New directions in discourse processing (pp. 53–120). Ablex.

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, 76(1), 116–129.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wang, Q. (2021a). The cultural foundation of human memory. Annual *Review of Psychology,* 72, 151–179.

Wang, Q. (2021b). Cultural pathways and outcomes of autobiographical memory development. *Child Development Perspectives,* 15(3), 196–202. https://doi.org/10.1111/cdep.12423

Yarkoni, T. (2022). The generalizability crisis. Behavioural and Brain Sciences, 45, e1.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin,* 123(2), 162–185.