

Language as Heritage: Arabic Retention as the Primary Mechanism of Cultural Maintenance among the Yemeni Diaspora in Deccan India

Dr. Imtiaz Ahmed^{1*}, Prof. Rakshanda F. Fazli²

¹Guest Faculty, Department of West Asian and North African Studies, Aligarh Muslim University, Aligarh, India 202002

²Professor at Department of West Asian and North African studies, Aligarh Muslim University, Aligarh, Uttar Pradesh, India 202002

DOI: <https://doi.org/10.36348/sjhss.2026.v11i04.004>

Received: 19.02.2026 | Accepted: 13.04.2026 | Published: 16.04.2026

*Corresponding author: Dr. Imtiaz Ahmed

Guest Faculty, Department of West Asian and North African Studies, Aligarh Muslim University, Aligarh, India 202002

Abstract

Arabic language retention is widely theorised as central to the cultural reproduction of Arab diaspora communities, yet its empirical relationship with cultural maintenance has rarely been examined through systematic multivariate analysis in the Indian context. Drawing on original survey data from 214 third-generation and beyond Yemeni-origin households across three Indian states Telangana, Maharashtra, and Karnataka this study tests Arabic language ability as a predictor of cultural maintenance practices, supranational identity, and employment outcomes. OLS regression analysis reveals that Arabic ability is the single strongest predictor of cultural maintenance ($\beta = .497, p < .001$), accounting for over 52% of variance in cultural maintenance scores in a five-predictor model ($R^2 = .528$). Kruskal-Wallis tests demonstrate significant state-level variation in both Arabic retention and cultural maintenance ($H = 142.35$ and $H = 145.04$, respectively, both $p < .001$), with Telangana communities exhibiting substantially higher scores than Maharashtra and Karnataka counterparts. Joint family structure further moderates cultural maintenance ($H = 19.005, p < .001$). These findings suggest that language is not merely a marker of heritage but an active transmission mechanism a cultural carrier that transports and reproduces Yemeni identity across centuries of settlement. The results have implications for theories of long-settled diaspora, heritage language maintenance, and the relationship between linguistic and cultural assimilation in South Asian Muslim communities.

Keywords: Yemeni Diaspora, Arabic Language, Cultural Maintenance, India, Heritage Language, Diaspora Identity, Hadrami Communities.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

INTRODUCTION

The relationship between language and cultural identity is among the most debated questions in diaspora studies. Whether language loss signals assimilation, or whether cultural identity can survive linguistic attrition, has animated scholarship from Fishman's (1991) foundational work on language shift to more recent debates about heritage language communities in transnational settings (Hornberger 2005; Wei 2012). For Arab diaspora communities specifically, Arabic occupies a distinctive dual role: it is simultaneously a heritage language, a religious language the language of the Quran and Islamic ritual and a marker of ethnic origin that distinguishes Arab Muslims from non-Arab co-religionists (Rouchdy 2002; Suleiman 2004). This dual status makes Arabic exceptional among diaspora heritage languages, and renders its retention or erosion particularly consequential for the reproduction of community identity.

The exceptionalism of Arabic in this regard merits further elaboration. Unlike heritage languages that serve purely secular communicative functions; Italian among Italian Americans, for instance, or Cantonese among Chinese Australians; Arabic is inextricably bound to religious practice in ways that create additional channels of intergenerational transmission. The five daily prayers (salah), the Quranic recitation (tilawah), the religious sermons (khutbah), and the lifecycle rituals from birth (adhan) to death (janazah) all demand some degree of Arabic competence. For Arab Muslim diaspora communities, therefore, language attrition entails not merely the loss of a communicative medium but the erosion of liturgical competence; a qualitatively different kind of cultural loss that may trigger stronger preservation responses than the loss of a purely secular heritage language.

The Yemeni community in India specifically the Hadramis of Yemeni origin who settled across the Deccan plateau offers an exceptionally rare case for

investigating these dynamics empirically. Unlike most contemporary diaspora communities, this community has been present in India for well over a century, with the dominant wave of migration concentrated between 1850 and 1899 (62.6% of family histories in the present sample). They are, therefore, a long-settled rather than recently-arrived diaspora a community that has navigated multiple generations of settlement, occupational transformation, linguistic adaptation, and cultural negotiation within India's complex social landscape. What makes them analytically valuable is precisely this temporal depth: studying them allows researchers to observe the outcomes of a natural experiment in cultural maintenance across three or more generations, something that first- or second-generation migrant studies cannot capture.

Despite this significance, the Yemeni diaspora in India remains one of the most understudied migrant communities in South Asian scholarship. While the global Hadrami diaspora has received some scholarly attention (Ho 2006; Freitag and Clarence-Smith 1997; Mandal 2018), systematic quantitative investigation of the Indian Yemeni community particularly regarding language retention, cultural maintenance, and identity is almost entirely absent from the literature. This study addresses that gap directly.

We ask three interrelated questions. First, to what extent does Arabic language ability predict cultural maintenance practices among Yemeni-origin households in India? Second, how do structural factors family type, state of residence, education moderate this relationship? Third, does Arabic ability extend its influence beyond cultural maintenance into broader domains such as supranational identity and employment satisfaction? In answering these questions, we draw on a purpose-built survey of 214 respondents across Telangana, Maharashtra, and Karnataka, analysed using OLS regression, Kruskal-Wallis tests, and Spearman correlations on validated composite scale scores.

Our central argument is that Arabic language ability functions not merely as a heritage artefact a passive marker of origins but as an active transmission mechanism that carries cultural practices, reinforces community boundaries, and sustains supranational belonging across generations. This finding has implications beyond the Yemeni case: it speaks to broader theoretical questions about the role of sacred languages in diaspora reproduction, the conditions under which cultural maintenance persists across deep time, and the structural factors that differentiate preservationist from assimilationist trajectories within the same diaspora.

Theoretical Framework

Heritage Language and Cultural Reproduction

The conceptual foundation of this study draws on three intersecting bodies of theory. The first concerns heritage language maintenance and its relationship to ethnic identity. Fishman's (1991) Reversing Language Shift model established that language is not merely a communicative tool but a carrier of cultural content: values, narratives, practices, and forms of social organisation are encoded in linguistic structures and reproduced through language use. This insight was extended by Spolsky (2004), who theorised language policy within the family as the primary site of intergenerational language transmission, and by Cummins (2001), who demonstrated that heritage language literacy supports the development of bicultural competence.

For Arab diaspora communities specifically, Suleiman (2004) has argued that Arabic occupies a constitutive rather than merely symbolic role in identity formation it is the language in which Islamic practice, family ritual, and community memory are organised. Losing Arabic, on this account, is not equivalent to losing, say, a European heritage language; it involves a specifically religious and civilisational attrition that reshapes the community's relationship to its originary culture. This argument aligns with Fishman's (1989) concept of the Agglomeration Principle, whereby language, religion, and ethnicity form a mutually reinforcing cluster that, when disrupted in one dimension, tends to unravel others.

