

# STATE-BUILDING IN A DIVERSE SOCIETY\*

Ruth Carlitz<sup>†</sup>

Ameet Morjaria<sup>‡</sup>

Joris Mueller<sup>§</sup>

November 2022

Diversity poses fundamental challenges to state-building and development. We study the effects of one of post-colonial Africa's largest policy experiments — the Tanzanian *Ujamaa* policy — which attempted to address these challenges. *Ujamaa* aimed to create a national identity and consolidate state authority by mandating a highly diverse population to live in planned villages, where children received political education. We combine differences in exposure to *Ujamaa* across space and age to identify long-term impacts of the policy. We show persistent, positive effects on national identity based on surveys and inter-ethnic marriages. We observe no systematic differences for cohorts that were above or below treatment-age during *Ujamaa*. Our preferred interpretation, supported by evidence that considers alternative hypotheses, is that changes to educational content drive our findings. Moreover, while *Ujamaa* contributed to establishing the Tanzanian state as a legitimate central authority, it appears to have lowered demands for democratic accountability.

**JEL Codes:** I28, J15, N47, P21, Z18

**Key words:** Diversity, Public education, State-building, Identity, Africa

---

\* We thank Sam Asher, Ralph Austen, Tim Besley, Georgy Egorov, Leander Heldring, Seema Jayachandran, Horacio Larreguy, Ted Miguel, Joel Mokyr, Nathan Nunn, Elias Papaioannou, Dan Posner, Nancy Qian, Luis Rayo, Bryony Reich, Cory Smith, Jörg Spenkuch, Christopher Udry, Léonard Wantchekon, Noam Yuchtman, and participants at the 2021 Working Group on African Political Economy (WGAPE), 10<sup>th</sup> European Meeting of the Urban Economics Association, Northwestern Development Lunch, Northwestern Economic History Lunch, and the 2<sup>nd</sup> Annual Conference on Governance and Local Development (Gothenburg) for useful discussions and feedback. We are grateful to Ted Miguel, Tina Green, Mike Shand, and Philip Osafo-Kwaako for sharing historical data and many conversations. For excellent research assistance we thank Elizabeth Cameron, Mary-Crescent Carter, Mathias Ngowi, and Tenny Tsang. We thank staff at the Institute for Democracy, Citizenship, and Public Policy in Africa, at the University of Cape Town for assistance with the Afrobarometer data, the Tanzanian National Archives and the Herskovits Library of African Studies at Northwestern University for assistance with historical government data. Financial support from Northwestern's Global Poverty Research Lab, Ford Motor Company Center for Global Citizenship at Kellogg School of Management and CIFAR's Azrieli Global Scholars Program is gratefully acknowledged.

† University of Amsterdam: [r.d.carlitz@uva.nl](mailto:r.d.carlitz@uva.nl)

‡ Kellogg School of Management, CEPR and NBER: [a.morjaria@kellogg.northwestern.edu](mailto:a.morjaria@kellogg.northwestern.edu)

§ National University of Singapore: [jorismueller@nus.edu.sg](mailto:jorismueller@nus.edu.sg)

## 1. Introduction

Diversity poses fundamental challenges to state-building and development. These include the under-provision of public goods due to heterogeneous preferences (Alesina et al., 1999), inter-group conflict reflecting inequality and competition for resources (Ray and Esteban, 2017), and difficulties with communication that hinder the establishment social trust (Wimmer, 2018). In more diverse countries, citizens may feel that public choices do not represent their preferences sufficiently and may thus not want to remain together (Alesina and Spolaore, 1997), posing a challenge to state legitimacy (Holsti, 1996). Such a lack of perceived state legitimacy is in turn associated with weak state capacity, poor economic performance, and even political violence (Weber, 1978; Besley and Persson, 2011).

How to overcome these challenges is a central question in political economy. Political leaders throughout history have worked to build new *states* through *nation*-building endeavors including the promotion of common values, identity, and language (often through public education), and the movement of people (often by force) (Tilly and Ardant, 1975; Smith, 1986; Hobsbawm, 1992).<sup>1</sup> However, such measures often require significant state capacity and infrastructure, the provision of which is hindered by diversity in the first place (Miguel and Gugerty, 2005). This dilemma poses a challenge to builders of new states. Leaders from early modern Europe to the post-colonial era have therefore relied on bundles of “homogenizing, territorializing, and mobilizing” activities (Smith, 1986).<sup>2</sup> It is an empirical question what the long-run results of these activities are.

It is difficult to identify the causal impact of such activities because they typically unfold over long time spans and eventually affect entire populations simultaneously. Some scholars have successfully addressed this empirical challenge by identifying causal effects of specific elements of states’ nation-building policies (e.g., Bazzi et al., 2019; Blouin and Mukand, 2019). In general, however, states do not implement such policies in a piecemeal manner. Moreover, potential interactions between different components of a given bundle of activities may lead to combined effects greater than the sum of their parts. Identifying such combined effects of a real-world bundled policy is the goal of this paper.

We aim to achieve this goal by studying a setting that provides sharp variation over time and within the population. Specifically, we investigate long-run effects of one of post-colonial Africa’s most ambitious policy experiments — the Tanzanian *Ujamaa* policy — on

---

<sup>1</sup> State-building is the construction of a state apparatus defined by its ability to establish a monopoly on the legitimate use of violence, to protect property rights, collect taxes, and to provide public goods in a given territory (Tilly and Ardant, 1975). Nation-building is a related concept, referring to the process leading to the formation of a national identity in which citizens feel a sufficient sense of emotional commonality that they wish to remain with each other. Nation-building may facilitate state-building (Alesina et al., 2021).

<sup>2</sup> Examples include efforts in the 19th-century to turn “peasants into Frenchmen” (Weber, 1976) and “make Italians” following that country’s unification (Duggan, 2008), as well as the early 20th-century construction of national identity in China (Wimmer, 2018) and the Soviet Union (Martin, 2001).

the development of a national identity and the consolidation of state authority in a highly diverse society.<sup>3</sup> We digitize historical administrative data and combine it with contemporary survey data to operationalize our empirical strategy.

While the political economy of state-building in diverse societies is globally relevant, it is particularly pertinent in the context of post-colonial Africa. Newly independent African states inherited artificial national borders (put in place by colonizers) that typically contained a multitude of ethnic groups with little history of centralized governance or strong shared identity (Gennaioli and Rainer, 2007; Alesina et al., 2011).<sup>4</sup> As a result, African post-independence leaders faced a number of challenges to the consolidation of state authority: a fragmented population scattered over inhospitable territories beyond the state's reach (Herbst, 2014), a mix of potentially hostile ethnic groups within arbitrary borders (Michalopoulos and Papaioannou, 2016), and no workable social contract between the state and population (Scott, 1998). These challenges have made it difficult for the new states to build capacity, expand public infrastructure, and develop common values around which to unite their diverse populations.

The *Ujamaa* policy, implemented from 1970 to 1981, aimed to address these challenges simultaneously by bringing the Tanzanian population into the state's reach and creating a national identity through the public education system. The key tool for expanding the Tanzanian state's reach was the so-called villagization program, which mandated much of the multi-ethnic rural population to settle in planned villages under central state administration. Villagization furthermore facilitated provision of infrastructure to dramatically increase access to public education with limited public investment.<sup>5</sup> These measures put education in the planned villages into the hands of the new central state, which quickly revamped the curriculum to reflect the state's political goals. The *Ujamaa* policy thereby allowed the state "to capture the peasantry" both geographically and ideologically (Hydén, 1980).

Exposure to the *Ujamaa* policy varied both across space and across age cohorts. Villagization intensity varied across districts since the implementation of the program was largely left to local officials, who implemented it with varying degrees of conviction. In addition, plausibly exogenous variation existed across age cohorts due to the policy's sharp timing and short-lived nature. Only the cohort young enough to enter school after the introduction of the policy was exposed to the new public education curriculum in the planned villages.

---

<sup>3</sup> *Ujamaa* roughly translates as "familyhood" (Mohiddin, 1968, fn. 1).

<sup>4</sup> Afrobarometer data reveals that over 50% of the African population identifies more with their ethnic identity than with the Tanzanian nation.

<sup>5</sup> Other elements of the *Ujamaa* policy included the abolition of individual titles to property, collectivization of agricultural production, and the nationalization of certain enterprises. These institutional reforms were critical to act as "equalizers" because unification of the nation could "not be completed or even advanced without an all-round development of the working population towards acquisition both of the new national identity and ideology" (Shengena, 1979, p. 175).

We are thus able to study impacts of *Ujamaa* using a difference-in-differences cohort design in the spirit of [Duflo \(2001\)](#). We neither assume that levels of villagization intensity varied randomly across space, nor that there were no underlying differences between different cohorts in the absence of villagization. Our empirical specification controls for systematic differences between districts and cohorts (with district and region-specific cohort fixed effects). Only the interaction between the two sources of variation constitutes exogenous variation in citizens' exposure to the *Ujamaa* policy under a parallel trends assumption. The sharp timing and short-lived nature of the policy allow us to compare differences in outcomes for cohorts that were of primary school age during the policy with cohorts that were slightly too young or slightly too old to be affected by the bundle of activities, for more and less intensely villagized areas. Our data also allows us to examine the development of outcomes across all cohorts before and after the policy, which we use to assess the plausibility of parallel trends.

