



Representation and Absence: Race and Ethnicity in the World's Best-Selling ELT Coursebook

Derek Hopper¹

¹University of Doha for Science and Technology, Qatar

Corresponding Author Email: derekmhopper@gmail.com

DOI: <https://doi.org/10.36892/ijlls.v8i6.2806>

APA Citation: Hopper, D. (2026). Representation and Absence: Race and Ethnicity in the World's Best-Selling ELT Coursebook. *International Journal of Language and Literary Studies*, 8(6).197-212. <https://doi.org/10.36892/ijlls.v8i6.2806>

Received:

05/08/2026

Accepted:

01/10/2026

Keywords:

ELT Materials,
Coursebook
Analysis,
Racial and
Ethnic
Representation,
Headway,
English
Language
Learners

Abstract

Claims about racial and ethnic representation in ELT materials are frequently made without specifying the demographic standard against which representation is being judged. As a result, the same data may support different conclusions. This study applies three explicit benchmarks — United Kingdom publisher-country demographics, global population demographics, and global English language learner demographics — to the six Student Books of the Headway 5th edition, a series marketed as the world's best-selling adult English course. Every photographed human figure whose racial or ethnic identity was visually legible was counted and assigned to one of eight categories. Of 1,877 photographed figures, 1,823 were codeable and 54 were recorded as ambiguous and excluded from benchmark comparisons. A second coder independently recoded a systematic sample of 197 pages (20.2% of the corpus), identifying 414 figures against the primary coder's 415, with 98.3% category-overlap agreement. White figures account for 78.2% of the codeable corpus and Black figures for 9.0%. The three benchmarks yield substantially different readings of these results: against United Kingdom census data the White share sits marginally below the national proportion and the Black share at more than double it, while against global population and global learner benchmarks the White share is several times larger than any corresponding demographic proportion. Southeast Asian figures account for 1.2% of the codeable corpus, and no figure in the series was coded as Central Asian, despite Southeast and Central Asian populations forming a substantial and growing share of English language learners worldwide. The study argues that specifying demographic benchmarks is a precondition for interpretable claims about representation, and documents a pattern of visual absence that a Black/White framing of diversity does not capture.

1. Introduction

The Headway series is marketed as 'the world's best-selling adult English course' (Oxford University Press, n.d.-a), and few ELT series appear to rival its global reach and longevity. With over 70 million copies sold (Ozóg, 2018) across multiple editions spanning four decades, and its authors Liz and John Soars having become, in the words of OUP itself, 'synonymous with English language teaching and learning' (Oxford University Press, n.d.-b), Headway occupies a position in the ELT industry that has been described as 'a remarkable publishing phenomenon' (Gray, 2007, p. 137). It is a highly visible reference point against which other global coursebooks are often measured, and is likely to be encountered by a substantial proportion of the estimated two billion English language learners worldwide (British Council, 2013; Patel et al., 2023).

That canonical status makes *Headway* a significant and frequently studied object of research. Scholars have examined its cultural content (Melliti, 2013), its construction of social class (Gray & Block, 2014), its standard language ideologies (Gerday, 2019), and its representation of gender (Keles et al., 2023). Yet despite this substantial body of work, systematic quantitative analysis of racial and ethnic representation across the complete series, benchmarked against relevant demographic standards, is essentially absent from the literature. More specifically, no study has yet asked whether *Headway's* visual world reflects the composition of its global learner base, or examined where, within its non-White representation, the most significant absences lie.

This paper attempts to address that gap. Drawing on quantitative frameworks developed in Hopper (2026), it presents a racial and ethnic headcount of the *Headway* 5th edition Student Books across all six levels, benchmarking the results against multiple demographic standards, including the demographics of global English language learners. Its central argument is that the representation problem in *Headway* is not simply one of White overrepresentation, a finding that would be unsurprising and largely consistent with the existing literature. More specifically, it identifies an invisibility that the diversity conversation in ELT has largely overlooked: Southeast and Central Asian populations, which together account for a substantial and growing share of the global EFL learner population, are almost entirely absent from its pages.

Two research questions guide the analysis:

1. How are racial and ethnic groups distributed across the photographed figures in the *Headway* 5th edition Student Books, and how does that distribution compare to relevant demographic benchmarks?
2. Where, within the series' non-White representation, do the most significant absences lie?

2. Literature Review

2.1 Representation in ELT Materials

Critical examination of racial and ethnic representation in ELT materials has a substantial history, yet it has been characterised by a persistent methodological limitation: a reliance on qualitative critique at the expense of quantitative rigour. Researchers frequently invoke concepts such as 'underrepresentation' and 'diversity' without specifying the demographic benchmarks against which representation is being measured, rendering their claims difficult to verify, compare, or act upon (Taylor-Mendes, 2009; Hilliard, 2014; Durand, 2021). This problem extends across multiple dimensions of representation (gender, for example: Selvi & Kocaman, 2021; culture: Grant & Wong, 2018; sexuality: Gray, 2013) but is particularly acute in the domain of racial and ethnic identity, where the influence of critical race theory has created additional resistance to quantitative measurement (Sablan, 2019). This study applies an explicitly benchmarked quantitative approach of the kind such critiques have lacked.

2.2 Quantitative Approaches to Representation Research

The relatively small number of studies that have attempted systematic quantitative analysis of racial and ethnic representation in ELT materials illustrates both the value and the difficulty of such work. Roohani and Molana (2013) provided one of the most detailed numerical analyses in the literature, cataloguing racial representation across the *Interchange* series and finding that White individuals constituted approximately 79% of images. However, as Hopper (2026) noted, the study's conclusions were undermined by the absence of an explicit demographic benchmark, and the claim that Black individuals were underrepresented was inconsistent with both UK and US census data. Melliti's (2013) analysis of *Headway Intermediate* similarly provided descriptive data without a clear comparative framework, limiting the interpretive reach of its findings. (It is notable, however, that Melliti's learner

respondents themselves identified Asian cultures as among the most poorly represented in the coursebook, a perception the study was not equipped to quantify or benchmark.)