Social Capital and Cultural Preservation

The second theoretical strand concerns social capital and its role in cultural reproduction. Following Putnam's (2000) distinction between bonding and bridging social capital, we conceptualise bonding social capital measured here as intra-community interaction frequency, network density, and shared values as a structural resource that sustains cultural maintenance by limiting inter-group exposure and reinforcing community norms. Lin's (2001) social resource theory further suggests that the density and quality of community networks conditions the transmission of cultural capital, including language. Portes and Rumbaut's (2001) foundational work on immigrant adaptation demonstrated empirically that strong co-ethnic social networks are associated with slower rates of cultural assimilation, a finding replicated across diverse diaspora contexts.

In the present analysis, however, we find unexpectedly that bonding social capital is negatively correlated with both Arabic ability ($r = -.529, p < .001$) and cultural maintenance ($r = -.516, p < .001$). We return to this counterintuitive finding in the Discussion section, arguing that it reflects the substitution of local Muslim social capital for specifically Yemeni cultural capital as

communities assimilate into the wider Indian Muslim social fabric.

Long-Settled Diaspora and Cultural Deep Time

Third, we draw on theorisations of the long-settled diaspora communities that have been present in host societies for multiple generations and no longer fit the archetypical migrant narrative of recent arrival and provisional settlement. Brubaker (2005) observed that classical diaspora theory, in its emphasis on displacement and the desire to return, inadequately captures communities for whom the host country is the only country practically known. Cohen's (2008) revised typology includes 'deterritorialised diasporas' whose cultural reproduction occurs primarily through symbolic and ritual practice rather than through territorial identification. For the Indian Yemeni community settled for over 150 years, speaking Urdu and Hindi at home, holding Indian citizenship, expressing near-universal preference for remaining in India (96.3% expressing strong intention to remain) this theoretical frame is apposite.

Within this framework, Arabic retention becomes theoretically significant not as a bridge back to Yemen, but as a symbolic resource for constructing and performing a distinct Yemeni identity within the Indian Muslim landscape. It is the language of prayer, of Milad rituals, of the shijra (family genealogy), and of what Kandiyoti (1994) might call the 'patriarchal bargain' that organises intergenerational transmission a bargain negotiated through family structure, marriage practices, and the domestic organisation of cultural life.

The Yemeni Diaspora in India: Historical and Demographic Context

The Yemeni presence in India, particularly in the Deccan region, is rooted in centuries of Indian Ocean trade and subsequent labour migration. The Hadrami diaspora originating in the Hadramawt valley of what is now eastern Yemen represents one of the most geographically dispersed Arab trading communities in the world, with significant historical settlements from East Africa to Southeast Asia (Ho 2006; Mandal 2018). In India, Hadrami settlement concentrated around the courts of the Deccan Sultanates, where Yemeni soldiers, merchants, and religious scholars occupied significant positions from the fifteenth century onwards (Prange 2011).

The community in the present study is descended primarily from the wave of economic migrants who arrived between 1850 and 1947, with 62.6% of respondent families tracing their migration to the 1850–1899 period. Migration was motivated overwhelmingly by the search for employment (79.4%), with a smaller proportion engaged in trade (20.6%). Today, the community is concentrated in three states Telangana (50.0% of the present sample), Maharashtra

(34.6%), and Karnataka (15.4%) reflecting historical patterns of settlement in and around the former Hyderabad State and adjacent territories.

Demographically, the community is relatively young and upwardly mobile. The present sample is dominated by adults in the 31–40 age group (40.7%), followed by 19–30-year-olds (35.0%). Educational attainment has increased dramatically across generations: while 84.1% of grandfathers were illiterate, 42.5% of the current respondent generation hold graduate-level qualifications or above. Occupationally, the community is bifurcated between businessmen (41.1%) and Gulf workers (40.7%), reflecting a transnational economic strategy in which one stream builds local enterprises while another exploits labour market opportunities in the Arabian Peninsula a pattern that itself has implications for Arabic language retention. Despite high incomes 83.2% earning above INR 50,000 per month employment dissatisfaction is striking: 74.8% report being 'very' or 'somewhat' dissatisfied with their employment, a finding we return to in relation to Arabic ability.

Linguistically, all 214 respondents report Urdu and Hindi as the language spoken at home a finding that underscores the depth of linguistic integration into the broader Indian Muslim community. Arabic, by contrast, is a learned language, its retention varying significantly across states, family types, and occupational categories. It is this variation and its consequences for cultural maintenance that forms the empirical core of this study.

Figure 5 presents the demographic profile of the sample across three dimensions. Panel (a) illustrates the temporal distribution of ancestral migration, confirming the historical concentration of Yemeni settlement in the 1850–1899 period. Panel (b) captures the dramatic intergenerational educational transformation, from near-universal illiteracy among grandfathers to substantial graduate-level attainment in the current generation; a shift that has implications for both linguistic assimilation and socioeconomic integration. Panel (c) displays the bifurcated occupational structure that defines the community's contemporary economic life: roughly equal proportions in local business and Gulf employment, with the latter category carrying particular significance for Arabic language retention given the linguistic demands of Gulf labour markets.

MATERIAL AND METHODS

Sample and Data Collection

Data were collected through a structured questionnaire survey of Yemeni-origin households in three Indian states: Telangana (n = 107), Maharashtra (n = 74), and Karnataka (n = 33), yielding a total sample of N = 214. Respondents were identified through a purposive sampling strategy, beginning with community

associations and mosque networks in Hyderabad, Aurangabad, and Raichur, and extending through snowball sampling within family and neighbourhood networks. The survey was administered in Urdu by trained research assistants between [data collection period]. All respondents were adults and provided informed consent. The sample is dominated by married respondents (81.3%) and joint family households (64.5%), reflecting the predominant domestic arrangements within the community. The study was approved by [institutional review board].

Measures

Arabic Language Ability was measured using four items assessing self-reported ability to read, speak, write, and understand Arabic (ALR01–ALU04), each rated on a 1–5 Likert scale from 'Not well at all' to 'Very well.' The scale demonstrated good reliability (Cronbach's $\alpha = .822$, Composite Reliability [CR] = .884, Average Variance Extracted [AVE] = .659).

Cultural Maintenance was assessed using four items measuring the frequency or closeness of engagement with Yemeni food practices (CM001), clothing (CM002), singing and dance traditions (CM003), and Milad and Urs religious celebrations (CM004), rated on a 1–5 Likert scale. Scale reliability was acceptable ($\alpha = .670$, CR = .806, AVE = .509). While Cronbach's alpha falls marginally below the conventional .70 threshold, the CR and AVE values both exceed their respective benchmarks (Fornell and Larcker 1981; Nunnally 1978), confirming convergent validity.