In our first set of results, we find large, significant, and persistent positive effects of the *Ujamaa* policy on different measures of national identity. In surveys conducted around two decades after the policy ended, members of the primary school-age cohort in more intensely villagized districts are significantly more likely to report identifying primarily with the Tanzanian nation rather than with their ethnic group. The difference between average expressed national identity of the treated and the control cohort, for a district with complete villagization compared to a district without any villagization, corresponds to the difference between a respondent "feeling more Tanzanian" and "feeling more a member of her ethnic group [than Tanzanian]." In addition, we find that treated cohorts are significantly more likely to marry across ethnicity, suggesting a decreased salience of ethnic divisions. The latter revealed-preference evidence provides support for the idea that *Ujamaa* had real impacts on citizens' behavior and nation-building in the long term.

Several empirical concerns challenge our interpretation of this first set of results, including a violation of the parallel trends assumption, selective migration, and survey reporting bias. The results pass a battery of robustness checks. We find no differences in the trends across cohorts of primary schooling age before and after the policy period for more and less intensively villagized districts. In addition, our results are robust to controlling for various pre-policy district characteristics interacted with cohort fixed effects. We also conduct a bounding exercise which shows that selective migration is unlikely to explain a large part of the estimated effects. Finally, we control for respondents' beliefs about surveyors to reduce concerns related to social desirability bias.

Our finding that villagization bolstered national identity is perhaps surprising at first. In other, comparable contexts, the forced coexistence of diverse ethnic groups has led to backlash

and even spurred inter-group conflict.<sup>6</sup> The lack of backlash to *Ujamaa* may speak to the bundled nature of the policy, highlighting the importance of possibly *interacting* activities. Public education may have helped offset the potentially negative impact of bringing diverse groups together. At the same time, an education reform alone may not have been successful without infrastructure to implement it.<sup>7</sup>

While villagization alone may have affected national identity across all cohorts, only those of primary schooling age during *Ujamaa* were exposed to both villagization and the revised curriculum. Hence, our main coefficient can be interpreted as the differential effect of the bundled treatment compared to the effect of villagization alone.<sup>8</sup> Our results thus imply that education reform *together* with villagization was a sufficiently “big push” to bolster national identity. This is our core finding and object of interest. Answering the question of whether villagization or education alone had a bigger effect would require a different context and identification strategy.

Nevertheless, we conduct several exercises to examine potential channels underlying our main result. First, the *Ujamaa* policy may have primarily shaped the identities of students through public education provided in planned villages, or, second, through factors unrelated to education that differently affected the primary-school age cohort in the villages. We find evidence consistent with the first but little evidence for the second. In particular, we find that the *Ujamaa* policy increased national identity only among those who actually attended primary school. Furthermore, we provide evidence consistent with the *Ujamaa* policy affecting students’ attitudes through indoctrination and the use of a national language of instruction rather than through other channels such as occupational choice. We only find an effect of the *Ujamaa* policy on the cohort that was of primary school age during the policy period, but not on the cohorts that went to school after the policy ended, which suggests that changes in the content of education rather than a general increase in the supply of education explain our results.

Having established that the *Ujamaa* policy helped bolster national identity, we next examine whether it contributed to state-building by fostering state legitimacy in the long run. The answer to this question is ex-ante ambiguous and highly debated. On the one hand, the policy may have strengthened the acceptance of the state as a central authority

---

<sup>6</sup> Two notable examples in Africa are post-independence socialist governments of Mozambique and Ethiopia, which both ‘villagized’ 2 and 12 million people respectively (Lorgen, 2000). In both cases, forced coexistence eventually led to the downfall of the instigated political parties and intense conflict over long periods. In Ethiopia, it ended the Derg regime which brought to power the revolutionary TPLF, and in Mozambique, the socialist party Frelimo, which fought for independence from the Portuguese, was embroiled in a civil war from 1977-1992 with Renamo, the anti-communist political party.

<sup>7</sup> Note that the case of education without villagization is not observed, as very few had access to schooling outside of the planned villages.

<sup>8</sup> The effect of villagization alone appears to be small. Average levels of national identity are no higher in more intensively villagized districts for cohorts outside of primary schooling age during the *Ujamaa* policy.

through efforts to cut across societal cleavages and forge a common identity (Smith, 1986; Wimmer, 2018). On the other hand, forced exposure to other ethnic groups could have ignited internal conflict and weakened the legitimacy of the new state (e.g., Dippel, 2014).<sup>9</sup>

We find that cohorts exposed to *Ujamaa* are more likely to respect state authority and approve of one-party rule, and to have higher trust in central government institutions such as state media (in contrast to independent media). Interestingly, we find little effect on generalized inter-ethnic trust. These findings suggest that the *Ujamaa* policy first and foremost contributed to the new Tanzanian state’s ability to consolidate power. However, it arguably also led to lower demands for democratic accountability, individual autonomy, and diversity. This implies that comprehensive nation-building activities may come at a cost, and speaks to the “narrow corridor” characterizing the most successful states, in which a balance between state and societal interests has been achieved (Acemoglu and Robinson, 2020).

State-building in diverse societies is a critical issue for many nations as they face cleavages based on history, geography, and culture that can be antithetical to economic prosperity (see e.g., Alesina and La Ferrara, 2005; Ray and Esteban, 2017). However, policies aimed at strengthening the state often rely on some existing state capacity in the first place.<sup>10</sup> This dilemma could help explain why many developing, and especially post-colonial, countries have experienced great difficulty in maintaining strong states. Our paper aims to understand how states can overcome such challenges, and thus builds on a rich theoretical and historical literature on the processes of nation- and state-building (e.g., Breton, 1964; Tilly and Ardant, 1975; Smith, 1986; Anderson, 1991). A more recent economics literature has provided important empirical evidence in support of these ideas and theories. A first set of recent papers focuses on determinants of national identity that are not the direct result of state policies. In terms of context, our work is most closely related to Depetris-Chauvin et al. (2020), who investigate the short-term impact of national football teams’ victories on national identity in Sub-Saharan Africa.<sup>11</sup> We contribute to this literature by showing that leaders can actively shape national identity and foster state legitimacy through public policy in the long run.

---

<sup>9</sup> Other studies have examined effects of diversity and nation-building on conflict (e.g., Depetris-Chauvin et al., 2020) and public goods provision (e.g., Miguel, 2004). Our identification strategy would require data at the individual or cohort-district level to examine these outcomes, which is not available in our context.

<sup>10</sup> As Depetris-Chauvin et al. (2020, p. 1575) point out, nation-building policies “[...] require a substantial level of state capacity, and, as such, pose a clear conundrum: if internal fractionalization is an obstacle to the consolidation of a capable state, it can itself become a major hurdle for the adoption of policies that could overcome it. On the other end of the spectrum are other “softer” and less tangible nation-building tools involving the use of symbols, such as the national flag or the national anthem, meant to evoke and reinforce national identity. Yet, while such “banal nationalism” (Billig 1995) may help maintain a national identity that has already been established, it is unlikely to be strong enough to make one emerge”.

<sup>11</sup> Other contemporary work on nation-building examines how religious pilgrimage (Clingingsmith et al., 2009); war, occupation and repression (Dehdari and Gehring, 2022); foreign enemies (Dell and Querubin, 2018); and national leaders (Assouad, 2021) contribute to national identity.

A second set of papers carefully examines the causal effects of specific aspects of broader nation-building policies. [Bazzi et al. \(2019\)](#) use a population resettlement program in Indonesia to identify effects of intergroup contact on national integration.<sup>12</sup> Our paper differs from theirs in that we study the interaction between a bundle of measures aimed at bolstering national identity and consolidating state power. In contrast to their work, we find little evidence for inter-group contact as an important channel in our context. [Blouin and Mukand \(2019\)](#) focus on a different channel the state can use for nation-building, examining how propaganda broadcast over radio helped to change inter-ethnic attitudes in post-genocide Rwanda.<sup>13</sup> A further related paper is [Okunogbe \(2021\)](#), who studies intergroup contact and national integration in the context of national service in Nigeria.<sup>14</sup> Our study further contributes to this literature by demonstrating consequences of a comprehensive state-building policy in the context of a new state with challenging initial conditions.

Our paper also contributes to an emerging economics literature that studies the determinants of culture, values, and preferences. While economics has traditionally focused on the consequences of culture and identities (e.g., [Akerlof and Kranton, 2000](#); [Guiso et al., 2006](#)), scholars have more recently turned to investigating the endogenous formation of identity. This includes studies of the development of caste identity in India ([Atkin et al., 2021](#); [Cassan, 2015](#); [Cassan et al., 2021](#)), religious identity in medieval Europe ([Botticini and Eckstein, 2012](#)), ethnic identity in contemporary China ([Jia and Persson, 2013](#)), racial and ethnic identity in the United States ([Antman and Duncan, 2015](#); [Bandiera et al., 2018](#); [Dahis et al., 2019](#)), racial identity in contemporary Brazil ([Cornwell et al., 2014](#)) and Catalan identity in Spain ([Clots-Figueras and Masella, 2013](#)). Economists have also begun to study how education shapes beliefs, preferences, and political ideology ([Cantoni and Yuchtman, 2013](#); [Alesina and Reich, 2015](#); [Friedman et al., 2016](#); [Cantoni et al., 2017](#); [Bazzi et al., 2020](#)). We complement this literature by providing quantitative evidence that the salience of ethnic identity can be manipulated by state leaders, even in environments with extraordinarily challenging initial conditions and high levels of diversity.