The methodological gap identified across this literature is not merely technical. Without explicit benchmarks, representation research risks circular critique: materials are judged against unstated standards that reflect the researcher's assumptions rather than defensible demographic realities. The three frameworks proposed in Hopper (2026) — publisher-country demographics, global demographics, and target market demographics — provide the analytical infrastructure this study employs. Of the three, the global English language learner demographics framework is the most directly relevant to a series marketed globally and used predominantly by non-native speakers of English: the question it poses — does the visual world of a globally marketed series reflect the composition of the global learner population? — is both the most pointed and the most answerable. All three benchmarks are applied in Section 4, but it is against this standard that the study's central findings are interpreted.

2.3 Asian Representation in Global ELT Materials

The existing literature on racial and ethnic representation in ELT materials has largely organised itself around a Black/White binary (Brittain, 2020; Hawthorne, 2020; Anya, 2024). This binary is often attributed to the Anglo-American publishing contexts from which global coursebooks predominantly originate, but that explanation does not withstand scrutiny: even within those contexts, Black populations are outnumbered by Hispanics in the United States and by people of South Asian descent in the United Kingdom. The salience of the Black/White binary is therefore better understood as a function of the critical race frameworks that have shaped this research tradition than as a reflection of any demographic reality, whether domestic or global. Asian learners, despite comprising the largest regional share of the global EFL population (Wei & Su, 2012; Patel et al., 2023; cf. Section 3.6), have received comparatively little sustained attention, and where Asian representation has been noted, it has typically been subsumed within undifferentiated non-White categories or treated as incidental to a broader critique of White dominance.

The consequences of this are practical as well as conceptual. East Asian learners (Chinese, Japanese, Korean) have at least been acknowledged as a major market whose preferences shape editorial decisions in global publishing (Harwood, 2014). Many other Asian populations, by contrast, have received almost no attention in the materials analysis literature, despite the scale of English language learning in countries such as India, Bangladesh, Vietnam, Indonesia, Thailand, and the post-Soviet Central Asian republics. Even studies explicitly concerned with racial representation have tended to foreground Black presence and absence while leaving the near-total invisibility of these populations unremarked (Taylor-Mendes, 2009; Dos Santos & Windle, 2021; Bryan et al., 2022).

This study addresses that gap directly. Rather than treating Asian representation as a monolithic category, it disaggregates visual representation across eight racial/ethnic categories: White/European, Black/African, East Asian, South Asian, Southeast Asian, Central Asian, Middle Eastern/Arab, and Latin American. This granularity is deliberate: the study's central argument is that the representation problem in *Headway* is not simply one of White dominance, but of a more specific invisibility operating within the non-White portion of the corpus. South, Southeast, and Central Asian figures, this study contends, are the true marginalised in the world's best-selling ELT coursebook, absent not because their populations are small, but because they fall outside the racial imaginaries of both the Anglo-American publishing tradition and the diversity conversation that has grown up in response to it.

3. Methodology

3.1 Corpus

The corpus comprises the six Student Books of the *Headway* 5th edition series (Soars et al., 2019a-f): Beginner, Elementary, Pre-Intermediate, Intermediate, Upper-Intermediate, and

Advanced. Student Books were selected as the primary site of analysis because they constitute the learner-facing component of the series and are the materials through which *Headway's* visual world is most directly encountered by those learning from it. Workbooks, teacher's guides, and digital supplementary components were excluded from analysis. The focus of this study is the representation of learners, and the Student Book is the most appropriate unit of analysis for that purpose.

3.2 Coding Unit

The coding unit was the individual photographed person. Every human figure appearing in a photograph whose racial or ethnic identity was legible to the coder was counted and assigned to one of the coding categories described in Section 3.3. Figures were counted each time they appeared; a person photographed twice on the same page was counted twice, as the concern is with visual presence across the corpus rather than character diversity. Figures whose faces were partially obscured or photographed largely from behind were included where racial identity remained visually legible. In dense scenes such as crowds or busy street views, only prominent or foregrounded figures were counted, even where the racial identity of other individuals was legible; this prevents incidental background figures in a small number of large-group images from distorting the aggregate counts. Illustrated figures (cartoons and other non-photographic representations of humans) were excluded from analysis. Although illustrated figures appear in the series, their number is comparatively small, and the stylised conventions of illustration make consistent racial and ethnic coding considerably less reliable than is the case for photographs of real people. The proportion of illustrated to photographic content varies across levels, a point returned to in Section 4. Figures whose racial or ethnic identity could not be determined with reasonable confidence were assigned to a separate ambiguous category and excluded from benchmark comparisons.

3.3 Coding Scheme

Each legible figure was assigned to one of eight mutually exclusive racial and ethnic categories. Coding combined visual assessment with contextual cues, including names, settings, and accompanying text, particularly where visual information alone was insufficient. The categories were:

- *White/European*: figures visually coded as of European descent, including those appearing in Western or internationally non-specific settings without other contextual cues suggesting a different regional identity.
- *Black/African*: figures visually coded as of sub-Saharan African descent, including those of African diasporic backgrounds where contextually legible.
- *East Asian*: figures visually coded as of Chinese, Japanese, or Korean descent.
- *South Asian*: figures visually coded as of Indian, Pakistani, Bangladeshi, or Sri Lankan descent.
- *Southeast Asian*: figures visually coded as of Vietnamese, Thai, Indonesian, Filipino, Malaysian, or other Southeast Asian descent. Where visual coding alone was insufficient to distinguish Southeast Asian from East Asian figures, contextual cues were used; figures remaining ambiguous after contextual analysis were assigned to the ambiguous category.
- *Central Asian*: figures visually coded as of Kazakh, Uzbek, Kyrgyz, Tajik, Turkmen, or other Central Asian descent. This was the most difficult category to code, as Central Asian populations exhibit phenotypic overlap with European and East Asian groups; contextual cues were therefore relied upon heavily, and figures that could not be confidently distinguished were assigned to the ambiguous category rather than coded as Central Asian. This conservative rule means the Central Asian count should be understood as a lower bound, though, as Section 4 reports, the category is empty across the entire corpus.