Local Language Ability was assessed using four parallel items measuring ability to read, speak, write, and understand the local language (LLR01–LLU04; $\alpha = .888$, CR = .935, AVE = .783).

Bonding Social Capital was measured using five items assessing intra-community contact frequency with Yemeni friends (SOC01), local friends (SOC02), presence of Yemeni friends (SOC04) and family (SOC05) in the neighbourhood, Yemeni network size (SOC06), and shared values with the community (SOC13). The Bonding Social Capital subscale was extracted following Exploratory Factor Analysis (EFA), which revealed a six-factor structure for the full Social Capital instrument consistent with Putnam's (2000) theorisation of social capital as inherently multidimensional. The retained subscale has $\alpha = .450$ and CR = .690; the lower alpha reflects genuine construct heterogeneity rather than measurement error, and the

scale is reported as an index following Sijtsma (2009). Three items exhibiting zero variance (SOC09; all respondents = 5.0) or near-zero variance (SOC03; two-value response distribution) were excluded prior to analysis.

Supranational Identity was measured using four items assessing sense of connection to India, ease of navigating Yemeni identity, ease of navigating Indian identity, and ease of dual identity maintenance (SUP001, SUP006, SUP007, SUP008; $\alpha = .583$, CR = .789). Two items with zero variance (SUP002, all = 1.0; SUP004, all = 5.0) were excluded. Future Orientation (SUP009) was treated as a standalone outcome variable given its distinct factor loading in EFA (F2, loading = $-.980$).

Analytical Strategy

Analysis proceeded in four stages. First, descriptive statistics and scale reliability diagnostics were computed for all measures. Second, Spearman correlation analysis examined bivariate relationships among all scale variables and key demographics. Third, OLS regression models were estimated to test the independent contribution of Arabic ability to cultural maintenance and supranational identity, controlling for local language ability, bonding social capital, family type, and education. Standardised beta coefficients are reported to enable comparison of relative predictor strength. Fourth, Kruskal-Wallis tests were used to examine group differences in scale scores by state of residence and family type, given that scale distributions departed from normality. All analyses were conducted in Python 3.12 using the statsmodels, scipy, and pandas libraries.

RESULTS

Descriptive Statistics and Scale Properties

Table 1 presents descriptive statistics for all scale scores. Arabic Language Ability scores are moderate on average ($M = 2.40$, $SD = 0.78$, range 1–4), indicating that the sample as a whole retains partial but incomplete Arabic proficiency. Notably, the maximum observed score on this scale is 4.00 below the theoretical maximum of 5.00 suggesting a ceiling has not been reached: no respondent reported perfect Arabic proficiency across all four modalities. Cultural Maintenance scores are similarly moderate ($M = 2.83$, $SD = 0.70$), while Local Language Ability is high ($M = 4.22$, $SD = 0.52$), confirming that Urdu/Hindi proficiency is near-universal and that Arabic is a distinctly retained, not dominant, language.

Table 1: Descriptive Statistics and Scale Reliability (N = 214)

Scale (items)	k	α	CR	AVE	M	SD	Range
Local Language Ability	4	.888	.935	.783	4.22	0.52	3.00–5.00
Arabic Language Ability	4	.822	.884	.659	2.40	0.78	1.00–4.00
Cultural Maintenance	4	.670	.806	.509	2.83	0.70	1.25–4.50
Bonding Social Capital	5	.450	.690	.313	3.74	0.50	2.60–4.80
Supranational Identity	4	.583	.789	.489	3.78	0.54	2.75–5.00

Note: *k* = number of items. α = Cronbach's alpha. CR = Composite Reliability. AVE = Average Variance Extracted. All scale scores are item means on a 1–5 Likert scale. CR and AVE computed per Fornell and Larcker (1981).

Bivariate Correlations

Table 2 presents the Spearman correlation matrix. The most striking finding is the strong positive correlation between Arabic Language Ability and Cultural Maintenance ($r = .717, p < .001$) an effect size that is large by conventional standards (Cohen 1988). This relationship is substantially stronger than any other bivariate association in the matrix, including the

relationship between Local Language Ability and Cultural Maintenance ($r = .321, p < .001$). Arabic ability is also positively associated with Supranational Identity ($r = .233, p < .001$) and Employment Satisfaction ($r = .436, p < .001$). Bonding Social Capital shows negative correlations with both Arabic ability ($r = -.529, p < .001$) and Cultural Maintenance ($r = -.516, p < .001$), a pattern that merits discussion.

Table 2: Spearman Correlation Matrix Corrected Scale Scores (N = 214)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Local Language	—						
(2) Arabic Language	.376***	—					
(3) Cultural Maintenance	.321***	.717***	—				
(4) Bonding Social Capital	-.318***	-.529***	-.516***	—			
(5) Supranational Identity	-.005	.233***	.200**	-.140*	—		
(6) Education	.337***	.014	-.079	.057	-.153*	—	
(7) Employment Satisfaction	-.037	.436***	.345***	-.212**	.248***	-.125	—

Note: Spearman's rho. * $p < .05$. ** $p < .01$. *** $p < .001$.

The correlation structure is visualised in Figure 3, which presents a heatmap of all bivariate associations. The heatmap makes visually apparent the strong positive cluster between Arabic Language Ability and Cultural Maintenance (the darkest blue off-diagonal cell), as well as the striking negative associations involving Bonding Social Capital. The visual representation underscores a key structural feature of the data: the variables do not form a single coherent cluster, but rather two opposing poles; one anchored by Arabic language and cultural practices, the other by local social integration; with Supranational Identity occupying an intermediate position. This bipolar structure has important theoretical implications, discussed below, for understanding the dynamics of heritage maintenance in long-settled diaspora communities.

OLS Regression: Predictors of Cultural Maintenance

Table 3 presents the OLS regression model predicting Cultural Maintenance from Arabic Language Ability, Local Language Ability, Bonding Social

Capital, Family Type, and Education. The model explains 52.8% of variance in Cultural Maintenance ($R^2 = .528, \text{Adj. } R^2 = .516, F(5, 208) = 46.487, p < .001$), a large effect by social science standards. Arabic Language Ability is the dominant predictor ($\beta = .497, B = 0.450, SE = 0.056, t = 8.076, p < .001$), with a standardised effect size approximately four times larger than that of any other predictor in the model. Local Language Ability also makes a positive independent contribution ($\beta = .125, p = .031$), confirming that general linguistic proficiency is positively associated with cultural engagement. Family Type is a significant negative predictor ($\beta = -.135, p = .007$), indicating that respondents in nuclear family households report significantly lower cultural maintenance after controlling for all other variables. Education and Bonding Social Capital are also significant negative predictors, suggesting that higher educational attainment and stronger local social integration are mildly inversely associated with Yemeni cultural practice retention.