Finally, we contribute to a literature that traces the impact of pre- and post-colonial institutions on African development. Much of this work examines the negative effects of

---

<sup>12</sup> As in newly independent Tanzania, the Indonesian government was concurrently engaged in a number of other activities oriented toward nation-building. These interventions included resettlement ([Bazzi et al., 2019](#)), school construction ([Duflo, 2001](#)), a nationalization of state media ([Pisani, 2014](#)), and other measures.

<sup>13</sup> This was part of a bundled policy attempt to rebuild post-genocide Rwanda, which included setting up a National Unity and Reconciliation Commission, national radio ([Paluck and Green, 2009](#)); revising constitution to criminalize divisionism ([Beswick, 2010](#)); memorial sites to promote collective memory, annual commemorations during national mourning week, new national symbols, administrative restructuring and change of place names, rewriting history textbooks; and introducing of social and solidarity programs ([Thomson, 2013](#); [Straus and Waldorf, 2011](#)).

<sup>14</sup> The paper finds participants who were relocated outside their ethnic region, seven years after program completion are more likely to live outside their own region, know Nigeria more, more likely to be in inter-ethnic relationships, and express greater national pride compared to those who served in their own region.

artificial colonial border design (Easterly and Levine, 1997; Alesina et al., 2011; Michalopoulos and Papaioannou, 2016), particularly the resulting ethno-linguistic fractionalization.

Much less attention has been given to national institutions that aim at overcoming the problems resulting from colonization (Michalopoulos and Papaioannou, 2013). We make a contribution here while also pointing to potential costs associated with such institutions. In this context, our paper is one of the first to empirically examine the political-economic consequences of the *Ujamaa* policy. This gap in the literature is due to empirical challenges rather than a lack of scholarly interest, though scholars have focused on *Ujamaa*'s largely negative impacts on economic development (Hydén, 1980; Puttermann, 1986; Collier et al., 1986; Scott, 1998). An important exception is thesis work by Osafo-Kwaako (2012), who examines the impact of *Ujamaa* on various contemporary social and economic outcomes.<sup>15</sup> Miguel (2004)'s influential article, which examines how nation-building in Tanzania and neighboring Kenya has affected inter-ethnic cooperation, is also closely related to our work.<sup>16</sup>

This paper proceeds in seven sections. Section 2 describes the *Ujamaa* policy and contextual background. Section 3 presents the data we use to measure exposure to the *Ujamaa* policy and its consequences. Section 4 outlines our empirical strategy to address challenges to identification. Section 5 presents our main results on national identity, conducts several robustness checks on our core findings, and considers alternative channels. Section 6 discusses implications of our findings on views of the state and the legitimacy of state authority. Section 7 concludes.

## 2. Background

*Ujamaa* included a series of reforms in post-independence Tanzania. These included the institutionalization of one-party rule and equality across all spheres of society (including the economy),<sup>17</sup> and — most important for the present study — resettling the rural population into planned villages and harnessing public education as a nation-building tool. All these

---

<sup>15</sup> Osafo-Kwaako (2012) employs a cohort design and cross-sectional instrumental variables strategy to examine socio-economic and political-economic consequences of *Ujamaa*. He finds positive effects on schooling and public goods provision, but negative effects on consumption and perceptions of corruption. Our work focuses on state-building; in contrast to Osafo-Kwaako (2012), we find that *Ujamaa* strengthened national identity and support for a strong state.

<sup>16</sup> Miguel (2004) makes a cross-sectional comparison of two nearby districts in Tanzania and Kenya which are separated by a national border but are similar in terms of local ethnic diversity. The study finds a negative relationship between ethnic diversity and public goods in Kenya, but a positive relationship in Tanzania. This is explained by Tanzania having implemented effective nation building-policies to ameliorate ethnic divisions, whereas Kenya did not. Our study focuses on the case of Tanzania and finds direct evidence in support of this explanation. We further add evidence on changed attitudes towards the state, whereas Miguel (2004) focuses on public goods provision and inter-ethnic relations.

<sup>17</sup> The institutionalization of equality was manifested in socialist reforms, including the nationalization of key commercial industries and collectivization in agriculture.

activities arguably aimed at organizing the country around the new state (Nyerere, 1969a,b). The villagization program – implemented primarily between 1970-1981 – mandated that the rural population live in government-administered villages (Scott, 1998). Moreover, in those villages, the state expanded access to public services – especially primary schools, which taught children a new political education curriculum centered around building a national identity and establishing the Tanzanian state as a legitimate authority (Nyerere, 1982). We now discuss key elements of the policy. Appendix Table A.1 summarizes key events.

When Tanzania gained independence from Britain in 1961, Prime Minister Julius K. Nyerere encountered a challenge common to many new leaders of post-colonial states: how to unify diverse people, spread across a large territory, around a common identity and state. Nyerere’s challenge was particularly acute: the country’s population comprised over 120 ethnic groups with different languages or dialects (Omari, 1995), making Tanzania one of the world’s most diverse countries (Hudson and Taylor, 1972; Fearon, 2003; Alesina et al., 2003). Furthermore, much of the country was sparsely populated; Herbst (2014, p. 151) describes Tanzania as a “rim-land” country, in which “the less populated areas are within a circle partially defined by the more populated areas,” presenting a stark geographical challenge to any would-be state builder.<sup>18</sup>

Nyerere developed an ambitious agenda to unite such a diverse and geographically dispersed population, embarking on what Scott (1998) describes as one of the most ambitious nation-building programs undertaken in post-colonial Africa. In 1963, shortly after independence, Nyerere consolidated political power and declared the governing Tanganyika African National Union (TANU) as the sole legal party (Tordoff, 1967), a measure that was understood as a means of reducing societal divisions (Tripp, 1999).<sup>19</sup> Nyerere subsequently outlined his ideas for the Tanzanian state in his 1967 landmark speech, the *Arusha Declaration*. The speech presented the vision behind the *Ujamaa* policy, which included the central role of villages for development, and the reduction of divisions across groups.<sup>20</sup>

The villagization program aimed at bringing the rural population into the reach of the state and providing access to public goods, most importantly schools. Since the state in part mandated the population to provide these public goods themselves (including the construction of schools in many villages), the program allowed for the provision of access to public schools despite limited existing infrastructure and public funds.

---

<sup>18</sup> As of the 1967 Census, 12 million inhabitants were spread over nearly one million square kilometers, with nearly 95 percent of the population residing in rural areas (Wenban-Smith, 2014).

<sup>19</sup> Nyerere was not alone in this regard. As Bandyopadhyay and Green (2013, p. 111) note, “in several states the creation of a one-party system allowed leaders more space to pursue nation-building policies” in the early independence period. These include the Democratic Republic of Congo, Cameroon, Ghana, and Zambia, among others.

<sup>20</sup> The speech also outlined various reforms aimed at reforming the economy in line with socialist principles, such as economic self-reliance, nationalizing commercial farms and businesses, and creating co-operatives rather than individual family farms to avoid class formation.

The mandatory villagization period began after 1969, with no guidelines for implementation apart from declaring 1976 as the target year for full rural villagization.<sup>21</sup> Implementation was largely left to district officials, who pursued the endeavor with different degrees of conviction (McHenry, 1979) through a series of “Operations” (Von Freyhold, 1979). In 1975, the *Village and Ujamaa Village Act* was promulgated, mandating registration of villages with 250 or more households as legal entities (Bryceson, 1982). As a result of the villagization program, over 80 percent of the country’s rural population lived in registered villages by the end of 1976 (McHenry, 1979; Shao, 1982).<sup>22</sup> Figure A.1 depicts the rapid growth of the population in registered villages over the period after independence.

Concurrently, the state centralized and consolidated the education sector. It was clear to Nyerere that to create a strong national identity, the education system had to be part of the process (Miguel, 2004). As the Minister of Education Solomon Eliufoo put it in 1968, education was “a cementing matrix of *Ujamaa* and progress [to establish national unity]” (Peeples, 2018, p. 46).

The *Ujamaa* education reforms were outlined in Nyerere’s second key declaration of 1967, the *Education for Self-Reliance* (ESR) paper, and consisted of two key elements: bringing the education sector into the hands of the state and changing the school curriculum to help foster nation-building. To execute the ESR, the government nationalized all schools with the Education Act of 1969 (Samoff, 1991). While only 7 percent of the population had completed some public education by 1967, data from the 1978 census showed primary enrollment at almost 90 percent (Government of Tanzania, 1984). This was likely largely due to an increase in the supply of schooling in the registered villages. In unreported results, we find that school-age cohorts in more villagized districts are more likely to have obtained primary schooling and that the gross primary enrollment rate rose from less than 50% in the late 1960s to above 80% by the early 1980s (National Census data).<sup>23</sup>

Beyond increasing the supply of education, the Ministerial Circular of 1968 introduced Political Education (known in Swahili as *Elimu ya Siasa*) as a new subject to the primary school curriculum for students enrolled in Standards IV-VII (equivalent to 4<sup>th</sup> grade onwards). The introduction of Political Education meant new syllabi and new textbooks, and also impacted teaching in other social studies (history, geography, and civics).<sup>24</sup> Political education primarily served as a tool for ideological indoctrination, emphasizing the importance of the nation and the state over tribal identities (Komba, 1996). In addition, Education Circular No. 2 of 1967 mandated Swahili as the national language of instruction in all public schools

<sup>21</sup> Presidential Circular No. 1 (Mung’ong’o 1995:80-1 in Kikula, 1997).