- *Middle Eastern/Arab*: figures visually coded as of Arab, Persian, or broader Middle Eastern descent.
- *Latin American*: figures visually coded as of Latin American descent.

Figures not assignable to any of the above categories with reasonable confidence were recorded as ambiguous/cannot determine and reported separately, and were excluded from benchmark comparisons. The ambiguous category was also used for the small number of figures whose regional identity fell entirely outside the scheme's eight categories. For example, one book contained two photographs of a man identifiable as Melanesian (from Papua New Guinea), but since Melanesia does not correspond to any major constituency of the global EFL learner population and the scheme's categories are defined by reference to that population, such figures were recorded as ambiguous/cannot determine rather than forced into an ill-fitting category. Three terms are used consistently throughout. *Photographed figures* denotes all counted human figures in the corpus ($n = 1,877$); *codeable figures* denotes those assigned to one of the eight racial and ethnic categories ($n = 1,823$); and *ambiguous figures* denotes those that could not be so assigned ($n = 54$). All percentages reported in Section 4, and all benchmark comparisons, are calculated on the codeable total unless otherwise stated.

3.4 Coding Procedure

Books were analysed sequentially from Beginner to Advanced. The coder proceeded page by page through each Student Book, maintaining a running tally in an Excel spreadsheet with one column per coding category. At the end of each page, counts for that page were added to the running total for each category. This procedure was repeated for all six books, producing both per-book totals and a series aggregate.

3.5 Intercoder Reliability

To assess the reliability of the coding scheme, a second coder independently recoded a systematic sample of the corpus: every fifth page of each Student Book, from page 5 to page 160 in Books 1–4, to page 170 in Book 5, and to page 175 in Book 6. This yielded 197 sampled pages, 20.2% of the 976-page corpus. Pages containing no photographs were sampled and recorded as containing no codeable figures rather than omitted, so that the sample reflects the systematic interval rather than only image-bearing pages.

An initial reliability attempt with a different coder was discontinued. The divergence in that attempt indicated that the written scheme, applied without demonstration, did not convey the figure-inclusion criteria reliably, particularly for dense scenes. For the repeated exercise the primary coder therefore worked through five non-sampled pages with the second coder to demonstrate the application of the inclusion criteria. The second coder received the written instructions and category definitions and then coded the sampled pages independently, without sight of the primary coder's counts.

Because the procedure records category counts per page rather than classifications of individually identified figures, agreement is reported as percentage agreement at the page and category level rather than as a chance-corrected coefficient. Cohen's kappa was not computed: the unit of analysis in dense multi-person images cannot be independently and reproducibly individuated, so the figure-level paired classification that kappa requires is not available for this corpus. This is a recognised constraint of count-based visual content analysis rather than a remediable feature of the present design, and it is returned to in Section 5.6.

Agreement was high. The two coders independently identified almost identical totals across the sample: 415 codeable figures (primary coder) and 414 (second coder). On 174 of 197 sampled pages (88.3%), all nine category counts were identical between coders. Per-category page-level agreement was 99% or higher for every category except White (92.9%) and Ambiguous (94.4%). The categories on which the study's central findings rest showed near-complete agreement: Black (29 versus 29), Southeast Asian (15 versus 13), and Central Asian (0 versus 1). Overall category-overlap agreement across the sample was 98.3%.

The disagreement that did occur was concentrated, as anticipated in the coding scheme, in the White and Ambiguous categories within dense-population images such as crowd scenes and group photographs, where the determination of which background figures meet the inclusion threshold is least certain. This divergence bears on the White and Ambiguous aggregates but not on the study's central findings regarding the near-absence of Southeast and Central Asian figures, where independent agreement was effectively complete. The full coding spreadsheets for both coders are openly available (<https://doi.org/10.5281/zenodo.20270921>), permitting independent verification; blank cells denote zero.

3.6 Benchmark Selection

Before turning to the benchmarks themselves, it is worth stating a premise this study takes as given: racial and ethnic categories do not correspond to discrete, non-overlapping biological populations. Human variation is substantially clinal, and any classification scheme — whether the eight-category coding scheme used here or the demographic benchmarks against which it is compared — is necessarily an approximation of a continuous reality (American Anthropological Association, 1998; Templeton, 1998). This is not a limitation unique to the present study; it applies with equal force to any claim about racial or ethnic representation, including the impressionistic claims of over- or under-representation common in the literature reviewed in Section 2. A study that specifies its benchmarks explicitly and acknowledges their approximate character is not thereby less rigorous than one that offers no benchmark at all; it is more rigorous, because its imprecision is visible and its reasoning can be checked, where the imprecision of an unbenchmarked claim is simply invisible. The methodological question this study treats as central is therefore not whether a benchmark is exact (no benchmark for a clinally distributed variable can be), but whether it is transparent, defensible, and adequate to support the claim being made of it.

Results are interpreted against three demographic benchmarks, each corresponding to one of the quantitative frameworks proposed in Hopper (2026). The first is publisher-country demographics, using the United Kingdom's 2021 census data, on the grounds that *Headway* is an OUP product developed in Britain and may therefore be read as a cultural artefact of its publishing context. The second is global population demographics. The third, and the one most directly relevant to this study's central argument, is global EFL learner demographics.