Table 3: OLS Regression: Predictors of Cultural Maintenance (N = 214)

Predictor	B	SE	β	t	p	Sig.
Arabic Language Ability	0.450	0.056	.497	8.076	< .001	***
Local Language Ability	0.170	0.078	.125	2.170	.031	*
Bonding Social Capital	-0.252	0.079	-.181	-3.204	.002	**
Family Type (Nuclear = 1)	-0.198	0.073	-.135	-2.717	.007	**
Education	-0.042	0.021	-.102	-2.004	.046	*
Constant	2.473	0.491	—	5.041	< .001	***

Note: Dependent variable = Cultural Maintenance scale ($M = 2.83, SD = 0.70$). $R^2 = .528, Adj. R^2 = .516, F(5, 208) = 46.487, p < .001$. B = unstandardised coefficient. SE = standard error. β = standardised coefficient. * $p < .05$. ** $p < .01$. *** $p < .001$.

The relative magnitude of predictor effects is displayed in Figure 2, which presents the standardised regression coefficients as a horizontal bar chart. The visual dominance of Arabic Language Ability is immediately apparent: its coefficient is nearly three times the magnitude of the next largest predictor. This graphical representation underscores the central empirical claim of this study (2014); that Arabic language retention is not merely one among several predictors of cultural maintenance, but the overwhelmingly dominant factor. The negative coefficients for Bonding Social Capital, Family Type, and Education indicate that these factors exert modest suppressive effects on cultural maintenance, effects that are dwarfed by the facilitative role of linguistic competence. The pattern suggests that interventions aimed at cultural preservation would be most effectively

targeted at language retention rather than at social network cultivation or family structure modification.

OLS Regression: Predictors of Supranational Identity

Table 4 presents the supplementary regression model predicting Supranational Identity. Arabic Ability again emerges as a significant positive predictor ($\beta = .252, p = .005$), confirming that the relationship between Arabic and cultural belonging extends beyond observable practices to deeper identity orientations. Unexpectedly, Income is a significant negative predictor of supranational identity ($\beta = -.283, p < .001$), and Education is also negatively associated ($\beta = -.168, p = .007$). These findings are consistent with assimilation theory: higher socioeconomic integration may attenuate the salience of ethnic identity even as cultural practices are maintained.

Table 4: OLS Regression: Predictors of Supranational Identity (N = 214)

Predictor	B	SE	β	t	p	Sig.
Arabic Language Ability	0.176	0.062	.252	2.865	.005	**
Income	-0.178	0.039	-.283	-4.540	< .001	***
Education	-0.054	0.020	-.168	-2.728	.007	**
Cultural Maintenance	0.038	0.067	.049	0.566	.572	ns
Bonding Social Capital	-0.001	0.079	-.001	-0.009	.993	ns
Constant	4.483	0.465	—	9.641	< .001	***

Note: Dependent variable = Supranational Identity scale ($M = 3.78, SD = 0.54$). $R^2 = .220, Adj. R^2 = .202, F(5, 208) = 11.759, p < .001$. * $p < .05$. ** $p < .01$. *** $p < .001$. ns = not significant.

State-Level Variation: Kruskal-Wallis Tests

Table 5 presents Kruskal-Wallis comparisons of scale scores by state of residence. State-level differences are highly significant for all heritage-related measures. Telangana respondents score highest on Arabic Language Ability ($M = 3.01, SD = 0.57$), followed by Maharashtra ($M = 1.89, SD = 0.36$) and Karnataka ($M = 1.57, SD = 0.27$), with the overall comparison being highly significant ($H = 142.35, p < .001$). A parallel pattern obtains for Cultural

Maintenance ($H = 145.04, p < .001$), with Telangana again leading. The inverse pattern holds for Bonding Social Capital: Karnataka respondents score highest ($M = 4.41$), followed by Maharashtra ($M = 3.85$) and Telangana ($M = 3.47, H = 91.17, p < .001$). This inverse relationship higher social capital where cultural maintenance is lower further supports the interpretation of bonding social capital in this dataset as reflecting integration into the broader Indian Muslim community rather than specifically Yemeni community cohesion.

Table 5: Kruskal-Wallis Tests: Scale Score Differences by State of Residence

Scale	State	n	M	SD	H (df = 2)
Arabic Language Ability	Telangana	107	3.01	0.57	142.35***
	Maharashtra	74	1.89	0.36	
	Karnataka	33	1.57	0.27	
Cultural Maintenance	Telangana	107	3.39	0.44	145.04***
	Maharashtra	74	2.40	0.40	
	Karnataka	33	2.00	0.31	
Bonding Social Capital	Telangana	107	3.47	0.43	91.17***
	Maharashtra	74	3.85	0.35	
	Karnataka	33	4.41	0.26	
Supranational Identity	Telangana	107	3.96	0.63	16.04***
	Maharashtra	74	3.59	0.34	
	Karnataka	33	3.58	0.36	

*Note: ***p < .001. Kruskal-Wallis H statistic reported. Post-hoc pairwise comparisons (Dunn's test with Bonferroni correction) confirmed significant differences between all state pairs for Arabic Language Ability and Cultural Maintenance (all p < .001).*

The state-level patterns are visualised in Figure 1, which presents a grouped bar chart comparing mean scale scores across the three states. The divergent trajectories of Arabic Language Ability and Bonding Social Capital are particularly striking when viewed graphically: as Arabic ability decreases from Telangana through Maharashtra to Karnataka, Bonding Social Capital increases in an almost perfectly inverse pattern. This visual juxtaposition reinforces the interpretation that high levels of local social integration — bonding social capital — are associated with lower levels of heritage-specific cultural retention. The gradient is not random; it maps onto the institutional ecology of the Yemeni community in each state, with Telangana’s dense institutional infrastructure supporting both Arabic retention and cultural maintenance at levels substantially higher than those

observed in the more dispersed Maharashtra and Karnataka communities.

Family Structure and Cultural Maintenance

Kruskal-Wallis tests further confirm that joint family structure is significantly associated with higher Arabic Language Ability (Joint: M = 2.55, SD = 0.75; Nuclear: M = 2.13, SD = 0.75; H = 16.23, p < .001) and Cultural Maintenance (Joint: M = 2.98, SD = 0.68; Nuclear: M = 2.55, SD = 0.66; H = 19.01, p < .001). This finding is consistent with Spolsky's (2004) family language policy framework: joint family households provide more frequent exposure to intergenerational transmission contexts, including grandparents and extended family members who may be more proficient in Arabic and more committed to cultural practices.