<sup>22</sup> The average distance people moved was rather small - usually within eight kilometers (Thomas, 1982).

<sup>23</sup> Data on historical school construction is unavailable.

<sup>24</sup> In villages where the Political Education syllabus could not be obtained, teachers were told to use party documents — in particular, the *Arusha Declaration* and the ESR policy paper (Komba, 1996).

by November 1969. Figure A.2 depicts excerpts from a Swahili textbook of the period, illustrating how nationalist themes were interwoven throughout even “apolitical” subjects.

Analysis of educational materials during the *Ujamaa* period displays explicit emphasis on indoctrinating students on national identity centered around the state. As Komba (1996, p. 111-112) notes, “the dominant theme was, obviously, nationalism” and “the general tendency was in the direction of political indoctrination rather than genuine Political Education”. Figure A.3 enumerates the contents of a typical political education textbook, highlighting the themes that speak to nation-building that Komba (1996) flags.

In sum, whereas the pre-*Ujamaa* period was characterized by ethnic divergences in educational access, the nationalization of schools broke down regional and ethnic differences.<sup>25</sup> Beginning in 1970, several school cohorts were thus exposed to the government’s nation-building program, delivered through public schools in the state-administered villages.

However, perhaps as a consequence of the scope of its ambition, the *Ujamaa* policy was fairly short-lived. In part due to the collective farming that villagization entailed, the country saw a considerable decline in agricultural productivity (Hydén, 1980). Tanzania’s economy had entered “deep, serious crisis” by the early 1980s (Bagachwa and Naho, 1995, p. 1391), at which point the villagization initiative was largely halted. In 1982, the *Villages and Ujamaa Villages Act* was repealed, thus marking the end of the mandatory villagization period. While Political Education remained part of the official curriculum until 1992, the zeal with which it was implemented waned following the demise of villagization. As we show in what follows, this policy experiment – though spanning little more than a decade (1970 - 1981) – had a profound and lasting impact on the hearts and minds of Tanzanian citizens, particularly those exposed to public education in planned villages during the *Ujamaa* period.

### 3. Data

Our sample includes individuals responding to contemporary surveys, from which we obtain outcomes and birth dates to capture variation in their exposure to the *Ujamaa* policy over time. Using information on respondents’ place of residence, we link this sample to historical census data to capture variation in their exposure to villagization over space. Table 1 provides descriptive statistics. Appendix Tables A.2 and A.3 provide an overview

<sup>25</sup> As Court and Kinyanjui (1980, p. 47) note, “irrespective of where the school is located, students can be allocated to them from anywhere in the country. This itself is a measure of national integration.” Furthermore, establishing Swahili as the language of instruction reflected “the belief that Swahili was meant to be a language for national integration” (Komba, 1996, p. 116). Simpson (2008, p. 18) further notes, “the development and spread of Swahili in national and official language roles is generally regarded as having had very positive effects on nation-building in the post-independence era, and was extremely important for the creation of a sense of allegiance and belonging to a single, new Tanzanian nation”. Even English-medium Tanzanian textbooks under villagization “[de-emphasized] images of ‘ethnic or tribal’ identities and [focused] only on creating new identities and loyalties based on the nation-state” (Mbuyi, 1987, p. 24).

and description of the variables used in the paper.<sup>26</sup>

### 3.1. Historical District-level Data

We use newly digitized data from the 1978 population census ([Bureau of Statistics, Ministry of Planning and Economic Affairs, 1981](#)) to measure the historical extent of villagization across space.<sup>27</sup> The historical extent in an individual's district of residence proxies for their exposure to villagization. While we have also obtained data on villagization from the historical population censuses at finer geographical units, it cannot be linked to outcomes from contemporary surveys, which identify respondents at the district level at most.<sup>28</sup> Figure A.4 in the Appendix shows an excerpt of the 1978 census data. We measure villagization as the share of a district's rural population living in registered government villages in 1978:

$$V_d = \frac{P_{d,1978}^{\text{registered}}}{P_{d,1978}^{\text{rural}}}$$

where  $P_{d,1978}^{\text{registered}}$  denotes the number of individuals living in registered villages in district  $d$  and  $P_{d,1978}^{\text{rural}}$  denotes total rural district population in district  $d$  in 1978. We aggregate districts to their 1967 boundaries since this is the unit of variation of our controls.

As shown in Table 1 Panel A, the villagization measure has a mean of 0.93, and significant dispersion with a minimum of 0.50 and a maximum of 1. The average extent of villagization is relatively high because some geographic zones of the country were almost completely villagized.<sup>29</sup> However, there is considerable within-zone variation in other areas, which we exploit in our empirical analysis. Figure A.5 illustrates the variation.

The variation in the extent of villagization across space can largely be explained by the fact that its implementation was left to district officials (a further reason for our focus on the district as a unit of analysis), who enforced the policy with different degrees of conviction ([McHenry, 1979](#)). This variation may not be random across space (nor does it need to be for our empirical strategy to be valid, as we explain in Section 4). Historical accounts point to several factors that could also explain some of the observed variation in villagization. These

---

<sup>26</sup> Due to the rural nature of the policy, we exclude the capital Dar es Salaam as well as the semi-autonomous islands (Zanzibar, Pemba, Mafia) from our sample. The islands' governments have separate authority over a number of government functions, including education.

<sup>27</sup> We are grateful to Philip Osafa-Kwaako for generously sharing part of his historical data, and to the Herskovits Library of African Studies at Northwestern University for their assistance.

<sup>28</sup> We do not observe respondents' district of birth or district of residence during the *Ujamaa* policy. We discuss in detail further below how our using the current district of residence affects the interpretation of our estimates and potential threats to identification, including selective migration.

<sup>29</sup> Tanzania has two main subnational administrative units: the region and the district. A zone is a larger subnational geographic area that is not an official administrative unit but is commonly used by organizations such as the Demographic and Health Surveys (DHS) and the Tanzania Ministry of Health. A zone contains three regions or six districts on average.

factors, which we account for in our analysis, include the pre-*Ujamaa* degree of urbanization, primary school enrollment rate, local government capacity, ethnolinguistic fractionalization, geographic characteristics, availability of public infrastructure, and weather (Hydén, 1980). The baseline characteristics variables are mostly digitized from historical district-level government statistics (Jensen, 1968). We discuss these variables and their relationships with villagization in Appendix A.

### 3.2. Individual-level Survey Data

Our primary data on long-run outcomes and other individual-level characteristics is from the geo-coded Afrobarometer public opinion survey rounds 3 and 4 (2005 and 2008) (Afrobarometer, 2017; BenYishay et al., 2017).<sup>30</sup> The Afrobarometer surveys are widely used, nationally representative surveys conducted by a pan-African research institution unaffiliated with any national government. The survey rounds we use comprise a sample of over 1,400 individual respondents from across Tanzania. We focus on respondents born in 1948-1983, which ensures that the individuals in the sample would have been able to complete their primary education. We use information on individuals' districts as reported in the survey to match them to our district-level data. Our focal dependent variable captures national identity and is based on the following question:

*“Let us suppose that you had to choose between being a Tanzanian and being a [respondent’s previously reported ethnic group]. Which of the following statements best expresses your feelings?”*

Respondents could report that they feel *only/more/equally/less/not at all Tanzanian* as compared to their ethnic group. Our baseline outcome variable is coded on a 0 to 1 scale with quarterly increments, where 0 indicates that the respondent identifies only with her ethnic group and 1 indicates that she identifies exclusively with Tanzania. We interpret a higher score as reflecting a stronger national identity. This measure of national identity is standard in the literature (see, e.g., Depetris-Chauvin et al., 2020).

We use several other variables from the Afrobarometer survey data to measure outcomes, such as respondents' views of the state and trust in the media.<sup>31</sup> We describe these variables as we introduce them for the empirical analysis in Section 6. We also use respondents' birth years from the Afrobarometer survey to measure their temporal exposure to the *Ujamaa* policy. Finally, we construct a dummy variable for whether the respondent has completed at least some formal primary schooling, which we use in complementary analyses.

---

<sup>30</sup> Our choice of these rounds reflects data availability (key questions were not included or are incomparable in their wording in earlier and some later rounds) and time since the villagization policy ended.

<sup>31</sup> Note that the sample size for analysis of some of these questions is smaller because they were only asked in one rather than both Afrobarometer rounds used in this paper.

Table 1, Panel B shows individual-level descriptive statistics for the baseline sample used in our analysis. The mean measure of national identity is 0.89. National identity is high on average in Tanzania compared to the rest of Africa, presumably in part due to the *Ujamaa* policy.<sup>32</sup> The average birth year of respondents in our sample is 1962. Most of the respondents (89%) completed some schooling.

### 3.3. Other Data

We incorporate data from various other sources for additional analyses and robustness checks. These include Round 1 of the Tanzania National Panel Survey in 2008/2009 (TNPS) (National Bureau of Statistics, 2010), which is part of the World Bank's Living Standard Measurement Surveys (LSMS) and includes several thousand respondents from across Tanzania. We use this dataset in robustness checks related to migration.<sup>33</sup> We also use data on individuals' occupations from the TNPS to investigate channels. The majority of the respondents work in agriculture (78%).