The second benchmark presents a methodological difficulty that must be stated plainly. No single authoritative dataset of the world's population by race or ethnicity exists, and none can readily be constructed: national data instruments are mutually incommensurable, with some states recording race, others ethnicity or ancestry, and many collecting no such data at all, so global figures cannot be obtained by aggregating national censuses. This benchmark is therefore operationalised as a regional proxy. Using United Nations World Population Prospects 2024 regional population estimates (United Nations, 2024), the world's major regions are mapped onto the study's categories: sub-Saharan Africa as a proxy for Black/African, Latin America and the Caribbean for Latin American, the European population for White/European, and the principal Asian sub-regions for the four Asian categories, with Northern Africa and Western Asia approximating Middle Eastern/Arab. The South Asian and Central Asian figures were derived separately from constituent country data within the WPP 'Central and Southern Asia' region, which the dataset does not disaggregate at the headline level. This mapping is deliberately coarse and is presented as such. It disregards diaspora populations, internal heterogeneity, and intermarriage, and should be read as an order-of-magnitude reference point rather than a precise demographic standard. It is included not because it can be made exact, which it cannot, but because even a coarse global benchmark makes visible the scale of the divergence between *Headway's* visual world and the world's population, a divergence too large to be an artefact of the approximation.

Precise figures for the third benchmark are not available from a single authoritative source, a limitation acknowledged explicitly here. However, the broad contours are well established

across the most credible available sources. The British Council estimates that around two billion people were using or learning English by 2020, rising to some 2.3 billion by the early 2020s (British Council, 2013; Patel et al., 2023). This learner population is overwhelmingly Asian. A national language survey in China, the largest of its kind conducted there, established that approximately 390 million people in mainland China had learned English (Wei & Su, 2012), a figure the British Council now places at around 400 million. The British Council's own regional consultations treat South Asia, East Asia, and the ASEAN countries (Indonesia, the Philippines, Thailand, Vietnam, and others) as among the largest and fastest-growing English-learning constituencies in the world (Patel et al., 2023), and earlier British Council forecasting singled out Indonesia, alongside other large non-Western states, for double-digit projected growth in demand for English (British Council, 2013).

This benchmark is therefore treated as directional rather than precise. The argument it supports does not depend on exact figures, which no source can currently supply, but on a claim that is clear, defensible, and consistent across every credible source from 2012 to the present: that Asia, and South and Southeast Asia within it, constitutes the plurality of the global EFL learner population, and that this has been true and growing throughout the period in which the *Headway* 5th edition has been in circulation. One further qualification should be stated explicitly, as it constrains every comparison drawn against this benchmark. The benchmark describes the global population of English language learners; it does not describe the users of *Headway*. Oxford University Press does not publish sales or user demographics for the series, and no independent dataset of its readership exists. What can be established from public sources is that the series is marketed and sold internationally, including through region-specific editions for markets outside the United Kingdom (Oxford University Press, n.d.-a). The comparison drawn in Section 4.2 is therefore contextual rather than direct: it asks how the series' visual composition stands in relation to the world's learner population, not how it stands in relation to its own purchasers.

4. Results

4.1 Racial and Ethnic Distribution Across the Series

Table 1 presents the distribution of coded figures by racial and ethnic category across all six *Headway* 5th edition Student Books, together with series totals and percentages.

Table 1: Racial and Ethnic Distribution of Photographed Figures in Headway 5th Edition Student Books

Category	Beginner	Elementary	Pre-Int	Intermediate	Upper-Int	Advanced	Total	%
White	323	311	226	214	201	150	1,425	78.2
Black	36	41	35	18	24	10	164	9.0
East Asian	27	4	6	9	12	9	67	3.7
South Asian	11	12	3	48	29	2	105	5.8
Southeast Asian	0	4	1	13	0	4	22	1.2
Central Asian	0	0	0	0	0	0	0	0.0
Middle East/Arab	15	3	2	2	3	0	25	1.4
Latin American	6	7	1	1	0	0	15	0.8
Codeable total	418	382	274	305	269	175	1,823	
Ambiguous	12	15	5	4	15	3	54	
Total	430	397	279	309	284	178	1,877	

White figures dominate the corpus at every level, ranging from 75.1% at Beginner to 85.7% at Advanced, with a series aggregate of 78.2%. Black figures constitute the second largest group at 9.0%, followed by South Asian at 5.8% and East Asian at 3.7%. Southeast Asian figures account for 1.2% of the codeable corpus, appearing in only four of the six books and never exceeding 13 figures at any single level. Central Asian figures are entirely absent from the series. The total number of photographed figures declines across levels, consistent with the

increasing proportion of text-based content in higher-level materials, but the representational pattern remains broadly consistent throughout.

Two qualitative observations warrant noting alongside the quantitative data. First, the South Asian figures at Intermediate level are substantially inflated by a single family album montage featuring an Indian family, which accounts for a disproportionate share of that level's South Asian count. Second, the majority of Southeast Asian figures at Intermediate level appear in photographs depicting a White woman conducting charity work in Vietnam, raising questions about the representational framing of those figures that the quantitative data alone does not capture. Both observations bear on the interpretation of these results and are discussed further in Section 5.

4.2 Comparison Against Demographic Benchmarks

Table 2 presents the series aggregate figures alongside the three demographic benchmarks described in Section 3.6.

Table 2: Headway 5th edition series aggregate against three demographic benchmarks

Category	Headway (%)	UK census (%)	World population (WPP 2024, approx.)	Global EFL learner base
White/European	78.2	81.7	≈14	Minority and declining share
Black/African	9.0	4.0	≈15	Small but growing share
East Asian	3.7	≈0.7 (census "Chinese")	≈20	Largest single national bloc (China dominant)
South Asian	5.8	6.9	≈24	Very large; English has official status across the region
Southeast Asian	1.2	Not separately enumerated	≈8–9	Very large and rapidly growing market
Central Asian	0.0	Not separately enumerated	≈1	Small but rapidly emerging
Middle Eastern/Arab	1.4	≈0.6 (census "Arab")	≈5–6	Significant and stable
Latin American	0.8	Not separately enumerated	≈8	Significant; large national programmes in Brazil and Mexico

The three benchmarks yield substantially different interpretations of *Headway's* representational profile. Against UK publisher-country demographics, White figures appear at a rate marginally below their share of the population, while Black figures appear at more than twice their demographic proportion. South Asian figures, despite comprising the largest non-White ethnic group in the UK, appear at a rate slightly below their demographic weight. The remaining categories expose a limit of this benchmark that is itself revealing: Southeast Asian, Central Asian, and Latin American populations are not separately enumerated in the UK census at all, falling within residual classifications that together amount to roughly 3.2% of the population. The invisibility this study documents in the coursebook is, for these groups, continuous with an invisibility in the demographic instruments of the publishing country itself.