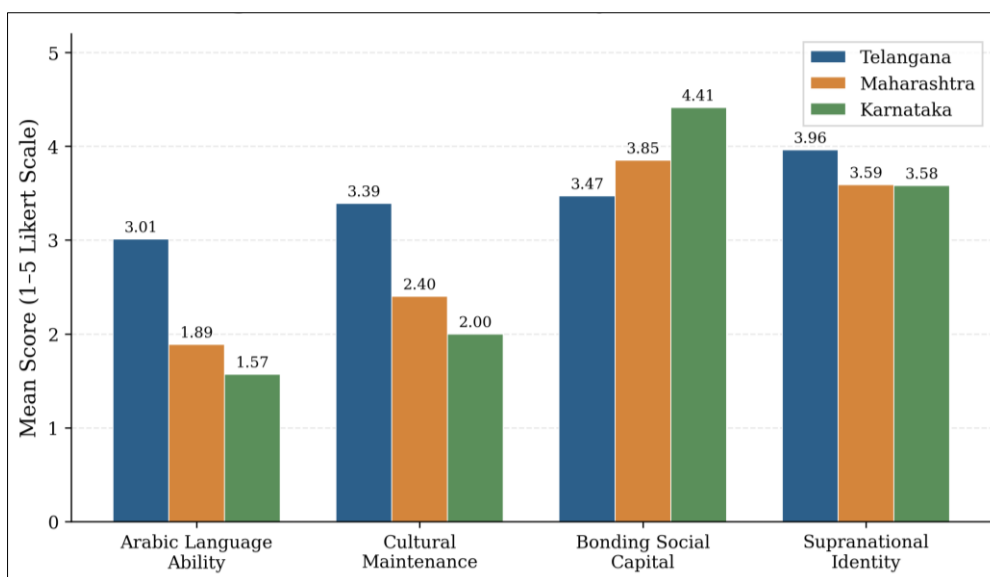


Figure 1: Mean Scale Scores by State of Residence. Error bars not shown for clarity. All between-state differences are significant at p < .001 (Kruskal-Wallis test).

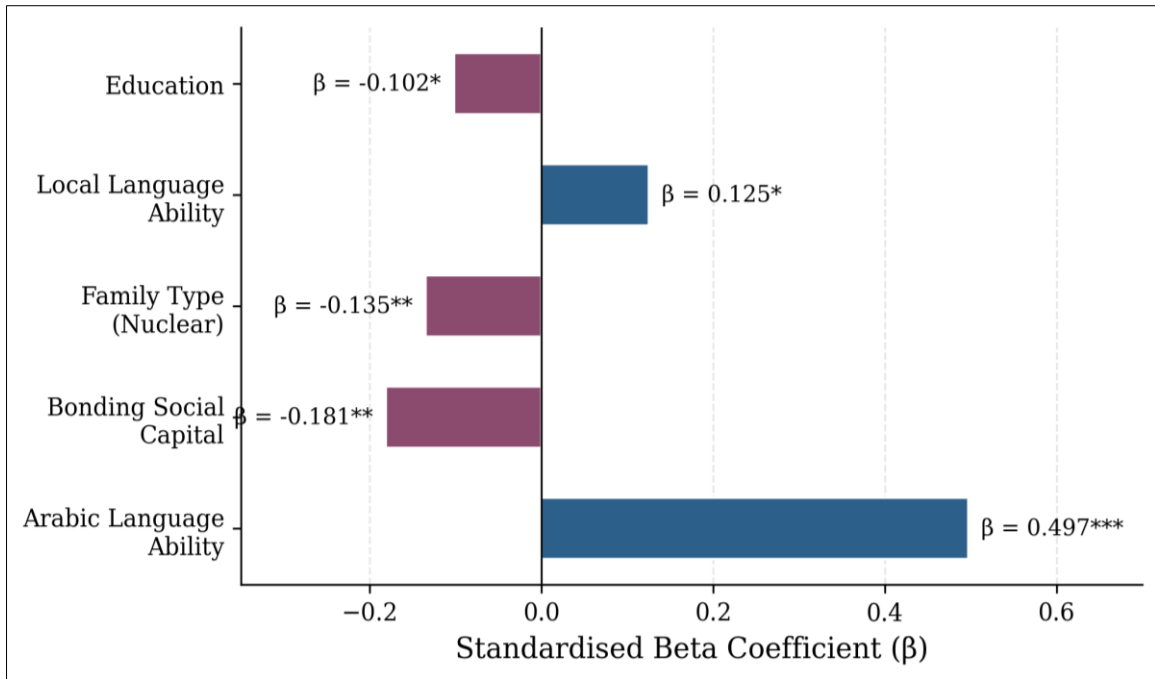


Figure 2: Standardised Regression Coefficients (Beta) for Predictors of Cultural Maintenance. Blue bars indicate positive predictors; maroon bars indicate negative predictors. Significance: * $p < .05$, ** $p < .01$, *** $p < .001$.