We also use data from the Demographic and Health Surveys (DHS) in 1991 and 1996 (Ngallaba et al., 1993; Bureau of Statistics/Tanzania and Macro International, 1997) to study intermarriage as a revealed-preference measure of national identity (see, e.g., Bazzi et al., 2019). These two survey waves report married couples' ethnic affiliations (later waves do not report ethnicity). We code a dummy that equals 1 if a male respondent shares his wife's ethnic group, and 0 otherwise. In our sample, close to 70% of marriages occur within the same ethnic group.

Finally, to examine the role of the national language of instruction in public schools, we compute each district's weighted average linguistic distance to Swahili (the national language) based on the district's ethnic group shares in 1967. Data on linguistic distances between ethnic groups are from the LEDA open-source software package (Müller-Crepon et al., 2022).<sup>34</sup> The mean district linguistic distance to Swahili in our sample is 0.21. A zero on the LEDA scale indicates zero linguistic distance, i.e. no difference to Swahili, while a one indicates no commonality between the language and Swahili.

## 4. Empirical Strategy

This section outlines our empirical strategy to estimate effects of the *Ujamaa* policy, which included villagization and a public education reform, on national identity and attitudes

<sup>32</sup> Tanzanians are ranked the third-highest on the continent for their feelings of national identity after Burundi and Guinea. The lowest national identity is held by Nigeria at 0.54 (Afrobarometer data).

<sup>33</sup> The Afrobarometer data do not contain information about migration history or birth district.

<sup>34</sup> For additional details on LEDA, refer to <https://github.com/carl-mc/LEDA>, accessed July 2021.

towards the state in the long run. The ideal experiment would randomly assign *Ujamaa* to some individuals but not to others, and then compare their outcomes. However, the policy, like most reforms throughout history and around the globe, was not carried out in such a manner and thus there may be joint determinants of villagization, access to education, and outcomes of interest. Our approach follows the logic of a difference-in-differences specification to address such confounders. The first difference comes from varying intensity in villagization across individuals' districts. The second difference comes from plausibly exogenous variation in the temporal exposure of different age cohorts induced by the timing of the *Ujamaa* policy. We discuss our identifying assumptions and potential threats to identification after first introducing our specification.

Our empirical strategy relies on the fact that the date of birth and villagization intensity in an individual's district jointly determine her exposure to *Ujamaa*. Individuals born in 1959 or earlier were older than the official primary schooling age when the *Ujamaa* policy took effect with the start of villagization in 1970. They should thus not be affected by the full "treatment." In contrast, individuals born between 1960 and 1971 were old enough to be exposed to the government's new political education curriculum under villagization, which was taught from 4<sup>th</sup> grade. They were also young enough to attend 4<sup>th</sup> grade before the villagization period officially ended in 1982. In the main specification, our "control" cohort was thus born in 1948–1959, and our "treated" cohort was born in 1960–1971. In complementary analyses, we also consider other cohorts. The two cohorts in our baseline sample consist of 862 individuals in 54 districts.

#### **4.1. Identifying Effects of *Ujamaa***

For our main results, we estimate the following specification:

$$y_{idzt} = \beta(V_{dz} \cdot \text{treatedcohort}_t) + (\mathbf{X}'_{dz} \cdot \text{treatedcohort}_t)\Gamma + \alpha_{dz} + \delta_{zt} + \epsilon_{idzt} \quad (1)$$

where  $y_{idzt}$  is an outcome of individual  $i$  in district  $d$ , zone  $z$ , and schooling cohort  $t$ .  $V_{dz}$  is our district-level measure of historical villagization as described in Section 3, and  $\text{treatedcohort}_t$  is a dummy that equals 1 if individual  $i$  was born in 1960–1971 (0 if born in 1948–1959).  $\alpha_{dz}$  denotes district fixed effects (which also include zone fixed effects),  $\delta_{zt}$  denotes zone-cohort fixed effects and  $\mathbf{X}'_{dz}$  is a vector of controls.<sup>35</sup> We cluster standard errors at the district.

Our coefficient of interest is on the first interaction term. Intuitively,  $\beta$  captures the average difference in outcomes between high- and low-villagization districts for individuals

---

<sup>35</sup> Our preferred specification includes the district-level primary schooling rate and share of population living in urban areas in 1967, each interacted with the cohort dummy, as control variables. We control for other variables in robustness checks. See Results Section 5.

of schooling age during the *Ujamaa* period.<sup>36</sup> Since we do not directly observe individual exposure to villagization but villagization intensity at the district level, we interpret  $\beta$  as an intent-to-treat effect, which is likely smaller than a treatment effect on the treated.

The baseline specification is appealing due to its simplicity and because pooling individuals into two cohorts increases statistical power. However, we also estimate a specification that allows treatment effects by cohorts to vary more flexibly over time. The main advantage of this second specification is that it allows us to assess pre-trends. In addition, we can test whether treatment effects are stronger for those that were of primary schooling age during the entire villagization period than for those with only partial temporal overlap in exposure to the *Ujamaa* policy. The flexible specification is as follows:

$$y_{idzt} = \sum_{t=1}^9 \beta_t (V_{dz} \cdot cohort_t) + \sum_{t=1}^9 (\mathbf{X}'_{dz} \cdot cohort_t) \boldsymbol{\Gamma}_t + \alpha_{dz} + \delta_{zt} + \epsilon_{idzt} \quad (2)$$

where  $cohort_t$  is a dummy that indicates whether individual  $i$  belongs to cohort  $t$ . We divide our sample into nine 4-year cohorts (born in 1948-1951, 1952-1955, ..., 1980-1983). The cohort born between 1956 and 1959, the last school cohort before the policy took effect, is the omitted category. We interpret each of the parameters  $\beta_t$  as the impact of the *Ujamaa* policy on cohort  $t$ . Since villagization lasted from 1970 to 1981, we expect the coefficients for the cohorts born between 1960 and 1971 to be greater than 0 and the coefficients for the cohorts born before 1960 to be equal to 0. The coefficients for cohorts born after 1971 are somewhat ambiguous ex ante but we expect them to decrease over time.<sup>37</sup>

#### *4.2. Assumptions and Threats to Identification*

The difference-in-differences estimate  $\beta$  in equation (1) can be interpreted as a causal ITT effect of the *Ujamaa* policy under the parallel trends assumption that, in the absence of the policy, the changes in the outcome variables across cohorts would not have been systematically different in low and high villagization districts within a zone. We now discuss the most important potential challenges to this assumption.<sup>38</sup>

---

<sup>36</sup> In the baseline, we use a continuous measure of villagization rather than a high-/low-villagization dummy.

<sup>37</sup> The emphasis on nation-building was removed from the official curriculum in 1992 (and received considerably less emphasis after villagization was officially abandoned in 1982). However, those who remained in the registered villages still enjoyed better access to public schooling than those in other areas, and the textbooks used in schools likely remained the same for a few years given a lack of alternatives.

<sup>38</sup> We are not concerned about the identification challenges related to difference-in-differences settings with staggered treatment highlighted in the recent econometrics literature (e.g., De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). The treatment in our main specification is not staggered, and we do not estimate dynamic treatment effects.

**Level differences across districts.** Districts with lower pre-existing levels of national identity may have implemented the villagization policy less intensively. Alternatively, the government may have enforced the *Ujamaa* policy to a higher degree in districts with lower pre-existing levels of national identity. In addition, there may be other joint determinants of villagization and levels of national identity across districts, as discussed in [Appendix A](#). Such level differences, insofar they are constant across cohorts, are subsumed by district fixed effects and are not a threat to identification in our baseline specification.

**National or regional policies and other differences between cohorts.** The timing of the *Ujamaa* policy may have been correlated with other factors that shaped national identity across the country, such as macroeconomic developments, trends in education, or regional policies. Such factors, insofar as they affected the country or the districts within a sub-national geographic zone similarly, are subsumed by cohort-zone fixed effects. Similarly, general differences in national identity between cohorts (for example due to age effects), insofar they are constant within zones, are not a concern given the cohort fixed effects.

**Different trends over cohorts across districts.** The development of national identity may have followed different pre-existing trends over cohorts across districts with differential intensity of villagization. For example, school cohorts in more remote districts, that were treated less (or more) intensively due to their distance from government institutions, may have had lower levels of national identity initially, and would have caught up with other districts post-independence even in the absence of the villagization policy. If not accounted for, such trend differences could violate the parallel trends assumption and thus pose a threat to identification (as opposed to level differences). The descriptive evidence in [Figure 1](#) shows that this is unlikely to be a major concern, as low- and high-intensity districts follow parallel trends over cohorts that were of primary schooling age pre- and post-policy. We also look for signs of differential pre-trends more systematically in [Section 5](#).

In addition, we interact the cohort fixed effects with numerous pre-*Ujamaa* district characteristics that may be correlated with the intensity of villagization and that may affect national identity differently for different cohorts in the long run (see [Appendix A](#)). Our baseline includes the share of urban district population and primary school enrollment rate in 1967. In a battery of robustness checks, we also include a number of other controls in 1967 interacted with cohort fixed effects, such as district government revenue to measure local state capacity, distance to the capital, ethnolinguistic fractionalization, a host of geographic characteristics, public health infrastructure, and weather shocks. Additional details are in [Section 5.4](#).