Against global population demographics, the picture shifts. White/European figures occupy 78.2% of the corpus against a regional-proxy world share in the mid-teens, while South and East Asian figures, together corresponding to roughly 44% of the world's population, account for less than 10% of *Headway's* photographed corpus. As Section 3.6 notes, this benchmark is a coarse regional proxy; but a divergence of this magnitude could not be produced by any plausible refinement of the approximation.

It is against the third benchmark that the most consequential gap emerges. The global EFL learner population is, on every credible estimate, overwhelmingly Asian: mainland China alone accounts for approximately 390 to 400 million English learners (Wei & Su, 2012; Patel et al., 2023), and South Asia and the ASEAN countries are consistently identified as among the

largest and fastest-growing English-learning constituencies in the world (Patel et al., 2023). Against this benchmark, Southeast Asian figures at 1.2% and the absence of any Central Asian figure represent a shortfall in visual representation of a different order from that indicated by the other two standards: the populations accounting for the largest share of English language learners worldwide are among the least visible in the series. As set out in Section 3.6, this is a contextual comparison with the global learner population, not a comparison with the readership of the series itself.

5. Discussion

5.1 The Headline Finding

The data presented in Section 4 admit of a straightforward summary: the world's best-selling ELT coursebook series is, visually, an overwhelmingly White world. Across 1,823 codeable figures in six Student Books, 78.2% are White and 9% are Black. The remaining 12.8% is distributed across five further categories, two of which, Southeast Asian and Central Asian, are effectively absent. This finding is, in one sense, unsurprising; White dominance in globally circulating ELT materials has been documented for decades. What the data reveal that is genuinely new is the specific texture of the non-White representation that does exist, and what it implies about whose presence the series imagines as worth representing.

5.2 What the Benchmarks Show

The three benchmarks applied in this study yield substantially different readings of *Headway's* representational profile, and that divergence is itself analytically significant. This demonstrates empirically the point argued in Hopper (2026): that claims about representation are meaningless without explicit reference to a demographic standard.

Against UK publisher-country demographics, *Headway's* White representation is in fact marginally below the national proportion of 81.7%, while Black representation at 9% is more than double the UK's 4% (Gov.uk, 2024). This is not a trivial finding. It means that anyone applying the standard diversity critique (i.e., that globally circulating ELT materials overrepresent White people) would need to specify what they mean, because against the most obvious publishing-context benchmark, White figures are if anything slightly underrepresented. South Asian figures, comprising the largest non-White ethnic group in Britain at 6.9% of the population, appear at 5.8% in the series. This is close to demographic parity, but it is notable given that nearly half of those figures derive from a single image sequence in one book.

Against the global population benchmark the picture is entirely different. On the regional-proxy mapping described in Section 3.6, White/European figures occupy 78.2% of *Headway's* visual space against a European share of the world's population in the mid-teens, while South and East Asian figures, together corresponding to roughly 44% of the world's population, account for less than 10% of the corpus. Southeast Asians, on the same mapping close to a tenth of the world's population, appear at 1.2%; Central Asians do not appear at all.

It is against the third benchmark, however, that the most consequential gap emerges. The global EFL learner population, on every credible estimate from 2012 to the present, is overwhelmingly Asian: mainland China alone accounts for approximately 390 to 400 million English learners (Wei & Su, 2012; Patel et al., 2023), and South Asia and the ASEAN countries are consistently identified as among the largest and fastest-growing English-learning constituencies in the world (Patel et al., 2023). Vietnam, Indonesia, and Thailand together represent hundreds of millions more. Against this benchmark, Southeast Asian figures at 1.2% and Central Asian figures at 0% represent a shortfall far larger than that registered by either of the other two standards. The qualification entered in Section 3.6 applies throughout: what the divergence documents is the distance between the series' visual composition and the global learner population, not the experience of any identified group of its readers.

5.3 The Function of Black Representation

Black figures appear at a relatively consistent rate across all six levels, ranging from 5.7% at Advanced to 12.8% at Pre-Intermediate, and at every level they exceed the 4% Black share of the UK population while sitting at or below the corresponding global-population proxy. This benchmark-relative pattern warrants examination. It is tempting to read this as evidence of genuine diversity awareness on the part of the series' authors and editors, and it may partly reflect that. But the pattern is difficult to explain on pedagogical or demographic grounds alone.

Black people are not, in global terms, a major EFL learner population. Sub-Saharan Africa represents a relatively small share of global EFL enrollment compared to Asia, and English in many African contexts functions as an official or second language rather than a foreign one. The rate of Black representation in *Headway* does not, therefore, reflect the demographics of who learns from the book. Nor does it reflect the demographics of where the book is published: Black Britons comprise 4% of the UK population, yet appear at more than twice that rate in the series.

The more plausible explanation is that Black representation in *Headway* is performing a different function: it is responding to diversity pressures emanating from the Anglo-American publishing and educational context in which the series is produced and marketed. The critical race framework that has shaped diversity discourse in British and American education, and which, as Hopper (2026) noted, has also shaped the methodological assumptions of ELT representation research, foregrounds Black and White as the primary categories of racial analysis. *Headway's* representational choices appear to reflect this framework. The present data establish this correspondence as a pattern; they do not, on their own, establish the editorial mechanism behind it. Demonstrating that directly would require the kind of evidence this study does not have, such as publisher guidelines or editorial interviews of the sort Gray (2010) obtained. What can be said is that the pattern is strongly inconsistent with pedagogical or demographic explanation, and that alignment with Anglo-American diversity discourse is the most parsimonious account that remains. The series looks diverse by the standards of the conversation happening in London and New York. It does not look diverse by the standards of the conversation that would need to happen in Hanoi, Jakarta, or Tashkent.