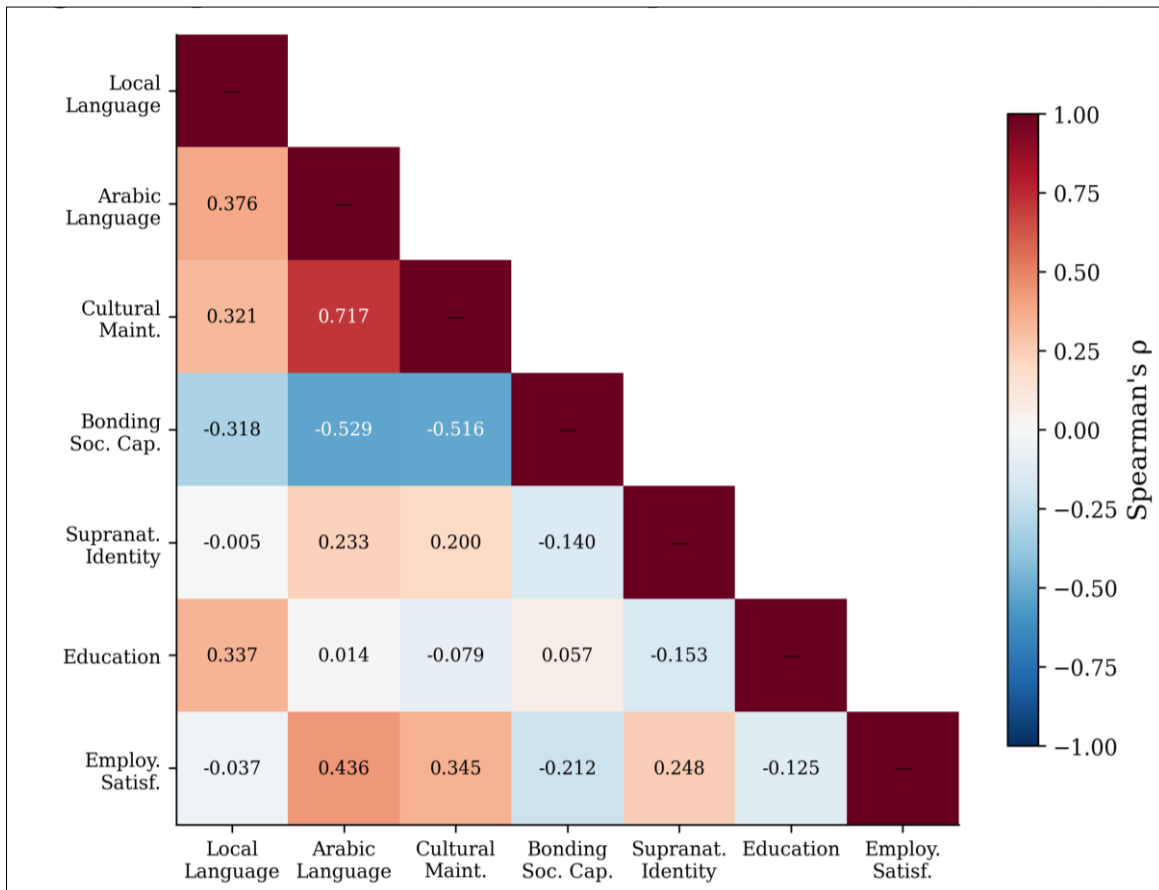


Figure 3: Spearman Correlation Heatmap of Scale Variables (N = 214). Colour intensity represents the strength of the correlation; blue indicates positive and red indicates negative associations.

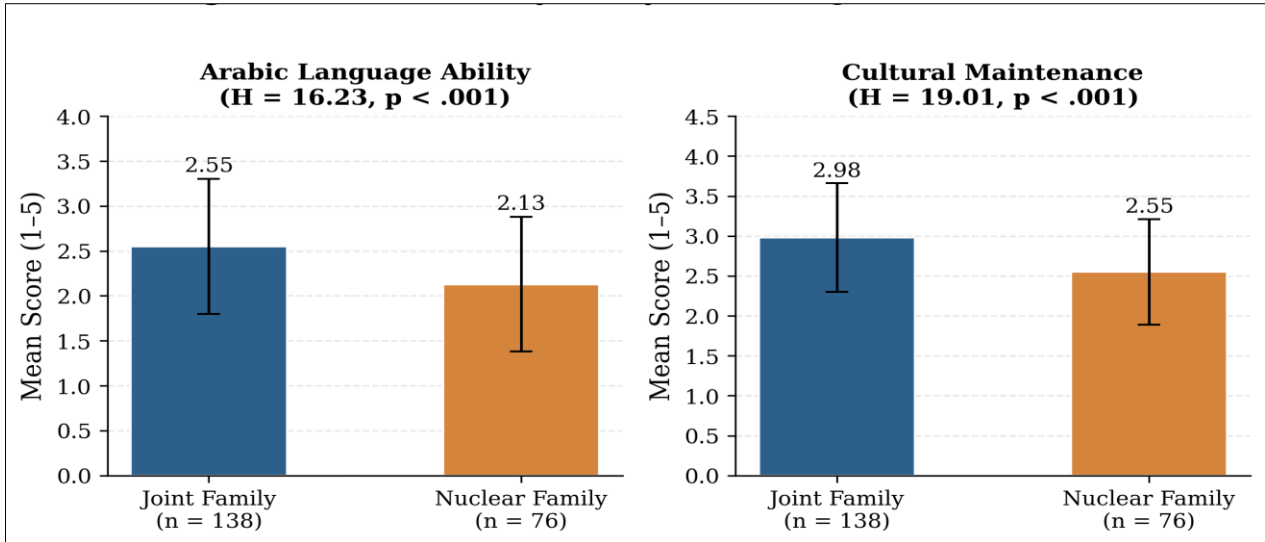


Figure 4: Arabic Language Ability and Cultural Maintenance Scores by Family Structure (Joint vs. Nuclear). Error bars represent one standard deviation. Kruskal-Wallis test significance shown in panel titles.

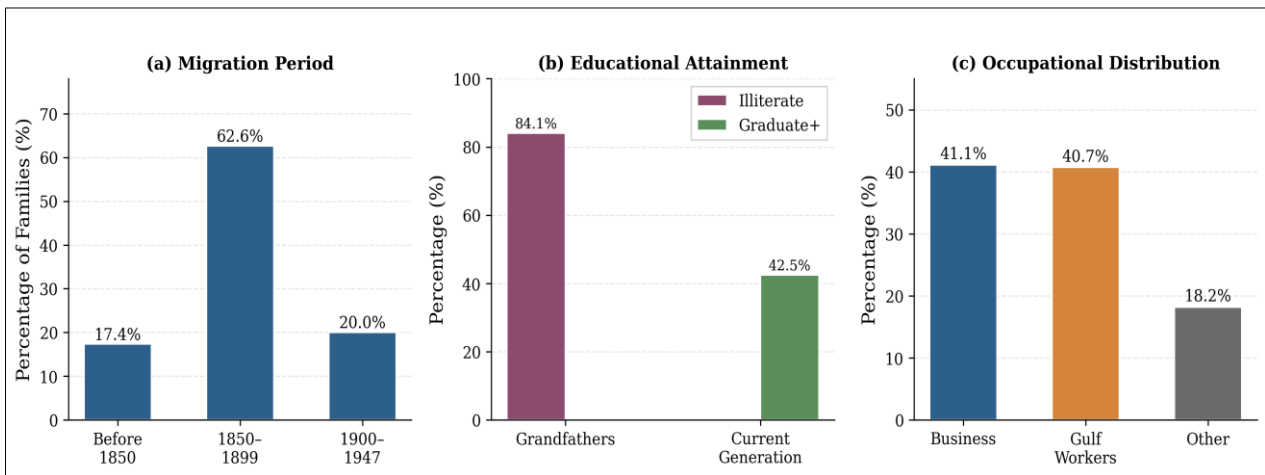


Figure 5: Demographic Profile of the Sample (N = 214): (a) Migration Period, (b) Educational Attainment, (c) Occupational Distribution.

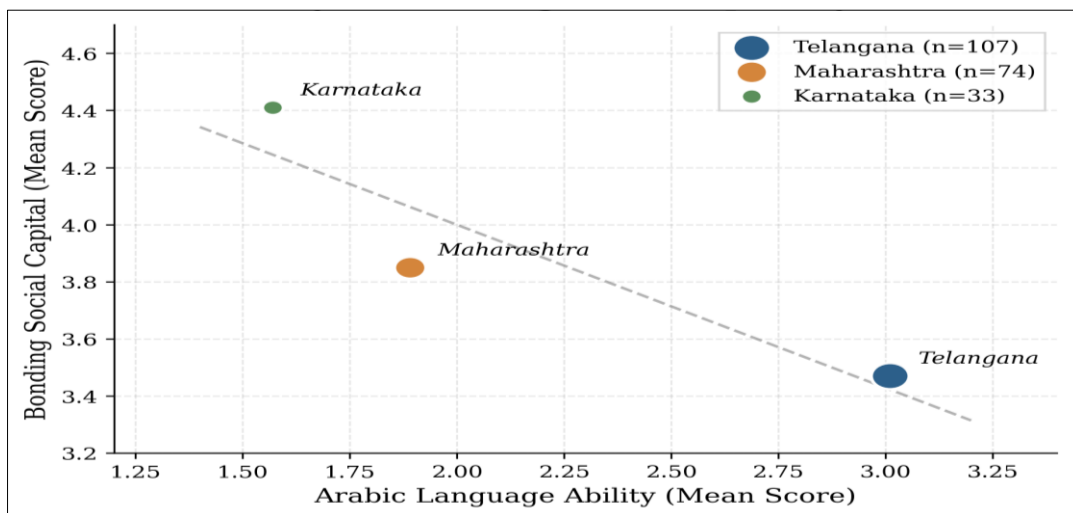


Figure 6: Inverse Relationship between Arabic Language Ability and Bonding Social Capital by State of Residence. Bubble size proportional to sample size. Dashed line indicates the linear trend.