**Migration.** We observe respondents' districts at the time the contemporary surveys were

conducted (1991–2008) but not at the time of villagization (1970–1981). If a respondent lived in a different district when they were of primary schooling age vs. when they responded to a survey, this could bias the estimate of our coefficient of interest in different directions, depending on what determines migration. However, not all forms of migration threaten the validity of our results. First, note that within-district migration does not affect our results as our measure of villagization varies at the district level. Second, if migration across districts is uncorrelated with villagization or national identity, this is akin to classical measurement error in the villagization variable, biasing our estimate towards zero. In that case, our coefficient estimate would be a lower bound on the true effect of the *Ujamaa* policy on outcomes. The kind of migration we are more concerned about is selective. For example, those who were receptive to public education during the *Ujamaa* period may have stayed in the registered villages after the policy ended. At the same time, those opposed to the government’s ideas may have moved to districts that had been less affected by the policy. Such selective migration would bias our estimate upwards. In Section 5.4, we provide several pieces of evidence to show that selective migration is unlikely to explain much of our results.

## 5. Results: Ujamaa and National Identity

### 5.1. Descriptive Evidence

Before discussing our main results, we show descriptive evidence on the levels of national identity by age cohort and intensity of villagization. Figure 1 plots the mean of our main measure of national identity in 2005/2008 by age cohort, relative to the level of the cohort that was of schooling age just before villagization (born 1956–1959), for high- and low-villagization districts separately (above and below mean villagization in the sample).<sup>39</sup>

We see no systematic differences across cohorts for low-villagization districts, where the *Ujamaa* policy was implemented to a lesser degree. In contrast, in high-villagization districts, national identity is higher on average for the cohorts that were of schooling age during villagization compared to the younger and older cohorts. Reassuringly, mean national identity of cohorts that were too old or too young during villagization (born 1948–1959 or 1972–1983) moves in parallel for high- and low-villagization districts, but diverges for cohorts that were of schooling age during villagization (those born 1960–1971).

### 5.2. National Identity: Difference-in-differences Estimates

Table 2 shows our main difference-in-differences estimates of the ITT effect of the *Ujamaa* policy in 1970–1981 on national identity in 2005–2008 (equation (1)). Column (1) shows

---

<sup>39</sup> For ease of interpretation, we standardize the measure of national identity to have a mean of 0 and a standard deviation of 1. We group respondents in 4-year cohorts to reduce noise.

that national identity is on average 0.18 standard deviations higher for the treated cohort compared to the control cohort for every 1 standard deviation increase in villagization. Column (2), our preferred specification, additionally controls for zone-cohort fixed effects to pick up any differential trends across larger sub-national geographic regions. Our main result in column (2) is that a one standard deviation increase in exposure to the *Ujamaa* policy increases national identity around two decades later by 0.24 standard deviations. The effect is statistically significant ( $p\text{-value} < 0.05$ ). The difference between average expressed national identity of the treated and the control cohort, for a district with complete villagization compared to a district without any villagization, corresponds to the difference between a respondent “feeling more Tanzanian” and “feeling more a member of her ethnic group [than Tanzanian]”.

In columns (3) to (8), we additionally control for the 1967 district characteristics indicated in the column heads interacted with the cohort dummy. These district characteristics are the potential correlates with villagization discussed in [Appendix A](#) (distance to capital, district revenue, ethnolinguistic fractionalization, geographical features, hospital beds, and weather shocks). All coefficients on the interaction between villagization and the cohort dummy remain qualitatively similar and are significant at the 5% or 10% level.<sup>40</sup>

### 5.3. Parallel Trends

[Figure 2](#) displays the regression coefficients of our flexible specification including all cohorts (equation (2)). Each coefficient shows the differential effect of a one standard deviation increase in villagization on national identity for the birth cohort indicated on the x-axis compared to the reference cohort (born in 1956–1959). In line with our main results, we find positive effects of villagization on national identity for the cohorts that were of primary-school age during the policy (largest and significant at the 5% level for the cohort born in 1968–1971, i.e., in 4th grade age at the height of villagization) but not for older cohorts. We find a positive, albeit statistically insignificant effect on the cohort that entered primary-school age right after villagization officially ended. This is not surprising given that the schools established in the registered villages remained operational after the policy ended. We note that for younger cohorts, i.e. those born in 1976–1983 and of primary school age long after villagization ended, there is no differential effect of villagization on national identity.

[Figure 2](#) provides support for the parallel trends assumption. We see no differential pre-trends in national identity between high- and low-villagization districts over cohorts that were of primary schooling age before the policy. Nor do we see differential trends for cohorts of primary schooling age many years after the policy ended. If our main results were explained

---

<sup>40</sup> Note that we cannot control for all district characteristics simultaneously, each interacted with cohort fixed effects, due to a lack of statistical power.

by unobserved factors correlated with villagization, they would need to have different effects for different cohorts following this specific pattern.

Appendix Table A.4 column (2) reports the point estimates and standard errors displayed in Figure 2. In addition, we re-estimate this specification, now controlling for the same variables as in our main Table 2, all interacted with cohort fixed effects. The coefficients of interest are qualitatively similar across all columns. The interaction for the birth cohort of 1968–1971, which was of schooling age at the height of villagization, is statistically significant at the 1% or 5% level in all columns except column (1) (significant at the 10% level), which includes only cohort fixed effects instead of cohort-zone fixed effects.

#### 5.4. Robustness

We address several other potential concerns regarding our empirical strategy by subjecting our results to various additional robustness checks. We return to our baseline specification in equation (1), which pools the treated cohorts born in 1960–1971 and compares them to the pooled cohorts born in 1948–1959. Table 3 shows the results of these robustness checks for our preferred specification, analogous to our preferred specification in Table 2 column (2). In the Appendix, we also report the results of these robustness checks subject to the same controls (interacted with the cohort dummy) as in Table 2 columns (3) to (8).

**Sample weighting.** In our preferred specification, we used survey weights from Afrobarometer to make our sample nationally representative. Table 3 column (2) and Appendix Table A.5 show the unweighted coefficients. The estimates from the weighted and unweighted regressions are very similar in size and significance.

**Social desirability bias.** Our main outcome measure, whether the respondent reports identifying more with her ethnic group or with the nation as a whole, may suffer from social desirability bias. The treated respondents were taught the importance of the nation and a sense of duty towards the state as part of the *Ujamaa* policy. Hence, treated respondents may believe that they are expected to answer in a certain way because they may believe that the surveyor is a representative of the national government. This could bias our coefficient of interest upwards. Afrobarometer spends considerable effort to mitigate such concerns by hiring independent surveyors unaffiliated with the government and from outside the survey districts. Nevertheless, almost 60% of the respondents in our sample report thinking that the surveyors were sent by a government institution.

Social desirability bias is unlikely to explain our results for at least two reasons. First, we directly control for whether a respondent believes that the surveyor was sent by a

government institution. Our estimates are unchanged when including this control variable (see Table 3 column (3) and Appendix Table A.6). Second, our results on inter-ethnic marriage (discussed below), which we view as a “revealed preference” measure of identity that should not suffer from potential reporting bias, are very consistent with the estimated effect on the main measure of national identity from the Afrobarometer.

**Sample choice.** Since villagization was a rural policy, those living in rural areas during the *Ujamaa* period should be affected by the treatment (the “compliers”).<sup>41</sup> Table 3 column (4) and Table A.7 report our coefficient estimates for the sample of respondents living in a rural area at the time of the survey. As expected, we find that the effect of *Ujamaa* on national identity is larger for this sub-sample (as they are more likely to be compliers).

**Migration.** Migration during the time between villagization and the survey could bias our estimates since we only observe the current district of respondents at the time of survey and not where they lived during the villagization period. However, as discussed in Section 4.2, not all types of migration are equally threatening to the interpretation of our results. The type of migration we are more concerned about is selective. In Appendix B, we conduct two exercises to assess robustness of our findings to potential selective migration. Table 3 column (5) drops all districts with either an in- or an out-migration rate in the highest deciles, with similar results to our baseline. As we explain in Appendix B, Columns (6) and (7) show that even under strong assumptions working against our findings, potential selective migration is unlikely to explain the majority of the positive relationship between the *Ujamaa* policy and national identity we find.

### 5.5. Intermarriage

So far, we have presented evidence suggesting that the *Ujamaa* policy bolstered self-reported national identity in the long run. We now examine whether this translates into real-world outcomes, examining the effect of the *Ujamaa* policy on ethnic intermarriage.

We follow the same empirical approach as previously but construct a new outcome variable using data from the DHS in 1991 and 1996.<sup>42</sup> We construct an outcome variable that equals 1 if both partners belong to the same ethnic group. The treatment variable is based on the

---

<sup>41</sup> Recall that prior to villagization, less than 5% of the Tanzanian population was living in urban areas.

<sup>42</sup> These two survey waves record the ethnic group of each respondent and identifiers that allow us to link couples. Subsequent waves did not ask about ethnicity. We include all married couples in the sample.

husband's current district and birth year.<sup>43</sup>

Table 4 presents the results. We find that a one standard deviation increase in villagization is associated with a 7.3 percentage point (0.17 standard deviations) decrease in the likelihood of partners sharing the same ethnic group for cohorts of primary schooling age during the villagization period, compared to pre-villagization cohorts. The effect is statistically significant at the 5% level. In other words, the *Ujamaa* policy increased the rate of inter-ethnic marriages. We interpret this as “revealed preference” evidence for a declining importance of ethnic identity relative to national identity as a consequence of the *Ujamaa* policy. The magnitude of the effect on intermarriage (0.17 standard deviations) is comparable to the effect for our main measure of self-reported national identity (0.23 standard deviations).