This asymmetry is consistent with what applied linguists have termed native-speakerism: the ideology within ELT that treats the native speaker — in practice, and despite the term's ostensibly linguistic definition, an implicitly White, Western figure — as the benchmark of legitimate English (Holliday, 2006). Native-speakerism offers a ready explanation for the specific shape of *Headway's* non-White representation: Whiteness is positioned as the default face of linguistic authority, while the populations who most numerically constitute the language's actual global user base are either absent or, as Section 5.4 discusses, present only in roles that do not carry that same authority.

This is not an argument against the inclusion of Black figures in ELT materials. It is an argument that the diversity logic driving those inclusions is parochial: calibrated to one racial context while remaining blind to another, larger one.

5.4 The Invisibility of Southeast and Central Asia

The near-total absence of South, Southeast and Central Asian figures from *Headway's* visual corpus is this study's central finding, and it requires examination at two levels: what the numbers show, and what the occasional appearances reveal.

At the numerical level, the finding is stark. Twenty-two Southeast Asian figures across six books and 1,823 codeable figures represents a presence so marginal as to be functionally invisible. Central Asian figures are entirely absent: not a single figure across the entire series is identifiable as Kazakh, Uzbek, Kyrgyz, Tajik, or Turkmen. This is not the result of ambiguous coding: the ambiguous category, which might in theory contain misclassified Southeast or Central Asian figures, totals 54 across the series. Even if every ambiguous figure were reassigned to these two categories, the finding would remain essentially unchanged.

What is perhaps more revealing than the zeros is the nature of the appearances that do occur. The majority of Southeast Asian figures at Intermediate level (the book where Southeast Asian presence peaks at 13 figures) appear in photographs depicting a White woman conducting charitable work in Vietnam. These figures are not learners, professionals, interlocutors, or autonomous social actors. They are recipients, positioned within a visual logic of benevolence and exoticism that reproduces rather than challenges the marginalisation of Southeast Asian people in globally circulating media. Presence of this kind may be worse than absence in one respect: it actively constructs a representation of Southeast Asians while foreclosing the possibility of a different one.

This visual logic of benevolence and exoticism is not unique to *Headway*; it is consistent with what critical commentary on international volunteering, aid, and charity photography has termed the White-Savior Industrial Complex (Cole, 2012), in which non-White populations are rendered visible primarily as objects of Western care rather than as autonomous social actors, learners, or interlocutors in their own right.

The South Asian figures at Intermediate level present a related problem. A substantial proportion of that level's South Asian count derives from a single family album montage featuring an Indian family. This is not so much systematic representation as an editorial accident that inflates the aggregate figures without reflecting any sustained authorial commitment to South Asian visibility. Stripped of that montage, South Asian representation across the series would be considerably thinner than the 5.8% aggregate suggests.

Central Asian absence is in some ways the most pointed finding of all, precisely because it is total. The Central Asian republics represent a growing EFL market as the region opens to global trade and reduces its dependence on Russian as a lingua franca. Their complete non-existence in *Headway*'s visual world is not a consequence of small population size; Central Asians comprise approximately 1.2% of the global population, comparable to several groups that do appear in the series. It reflects, rather, a failure of imagination: these are populations that simply do not register within the visual grammar of ELT publishing.

A possible objection is that this near-absence reflects not editorial choice but the commercial image libraries available to ELT publishers: if stock photography is itself skewed away from Southeast and Central Asian subjects, the pattern would be an upstream supply constraint rather than a representational logic internal to the series. This does not weaken the finding. The consequence for any learner drawn from these populations who does use the series is identical regardless of mechanism: the visual world presented to them contains almost no one of their own background, whether that absence originates with the editor or with the catalogue. Indeed, a supply constraint of this kind strengthens the case for the publisher-country benchmark specifically: a British publisher cannot reasonably be held to a global-learner standard for populations its available image stock does not contain, but it can be held to the demographics of the society in which it operates and publishes. Establishing where in the production chain the pattern originates, whether through analysis of the image libraries or interviews with publishers of the kind Gray (2010) conducted, is a worthwhile direction for future research, not a condition for taking the documented pattern seriously.

5.5 The Commercial Paradox

There is a commercial dimension to these findings that deserves brief acknowledgment. Oxford University Press markets and sells *Headway* extensively across Southeast Asia, the Gulf region, and increasingly Central Asia. The series' status as the world's best-selling ELT course is built substantially on sales in precisely the markets whose populations are most invisible in its pages. There is something worth remarking in the fact that OUP describes *Headway* as "the world's best-selling adult English course" while producing a series in which so many of the world's English learners are effectively unrepresented. Whether this paradox reflects deliberate editorial choices, the inertia of established publishing conventions, or simple

inattention to the demographics of global EFL is beyond the scope of this study to determine. What the data make clear is that it exists.

This pattern is consistent with what Phillipson (1992) termed linguistic imperialism: a structural arrangement in which the institutions and materials that define and export 'legitimate' English originate from, and disproportionately represent, a linguistic and cultural centre, while the periphery supplies the market — and, increasingly, the revenue — without commensurate representation in the materials it purchases.