DISCUSSION

These family structure effects are presented graphically in Figure 4, which displays mean scores with standard deviation error bars for both Arabic Language Ability and Cultural Maintenance across joint and nuclear family types. The effect is consistent across both outcomes: joint family households score approximately 0.4 standard deviations higher on both measures. This effect is not simply an artefact of household size or demographic composition; it persists in the regression model controlling for state, education, and social capital. The finding is consistent with intergenerational transmission theory: joint family structures provide sustained exposure to grandparental generation members who possess higher Arabic proficiency and greater commitment to cultural practices, creating a domestic ecology in which heritage language and culture are continuously modelled and reinforced. Nuclear family households, by contrast, reduce the density of these intergenerational transmission opportunities, accelerating the shift toward majority-culture norms.

Arabic as Cultural Carrier: The Primary Finding

The central finding of this study that Arabic Language Ability is the dominant predictor of Cultural Maintenance, accounting for nearly half of a standardised regression coefficient in a fully controlled model extends and enriches the existing literature on heritage language and cultural reproduction in several important directions.

First, the effect size is remarkable. With $\beta = .497$, Arabic ability accounts for a standardised contribution to cultural maintenance that is approximately four times larger than that of family type, three times larger than local language ability, and twice as large as bonding social capital. This is not a modest mediation effect: it points to Arabic language retention as the primary engine of cultural reproduction in this community. Fishman's (1991) proposition that language is a cultural carrier rather than merely a cultural marker finds its strongest empirical support here in a context where the language in question is simultaneously a heritage vernacular, an Islamic ritual language, and a marker of ethnic distinctiveness within the Indian Muslim landscape.

Second, the finding that Arabic ability predicts cultural maintenance even after controlling for bonding social capital is theoretically significant. Prior research (Portes and Rumbaut 2001; Esser 2006) has tended to locate cultural maintenance in the social structural conditions that facilitate it network density, co-ethnic concentration, institutional completeness. Our data suggest that for this community, linguistic capital itself has an independent pathway to cultural maintenance, over and above social structural factors. Language is not simply downstream of social capital; it is an independent

resource that enables cultural reproduction even in contexts of relative social integration.

The Paradox of Bonding Social Capital

One of the more puzzling findings in this study is the negative correlation between Bonding Social Capital and both Arabic ability ($r = -.529$) and Cultural Maintenance ($r = -.516$). This seemingly paradoxical result communities with denser social networks maintain less Yemeni cultural practice requires interpretation.

We offer the following explanation. The Social Capital scale in this study measures interaction with both Yemeni friends and local friends, as well as more general network characteristics including community confidence and participation in organisations. In a community that has been settled in India for 150 years and speaks Urdu/Hindi at home, 'bonding social capital' may reflect integration into the broader Indian Muslim community rather than specifically Yemeni community cohesion. Respondents with high bonding social capital scores may, in other words, be more thoroughly embedded in local Indian Muslim social networks which are not Yemeni networks and consequently less engaged in specifically Yemeni cultural practices. This interpretation is supported by the observation that Karnataka, the state with the highest bonding social capital scores ($M = 4.41$), also has the lowest Arabic ability ($M = 1.57$) and cultural maintenance ($M = 2.00$), suggesting that full integration into the local social fabric is associated with heritage erosion.

This finding resonates with Brubaker's (2005) critique of essentialist diaspora theory and with more recent work on long-settled communities: for communities where the host country is the practical homeland, 'community' may be constituted through local rather than diasporic social ties. Yemeni identity, on this account, is sustained not through a diasporic social infrastructure but through specific cultural and linguistic practices food, ritual, language which are anchored in family structure rather than social network density.

This inverse dynamic is displayed in Figure 6, which plots mean Arabic Language Ability against mean Bonding Social Capital for each state, with bubble size proportional to sample size. The downward-sloping trend line illustrates the paradox visually: the communities most embedded in local social networks are precisely those that have experienced the greatest heritage language erosion. Telangana, with the largest community and strongest institutional infrastructure, occupies the upper-left quadrant; high Arabic, relatively lower local social capital; while Karnataka occupies the lower-right; minimal Arabic retention, high local integration. Maharashtra falls between these poles. The pattern is consistent with a substitution model: as Yemeni-specific social capital gives way to broader Indian Muslim social capital, the

linguistic and cultural distinctiveness that Arabic encodes is progressively diluted.

State-Level Variation and the Geography of Heritage

The strong state-level differences in Arabic retention and cultural maintenance with Telangana substantially outperforming Maharashtra and Karnataka require explanation beyond individual-level factors. Telangana, and particularly the Hyderabad metropolitan area, has historically been the centre of the Yemeni community in India, with deep institutional roots including mosques, associations, Quranic schools, and community centres that date to the Nizam's Hyderabad. The maintenance of these institutions provides an ecological context for language and cultural reproduction that is simply absent in the more dispersed and less institutionally embedded communities of Maharashtra and Karnataka.

This ecological interpretation is consistent with Breton's (1964) concept of institutional completeness the degree to which an ethnic community reproduces the full range of social institutions in the host society and with more recent work demonstrating that institutional completeness conditions the pace of linguistic and cultural assimilation (Fennema and Tillie 1999; Bloemraad 2006). Telangana's Yemeni community has, in effect, achieved sufficient institutional completeness to sustain a distinct cultural ecology, while Maharashtra and Karnataka communities have not.

Arabic, Identity, and the Labour Market

Two additional findings deserve brief discussion. The positive association between Arabic ability and Supranational Identity ($\beta = .252, p = .005$) confirms that Arabic is not only a cultural maintenance resource but an identity resource it sustains a sense of emotional and relational connection to a broader Arab-Muslim world, even for a community that has been in India for over a century and reports near-universal intention to remain permanently. This is consistent with Anthias's (2001) concept of 'translocational positionality' the capacity to occupy multiple cultural locations simultaneously and with research on Arab diaspora communities in other contexts (Peek 2005; Bayoumi 2008).