### 5.6. *Channels*

Our core finding and object of interest is that the *Ujamaa* policy, which combined villagization with a public education reform, helped bolster national identity among the treated cohorts. Answering the question of whether villagization or education alone had a bigger effect would require a different context and identification strategy. Our question is about the interaction of villagization and education, as the education reform would have been virtually impossible to implement without the villagization program, and nation-building efforts in the real world come as such bundles of measures.<sup>44</sup>

Nevertheless, we conduct several exercises to examine two sets of plausible channels that could underlie our main result. First, the *Ujamaa* policy may have primarily shaped national identity through public education provided in planned villages, or, second, through factors unrelated to education that differentially affected the treated cohort in more intensely villagized districts compared to other cohorts.

To differentiate between these possibilities, we examine heterogeneous treatment effects by whether respondents attended primary school. In Table 5, we interact our treatment with a dummy that indicates whether the respondent completed at least some formal primary schooling. We also control for the un-interacted schooling dummy as well as its interactions

---

<sup>43</sup> Results are robust to using the wife's current district and birth year, albeit less precise (results available upon request). This could be explained by the fact that wives are typically younger than their husbands. This makes it more likely that a wife is also treated when the husband is treated, but less likely that a husband is treated when a wife is treated.

<sup>44</sup> While villagization alone may have affected national identity across all cohorts, only those of primary schooling age during *Ujamaa* were exposed to the bundled treatment (combining villagization with education). Hence, our main coefficient can be interpreted as the differential effect of the bundled treatment compared to the effect of villagization alone. The effect of villagization alone appears to be small. Average levels of national identity are no higher in more intensively villagized districts for cohorts outside of primary schooling age during the *Ujamaa* era. The case of education without villagization is not observed, as very few had access to schooling outside of the planned villages.

with the cohort dummy and the villagization variable. Column (2), our preferred specification, indicates that the coefficient on the treatment is 0.36 standard deviations larger for those that attended primary school compared to those that did not. There is no significant effect of the treatment on national identity for individuals who did not obtain any formal schooling. This result is consistent with public schooling as a channel through which the *Ujamaa* policy shaped national identity, rather than other differences between age cohorts (unrelated to schooling) that may affect national identity and that vary with villagization.

There are a number of channels through which public education during the *Ujamaa* period may have influenced national identity. First and foremost, the content of the new political education curriculum was clearly oriented toward establishing national identity, as we outline in the Background (Section 2). In addition, establishing Swahili as the language of instruction likely played a role as well – in particular to bring students who grew up speaking distinct ethnic languages at home under a common rubric. Public education may in theory have also facilitated inter-group contact across ethnic lines and may have had implications for human capital and occupational choice. We examine these channels below.

**Public education curriculum.** As we document in the Background section, changes in the curriculum aimed at instilling a sense of national identity were a key element of the education reforms under the *Ujamaa* policy. At the same time, the policy also led to an overall increase in the supply of education in official villages which could foster national identity in treated cohorts. However, recall from Figure 2 that we see no differential effect of villagization on cohorts of primary schooling age after the *Ujamaa* period. This finding suggests that changes in the content of education that was delivered under the *Ujamaa* policy, rather than a general increase in the supply of education, explain our results.<sup>45</sup>

**Swahili as the national language of instruction.** Around the world, the establishment of a common, national language has played an important role in strengthening national identity (Habermas, 1992; Simpson, 2008; Gordon, 2015; Alesina et al., 2021). Tanzania’s President Nyerere also recognized this and introduced Swahili as the national language of instruction in all primary schools. This arguably had more significant implications for some ethnic groups than others: on the one hand, the majority of ethnolinguistic groups in Tanzania are part of the Bantu language group, to which Swahili also belongs (Ethnologue, 2015). On the other hand, ethnolinguistic groups such as the Masai natively speak languages that are completely distinct from Swahili (Müller-Crepon et al., 2022). We capture this variation in Figure A.6, which shows the distribution of average ethnolinguistic distance to Swahili of the districts in

---

<sup>45</sup> The overall supply of public education increased for all cohorts, even after the *Ujamaa* policy ended, since the public schooling infrastructure remained. However, the elements of the policy that explicitly targeted nation-building waned with the end of the *Ujamaa* era.

our sample, based on ethnic group shares from the 1967 census and data on the linguistic distance between ethnic groups from the LEDA project.

If Swahili as the national language of instruction is an important channel through which the *Ujamaa* policy fostered national identity, we might expect its introduction to have a larger impact in areas that were historically more distant in terms of language. While we are under-powered to provide clear evidence for this hypothesis, we note that our baseline coefficient estimate drops from 0.243 to 0.180 when we drop the districts with the largest linguistic distance to Swahili ( $> 0.2$ ) from the sample (results available upon request). Supported by the strong qualitative evidence, we interpret this as suggestive quantitative evidence that introducing a national language of instruction may be one of the channels through which public education strengthened national identity.

**Intergroup contact.** An alternative and perhaps complementary channel through which the *Ujamaa* policy might have shaped national identity is through intergroup contact across ethnic lines. This idea has a long history in the social sciences, beginning with psychologist Gordon Allport (1954) specifying the conditions under which intergroup contact can reduce prejudice. Allport's (1954) hypothesis has been confirmed in more recent meta-studies (Hewstone et al., 2014) and economists have also begun to more rigorously identify the effects of intergroup contact – for example showing that collaborative intergroup contact between individuals from different social groups may foster cooperation and reduce the salience of group identity (Bazzi et al., 2019; Rao, 2019; Lowe, 2021). In the context of villagization, bringing children from different ethnic groups together in public village schools could have decreased the salience of ethnic identity in favor of national identity. However, scholars have also shown that intergroup contact can foster exclusionary attitudes (Enos, 2014) and challenge social solidarity (Putnam, 2007). In our context, it is also possible that intergroup contact could have *sharpened* the salience of ethnic identities, for example due to intergroup competition for limited resources in the villages.

If intergroup contact across ethnic lines played an important role in explaining the effect of our treatment on national identity, we would expect larger effects in places where such contact was more likely to take place. To test this idea, we further interact our treatment (cohort interacted with villagization) with district-level ethnolinguistic fractionalization (ELF) in 1967.<sup>46</sup> We also control for the lower-level interactions that are not included in the fixed effects. Table A.8 shows the results for our main measure of national identity and Table A.9 shows the results for intermarriage. In all columns except one, the triple interaction of villagization  $\times$  cohort  $\times$  ELF is small and statistically insignificant. Only the

---

<sup>46</sup> ELF is computed as 1 minus the Herfindahl concentration index of ethnolinguistic group shares in each district given by the 1967 population census data. A higher ELF index corresponds to a higher degree of ethnic diversity within a district.

coefficient in Table A.9 column (5), which controls for geographical characteristics interacted with the cohort dummy, is statistically significant ( $p$ -value = 0.011). However, it has a positive sign: if anything, the effect of the *Ujamaa* policy on intermarriage appears to have been *weaker* for more ethnically diverse districts. Taken together, we do not find strong evidence in favor of the intergroup contact channel in our setting.

**Occupational choice.** Villagization increased access to schooling for the treated cohort. Schooling not only bolstered national identity among the students through political education but may have also provided them with human capital. *Ujamaa* could thus have affected national identity through occupational choice due to human capital accumulation. For example, those who obtained public education during the *Ujamaa* period may be more likely to work in public sector jobs, which in turn may affect their feelings of national identity. We investigate this avenue in Table A.10 which is analogous to our baseline specification but with dummies for different occupations as outcome variables (employed in the government, employed in the private sector, agriculture or self-employed in other sectors).<sup>47</sup> All coefficients are neither sizeable nor statistically significant, indicating that the treatment had little to no effect on occupational choice.

## 6. Results: Ujamaa and State Legitimacy

Our results indicate not only that the Tanzanian state was able to strengthen the national identity over the long term but that in building the Tanzanian *nation*, *Ujamaa* also appears to have helped strengthen the nascent Tanzanian *state* by helping to establish it as a legitimate authority. As we detail below, treated cohorts are more likely to express attitudes in favor of a strong, central state. At the same time, however, the results suggest that reforms to promote national identity have also engendered acquiescence to authoritarianism, which has been on the rise in Tanzania (and the region) in recent years (Paget, 2021).

We follow the same empirical strategy as outlined in Section 4, replacing national identity with other outcome variables from Afrobarometer. Table 6 shows the results. As above, we report standardized coefficients to facilitate interpretation. The coefficient of interest, like in prior tables, is the interaction between the respondent's district-level measure of villagization and a dummy that indicates whether the respondent is in the treated cohort. We interpret this coefficient as the effect of the *Ujamaa* policy on the outcomes stated in the column heads.

In Panel A, column (1), the outcome is based on a question asking respondents which one of two statements regarding views of the state is closest to their views: *Statement 1*:

---

<sup>47</sup> The data is from the Tanzania National Panel Survey Round 1 (2008). We use this data, which also contains the birth years and district of respondents, since information on occupation is available for only a small subsample of respondents in the Afrobarometer data.