5.6 Limitations

Several limitations of this study warrant acknowledgment. First, the coding of racial and ethnic identity from photographs is an inherently interpretive and subjective act, and the Southeast Asian/East Asian distinction in particular presents coding challenges that contextual cues do not always resolve. The categories employed are, in some respects, socially constructed, and they are applied here based on visual and contextual inference rather than self-identification. Second, intercoder reliability was assessed on a systematic sample of the corpus (every fifth page, 20.2% of all pages) rather than its entirety, and is reported as percentage agreement at the page and category level rather than as a chance-corrected coefficient; as set out in Section 3.5, figure-level Cohen's kappa was not computed because dense multi-person images do not yield a stable, independently reproducible per-figure unit, a recognised constraint of count-based visual content analysis. Agreement on the categories central to the study's findings was near-complete, with the modest divergence concentrated in the White and Ambiguous counts in dense-population images; this bears on those aggregates but not on the central findings regarding Southeast and Central Asian near-absence. Third, the global EFL learner demographics benchmark, while directionally robust, lacks the precision of the UK census data; the argument it supports depends on approximate rather than exact figures. Relatedly, and more consequentially, no sales or user demographic data for the series are publicly available from the publisher, and this study did not obtain them. The learner benchmark therefore supports a contextual comparison with the global population of English language learners rather than a direct comparison with the people who use *Headway*, and the findings reported here should be read as describing the series' relation to the former rather than the latter. Fourth, the global population benchmark is a coarse regional proxy rather than a true measure of the world's population by race or ethnicity, for the reasons set out in Section 3.6. Fifth, this study is limited to the visual dimension of representation and does not examine how racially and ethnically coded figures are positioned within the series' narratives, dialogues, and exercises, a dimension future research might productively explore. Finally, the qualitative observations about the charity photographs and family montage, while analytically significant, are based on individual instances rather than systematic content analysis of representational framing across the corpus.

6. Conclusion

This study set out to apply the quantitative frameworks proposed in Hopper (2026) to a concrete empirical case: the racial and ethnic representation of photographed figures across all six Student Books of the *Headway* 5th edition series. The findings are clear. White figures dominate the corpus at every level, accounting for 78.2% of all codeable figures. Black figures appear at a level that is well above their share of the UK population but at or below their share of the global population, and far above their share of the global EFL learner base, so that Anglo-American diversity norms are the benchmark with which their representation is most nearly consistent. Southeast Asian figures, drawn from populations that constitute a major share of the global EFL learner base, appear at 1.2%. Central Asian figures do not appear at all.

The study's central contribution is not the documentation of White dominance: that finding, while confirmed here, is consistent with the existing literature and would surprise few readers. The contribution is the disaggregation of the non-White remainder. By applying a coding scheme fine-grained enough to distinguish Southeast Asian from East Asian figures, and by

benchmarking results against global EFL learner demographics rather than the Anglo-American racial categories that have dominated this research tradition, the study documents a pattern of visual absence that the existing literature has not examined: several of the populations most numerically significant to global English language learning are almost entirely unrepresented in one of the most widely sold series used to teach it. The series is not only predominantly White. Its non-White representation is patterned in a way that corresponds more closely to the diversity categories salient in its publishing context than to the demographic composition of English language learning worldwide.

The implications of this finding are relevant to three audiences. For publishers, the data show that the series' non-White representation is distributed along the lines salient in its domestic publishing context, and that several large learner populations are minimally or not at all represented. Increased Black representation registers as an improvement against the publisher-country benchmark but not against the global learner benchmark. What follows editorially from that is a judgement for publishers to make; the data establish the pattern, not its consequences. For researchers, the study demonstrates the value of the quantitative frameworks proposed in Hopper (2026): without explicit demographic benchmarks, the overrepresentation of Black figures (at least in terms of UK population share and learner demographics) would likely have been read as a diversity success rather than a symptom of the same parochialism that produces Southeast Asian absence. For educators and institutions in Southeast and Central Asia, the findings offer one piece of evidence relevant to materials selection: how far a given series visually represents the populations it is being used with. This study documents that representation; it does not test how learners respond to it, and nothing here should be taken to imply that altering representation would in itself affect language learning outcomes.

Future research might extend this analysis in several directions. A parallel study of other major global ELT series — Cambridge's *Interchange* or Pearson's *Cutting Edge*, for instance — would establish whether *Headway*'s representational profile is idiosyncratic or symptomatic of the industry as a whole. Qualitative analysis of how racially and ethnically coded figures are positioned within *Headway*'s narratives, dialogues, and exercises would complement the quantitative headcount undertaken here, addressing the question of not just who appears but in what roles and with what agency. And longitudinal analysis across *Headway*'s five editions would shed light on whether the series has become more or less representative of its global learner base over time, a question with obvious implications for how publishers respond to diversity critique.

What this study cannot claim is that representation of the kind documented here directly harms learners, or that its absence directly impedes language acquisition. That empirical question remains largely unresolved in the literature, as Gray (personal communication, October 14, 2023) and Hopper (2026) have noted. What it can claim is that one of the world's most widely sold ELT coursebooks presents a visual world from which several of the largest English-learning populations in the world are almost entirely absent, and that this is worth knowing, measuring, and taking seriously.

7. References

- American Anthropological Association. (1998). AAA statement on race. *American Anthropologist*, 100(3), 712. <https://doi.org/10.1525/aa.1998.100.3.712>
- Anya, U. (2024). Examining materials and instruction in the practice of critical race pedagogy for world language teaching. In *Race, racism, and antiracism in language education* (pp. 101-124). Routledge. <https://doi.org/10.4324/9781003283492>
- British Council. (2013). *The English effect: The impact of English, what it's worth to the UK and why it matters to the world*. <https://www.britishcouncil.org/sites/default/files/english-effect-report-v2.pdf>