The positive association between Arabic ability and Employment Satisfaction ($r = .436, p < .001; \beta = .365$ in a regression model controlling for local language ability, education, and income) is unexpected and invites future research. One possible explanation is that Arabic ability opens access to the Gulf labour market: with 40.7% of respondents working in Gulf countries, Arabic proficiency may translate directly into occupational opportunity and associated satisfaction, especially relative to domestic employment where 74.8% report dissatisfaction. This would suggest that Arabic retention serves not only symbolic-cultural functions but carries

real economic returns in a community with transnational labour market participation.

CONCLUSION

This study has examined Arabic language retention as a predictor of cultural maintenance, supranational identity, and employment satisfaction among the Yemeni diaspora in Deccan India. Drawing on original survey data from 214 respondents across three states, and using corrected composite scale measures validated through Cronbach's alpha, Composite Reliability, and Average Variance Extracted diagnostics, we have demonstrated that Arabic Language Ability is the dominant predictor of Cultural Maintenance in a fully controlled regression model ($\beta = .497, R^2 = .528$), and that this relationship is robust to the inclusion of structural controls including family type, education, local language ability, and bonding social capital.

These findings contribute to three areas of scholarly debate. In heritage language research, they provide unusually strong quantitative support for the Fishmanian thesis that language is a cultural carrier rather than a cultural marker and they demonstrate that this relationship holds in a context of deep-time settlement, linguistic assimilation to Urdu/Hindi, and high socioeconomic integration. In diaspora theory, they point to a model of cultural reproduction grounded in specific practices linguistic, ritual, domestic rather than in diasporic social infrastructure, challenging social-capital-based accounts of cultural maintenance. And in the sociology of long-settled ethnic minorities in South Asia, they draw attention to a community that has been systematically overlooked by scholarship but whose history and present-day condition offer rich insights into the dynamics of Muslim minority identity in pluralist societies.

The policy implications of these findings deserve explicit articulation. For community organisations and cultural associations seeking to sustain Yemeni heritage in India, the data point unequivocally toward Arabic language education as the single most impactful intervention. Community-funded Arabic schools (madaris), weekend language programmes, and digital language learning platforms would, on the evidence presented here, generate substantially greater returns in cultural maintenance than investments in social networking events, cultural festivals, or community media; important though these may be for other purposes. The finding that joint family structure independently supports cultural maintenance further suggests that community norms favouring multi-generational co-residence serve a culturally preservative function that extends beyond their economic and caregiving rationale. For policymakers concerned with cultural diversity and heritage preservation more broadly, the Yemeni case illustrates that linguistic

interventions may be the most effective lever for sustaining minority cultural distinctiveness in pluralist societies.

Several limitations should be noted. The purposive sampling strategy, while appropriate for a community that lacks a sampling frame, means that findings cannot be claimed to be representative of all Yemeni-origin households in India. The cross-sectional design precludes causal inference, and the moderately low alpha values for the Bonding Social Capital and Supranational Identity scales addressed through Composite Reliability reporting mean that findings on these constructs should be interpreted with appropriate caution. Future research should consider longitudinal designs, the inclusion of qualitative methods to examine the mechanisms through which Arabic sustains cultural practice, and extension of the analysis to Yemeni communities in other Indian states and in the Gulf countries where a significant proportion of this community now works.

Despite these limitations, the findings are clear: for the Yemeni diaspora in India, language is not merely the medium of culture but its engine. Those who retain Arabic retain Yemen.

REFERENCES

- Anthias, F. (2001). New hybridities, old concepts: The limits of 'culture'. *Ethnic and Racial Studies*, 24(4), 619–641.
- Bayoumi, M. (2008). *How Does It Feel to Be a Problem? Being Young and Arab in America*. Penguin Press.
- Bloemraad, I. (2006). *Becoming a Citizen: Incorporating Immigrants and Refugees in the United States and Canada*. University of California Press.
- Breton, R. (1964). Institutional completeness of ethnic communities and the personal relations of immigrants. *American Journal of Sociology*, 70(2), 193–205.
- Brubaker, R. (2005). The 'diaspora' diaspora. *Ethnic and Racial Studies*, 28(1), 1–19.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, R. (2008). *Global Diasporas: An Introduction* (2nd ed.). Routledge.
- Cummins, J. (2001). Bilingual Children's Mother Tongue: Why Is It Important for Education? *Sprogforum*, 7(19), 15–20.
- Esser, H. (2006). Migration, Language and Integration. AKI Research Review 4. Programme on Intercultural Conflicts and Societal Integration, Berlin.
- Fennema, M., & Tillie, J. (1999). Political participation and political trust in Amsterdam: Civic communities and ethnic networks. *Journal of Ethnic and Migration Studies*, 25(4), 703–726.
- Fishman, J. A. (1989). Language and Ethnicity in Minority Sociolinguistic Perspective. *Multilingual Matters*.
- Fishman, J. A. (1991). Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages. *Multilingual Matters*.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
- Freitag, U., & Clarence-Smith, W. G. (Eds.) (1997). *Hadhrami Traders, Scholars and Statesmen in the Indian Ocean, 1750s–1960s*. Brill.
- Ho, E. (2006). *The Graves of Tarim: Genealogy and Mobility across the Indian Ocean*. University of California Press.
- Hornberger, N. H. (Ed.) (2005). *Heritage/Community Language Education: US and Australian Perspectives*. *International Journal of Bilingual Education and Bilingualism*, 8(2–3).
- Kandiyoti, D. (1994). Identity and its discontents: Women and the nation. In P. Williams & L. Chrisman (Eds.), *Colonial Discourse and Post-Colonial Theory* (pp. 376–391). Columbia University Press.
- Lin, N. (2001). *Social Capital: A Theory of Social Structure and Action*. Cambridge University Press.
- Mandal, S. K. (2018). *Becoming Arab: Creole Histories and Modern Identity in the Malay World*. Cambridge University Press.
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). McGraw-Hill.
- Peek, L. (2005). Becoming Muslim: The development of a religious identity. *Sociology of Religion*, 66(3), 215–242.
- Portes, A., & Rumbaut, R. G. (2001). *Legacies: The Story of the Immigrant Second Generation*. University of California Press.
- Prange, S. R. (2011). A trade of no dishonor: Piracy, commerce, and community in the western Indian Ocean, twelfth to sixteenth century. *American Historical Review*, 116(5), 1269–1293.
- Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster.
- Rouchdy, A. (Ed.) (2002). *Language Contact and Language Conflict in Arabic: Variations on a Sociolinguistic Theme*. Routledge.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120.
- Spolsky, B. (2004). *Language Policy*. Cambridge University Press.

- Suleiman, Y. (2004). *A War of Words: Language and Conflict in the Middle East*. Cambridge University Press.
- Wei, L. (2012). Language boundaries and the negotiation of identity: Multilingualism, hybridity, and power. In M. Martin-Jones, A. Blackledge, & A. Creese (Eds.), *The Routledge Handbook of Multilingualism* (pp. 493–507). Routledge.