- Brittain, E. (2020). Reinforcement of White native-speakerism: An analysis of English language teacher training materials. *GATESOL Journal*, 30(1), 16-31. <https://doi.org/10.52242/gatesol.105>
- Bryan, K., Romney-Schaab, M., & Cooper, A. (2022). The illusion of inclusion: Blackness in ELT. *The CATESOL Journal*, 33(1). <https://doi.org/10.5070/B5.35899>
- Cole, T. (2012, March 21). The white-savior industrial complex. *The Atlantic*. <https://www.theatlantic.com/international/archive/2012/03/the-white-savior-industrial-complex/254843/>
- Dos Santos, G. N., & Windle, J. (2021). The nexus of race and class in ELT: From interaction orders to orders of being. *Applied Linguistics*, 42(3), 473-491. <https://doi.org/10.1093/applin/amaa031>
- Durand, S. (2021). Rethinking under-represented ethnicities in ELT materials through a critical discourse analysis-oriented model. *MEXTESOL Journal*, 45(1), n1.
- Gerday, L. (2019). What is 'global' in an ELT global coursebook? A linguistic analysis of the New Headway Upper-Intermediate Student's Book. *English Text Construction*, 12(1), 59-83. <https://doi.org/10.1075/etc.00018.ger>
- Gov.uk. (2024). *Population of England and Wales*. Ethnicity facts and figures. <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest/>
- Grant, R., & Wong, S. (2018). Addressing cultural bias in ELT materials. *The TESOL encyclopedia of English language teaching*, 1–8.
- Gray, J. (2007). *A study of cultural content in the British ELT global coursebook: A cultural studies approach* (Doctoral dissertation, Institute of Education, University of London).
- Gray, J. (2010). *The construction of English: Culture, consumerism and promotion in the ELT global coursebook*. Palgrave Macmillan.
- Gray, J. (2013). LGBT invisibility and heteronormativity in ELT materials. In *Critical perspectives on language teaching materials* (pp. 40-63). London: Palgrave Macmillan UK.
- Gray, J., & Block, D. (2014). All middle class now? Evolving representations of the working class in the neoliberal era: The case of ELT textbooks. In *English language teaching textbooks: Content, consumption, production* (pp. 45-71). London: Palgrave Macmillan UK.
- Harwood, N. (Ed.). (2014). *English language teaching textbooks: Content, consumption, production*. Springer.
- Hawthorne, A. J. (2020). *Black in the British: Analysing the visual representation of Black people in British Council teaching materials* (Doctoral dissertation, Masters Dissertation. Lancaster University).
- Hilliard, A. D. (2014). A critical examination of representation and culture in four English language textbooks. *Language Education in Asia*, 5(2), 238-252. <http://dx.doi.org/10.5746/LEiA/14/V5/I2/A06/Hilliard>
- Holliday, A. (2006). Native-speakerism. *ELT Journal*, 60(4), 385–387. <https://doi.org/10.1093/elt/ccl030>
- Hopper, D. (2026). The Case for Quantitative Frameworks in Assessing Racial and Ethnic Representation in ELT Materials, *TESOL Journal*, 17, no. 3: e70142. <https://doi.org/10.1002/tesj.70142>.

- Keles, U., Mansouri, B., & Yazan, B. (2023). Perpetuation of gender bias in New Headway: A multimodal critical discourse analysis with a diachronic lens. *Pedagogy, Culture & Society*, 31(3), 379-401. <https://doi.org/10.1080/14681366.2021.1913442>
- Melliti, M. (2013). Global content in global coursebooks: The way issues of inappropriacy, inclusivity, and connectedness are treated in Headway Intermediate. *SAGE Open*, 3(4). <https://doi.org/10.1177/2158244013507265>
- Oxford University Press. (n.d.-a). *New Headway Plus*. https://elt.oup.com/catalogue/items/local/ae/new_headway_plus/?cc=global&selLanguage=en
- Oxford University Press. (n.d.-b). *Headway*. https://elt.oup.com/catalogue/items/global/adult_courses/headway/?cc=qa&selLanguage=en
- Ozóg, C. (2018). An Interview with Liz and John Soars. Meet the Authors. *IH Journal of Education and Development*, 44.
- Patel, M., Solly, M. & Copeland, S. (2023). *The future of English: Global perspectives*. British Council. <https://www.britishcouncil.org/future-of-english>
- Phillipson, R. (1992). Linguistic imperialism. Oxford University Press.
- Roohani, A., & Molana, E. (2013). An Investigation into Cultural Representation in Interchange Textbooks. *Issues in Language Teaching*, 2(2), 113-136.
- Sablan, J. R. (2019). Can you really measure that? Combining critical race theory and quantitative methods. *American Educational Research Journal*, 56(1), 178-203. <https://doi.org/10.3102/0002831218798325>
- Selvi, A. F., & Kocaman, C. (2021). (Mis-/Under-) Representations of gender and sexuality in locally-produced ELT materials. *Journal of Language, Identity & Education*, 20(2), 118-133. <https://doi.org/10.1080/15348458.2020.1726757>
- Soars, J., Soars, L., & McCaul, J. (2019a). *Headway: Beginner student's book* (5th ed.). Oxford University Press.
- Soars, J., Soars, L., & Hancock, P. (2019b). *Headway: Elementary student's book* (5th ed.). Oxford University Press.
- Soars, J., Soars, L., & Hancock, P. (2019c). *Headway: Pre-intermediate student's book* (5th ed.). Oxford University Press.
- Soars, J., Soars, L., & Hancock, P. (2019d). *Headway: Intermediate student's book* (5th ed.). Oxford University Press.
- Soars, J., Soars, L., & Hancock, P. (2019e). *Headway: Upper-intermediate student's book* (5th ed.). Oxford University Press.
- Soars, J., Soars, L., & Hancock, P. (2019f). *Headway: Advanced student's book* (5th ed.). Oxford University Press.
- Taylor-Mendes, C. (2009). Construction of racial stereotypes in English as a foreign language (EFL) textbooks: Images as discourse. *Race, culture, and identities in second language education: Exploring critically engaged practice*, 64-80.
- Templeton, A. R. (1998). Human races: A genetic and evolutionary perspective. *American Anthropologist*, 100(3), 632-650. <https://doi.org/10.1525/aa.1998.100.3.632>

- United Nations, Department of Economic and Social Affairs, Population Division. (2024). *World population prospects 2024: Summary of results* (UN DESA/POP/2024/TR/NO. 9). United Nations. https://population.un.org/wpp/assets/Files/WPP2024_Summary-of-Results.pdf
- Wei, R., & Su, J. (2012). The statistics of English in China: An analysis of the best available data from government sources. *English Today*, 28(3), 10-14. doi:10.1017/S0266078412000235