*Citizens should be more active in questioning the actions of leaders or Statement 2: In our country, citizens should show more respect for authority.* The outcome variable is coded in 1/4 increments, equalling 0 for respondents who agree very strongly with Statement 1, and 1 for respondents who agree very strongly with Statement 2. We find that exposure to the *Ujamaa* policy has a sizable and statistically significant effect on respect for state authority: a one standard deviation increase in the treatment increases the outcome variable by 0.22 standard deviations ( $p < 0.05$ ). Prior studies have interpreted this variable as measuring citizens' critical attitudes toward government (Doorenspleet, 2012), which has also been understood as an indicator of demand for democracy (Inglehart, 1997). We interpret these results as suggesting that citizens who were exposed to the *Ujamaa* policy are more likely to see the state as a legitimate central authority, but less likely to express demand for democracy.

In Panel A, column (2), we examine respondents' stated approval for a system of government where only one political party is allowed to stand for election and hold office. We find that exposure to the *Ujamaa* policy is positively correlated with approval of one-party rule: a one standard deviation increase in the treatment is associated with a 0.21 standard deviations higher measure of approval of one-party rule ( $p < 0.01$ ). Indeed, to Nyerere in Tanzania's early years, single party rule was seen as a necessity to foster national integration in a country characterized by substantial ethnic differences (Komba, 1996). Our results speak to the legacy of single-party rule, as well as the intertwined nature of the party and the state, which has persisted despite the (re-)introduction of multi-partyism (Paget, 2021).

In Panel A, column (3), the outcome is based on a question asking respondents which of two statements is closest to their views: *Statement 1: People are like children; the government should take care of them like a parent* or *Statement 2: Government is like an employee; the people should be the bosses who control the government.* The outcome variable is coded in 1/4 increments, equalling 0 for respondents who agree very strongly with Statement 2, and 1 for respondents who agree very strongly with Statement 1. We again find that the *Ujamaa* policy has a sizable and statistically significant effect on pro-government attitudes: a one standard deviation increase in the treatment is estimated to increase the outcome variable by 0.44 standard deviations ( $p < 0.01$ ). Such widespread trust in government can facilitate voluntary compliance with state policies. For instance, scholars point to extensive buy-in of public health measures such as malaria control (Croke, 2012) among Tanzanians in contrast to their neighbors. Such compliance is difficult to achieve without broad acceptance of the state as a legitimate authority.

In Panel A, column (4), the dependent variable indicates whether respondents got together with others to raise an issue to the government, which we interpret as another proxy for critical attitudes toward government. The negative coefficient (-0.1 standard deviations) is in line with the result in column (1) described above, in that it indicates greater acceptance of the state as a legitimate authority. However, the coefficient is not statistically significant

at conventional levels.

In Panel B, columns (1) to (4), respondents were asked how much they trust each of the following news sources: government newspapers (column (1)); government broadcasting service (TV/radio) (column (2)); independent newspapers (column (3)) and independent broadcasting services (TV/radio) (column (4)). We find that exposure to the *Ujamaa* policy is positively correlated with stated trust in government broadcasting services (0.12 and 0.25 standard deviations) but much less so with independent media services (0.04 and 0.07 standard deviations), even though the coefficients are imprecisely estimated. This is consistent with our findings on citizens' views of the state above.

In sum, we find that cohorts exposed to *Ujamaa* are more likely to respect state authority and approve of one-party rule, and have higher trust in central government institutions such as state media. These findings suggest that the *Ujamaa* policy contributed to putting in place the new Tanzanian state as a legitimate central authority. There is evidence to suggest that a strong shared identity and trust in a commonly accepted central state are important ingredients for avoiding fragility and civil conflict ([Besley, 2020](#)). Within Africa, intra-state conflict has been more prevalent in countries where a smaller proportion of the population identifies with the nation as a whole ([Besley and Reynal-Querol, 2014](#)), as we show in Figure A.7. This correlation is consistent with our narrative that creating a strong national identity may contribute to loyalty to the state and more political stability as a result. Tanzania today is a clear outlier amongst its peers, with a very low prevalence of intra-state conflict despite being one of the most ethnically diverse countries.<sup>48</sup>

However, we note that efforts to forge a strong, shared national identity do not necessarily generate more cooperation and trust among the population. While exposure to the *Ujamaa* policy generated higher levels of trust in state institutions, interestingly it did not have the same impact on generalized intra- or inter-ethnic trust. In unreported results, we find only small and statistically insignificant effects of the *Ujamaa* policy on trust in members of one's own ethnic group (0.09 SD) or in members of other ethnic groups (-0.15 SD). If anything, the *Ujamaa* policy generally *decreased* inter-ethnic trust. These results indicate that the creation of a national identity in a top-down manner may have first and foremost strengthened the one-party state's ability to govern but that the extent of social cohesion across ethnic groups may have remained relatively limited.

---

<sup>48</sup> We leave it to future research to further test this narrative by examining conflict directly. Data on conflict is required to be at the individual or cohort level given our identification strategy. Such data is not available in our context.

## 7. Conclusion

Political leaders throughout history have attempted to overcome the challenges to governance of diverse populations using ‘bundles’ of nation-building measures including the establishment of national languages, public education, and forced resettlement. We study the consequences of one of the largest of such efforts in post-colonial Africa — the Tanzanian *Ujamaa* policy, which combined the reorganization of millions of people with a public education reform. We show that individuals most affected by the policy are more likely to primarily identify with Tanzania as a whole rather than with their ethnic group and are more likely to marry across ethnic lines. The effects are persistent and substantive.

The positive impact on national identity we detect stands in contrast to previous studies of state-building policies that also involved forced resettlement of diverse populations (e.g., Dippel, 2014). We argue this may reflect the bundled nature of the *Ujamaa* policy – in particular, the fact that it combined resettlement with a public education reform that allowed the nascent Tanzanian state to project its national ideology to an equally young (and thus potentially more impressionable) population. *Ujamaa*’s sharp timing and variation across space help us address identification challenges that have hindered similar research to date, contributing to fill an important gap in the literature.

We also find that the Tanzanian state appears to have successfully established itself as a legitimate central authority in the eyes of the treated cohort. At the same time, this may have led to reduced demands for government accountability, highlighting the difficult balance inherent to effective state-building (Acemoglu and Robinson, 2020). Members of the treated cohort are more likely to trust state institutions and less likely to question state authority. However, we find few signs of increased cooperation and generalized trust among those most exposed to *Ujamaa*.

Furthermore, while the state efforts to assert control did not result in more violence (in contrast to a growing literature in economics and political science) the policy certainly came with its drawbacks.<sup>49</sup> These include the above-mentioned decreased demand for accountability, as well as economic and ecological losses that have been attributed to villagization (Scott, 1998), and arguably the erosion of local diversity.

Our examination of *Ujamaa* in Tanzania has broader implications for the study of state-building in diverse societies. The question of how to effectively govern diverse groups spread over large territories has challenged state leaders throughout history and into the present day. Although in essence facing similar challenges, the policy tools leaders have deployed and outcomes they have achieved vary considerably — as evidenced by the divergent trajectories of the United States, China, the former Soviet Union, and Somalia to give just a few examples.

---

<sup>49</sup> Note that we do not study the policy’s impact on violence directly, though Tanzania’s peaceful political settlement is widely accepted by close observers (Lofchie, 2014; Ibhwah and Dibua, 2003).

Our finding that efforts to build a national identity can also promote unquestioning support for state authority raises important questions related to the potential anti-democratic nature of state-building reforms. These questions are particularly important in light of widespread concern about democratic backsliding and institutional erosion around the world (Hyde, 2020; V-Dem Institute, 2022). Our results encourage further scrutiny of the political economy of state-building in diverse societies. We demonstrate that the choices leaders make on how to build a shared identity from diverse groups are first order and deserve further attention.

## References

- Acemoglu, D. and J. A. Robinson (2020). *The Narrow Corridor: States, Societies, and the Fate of Liberty*. Penguin Books.
- Afrobarometer (2017). *Afrobarometer Data, Tanzania*. Available at <http://www.afrobarometer.org>.
- Akerlof, G. and R. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics* 115(3), 715–753.
- Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics* 114(4), 1243–1284.
- Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth* 8(2), 155–194.
- Alesina, A., W. Easterly, and J. Matuszeski (2011). Artificial states. *Journal of the European Economic Association* 9(2), 246–277.
- Alesina, A., P. Giuliano, and B. Reich (2021). Nation-building and education. *The Economic Journal* 131(638), 2273–2303.
- Alesina, A. and E. La Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature* 43(3), 762–800.
- Alesina, A. and B. Reich (2015). Nation building. *Working paper, Harvard University*.
- Alesina, A. and E. Spolaore (1997). On the number and size of nations. *The Quarterly Journal of Economics* 112(4), 1027–1056.
- Allport, G. W. (1954). *The Nature of Prejudice*. Reading, MA: Addison-Wesley.
- Anderson, B. (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Antman, F. and B. Duncan (2015). Incentives to identify: Racial identity in the age of affirmative action. *The Review of Economics and Statistics* 97(3), 710–713.

- Assouad, L. (2021). Charismatic leaders and nation-building. *Working Paper, Paris School of Economics*.
- Atkin, D., E. Colson-Sihra, and M. Shayo (2021). How do we choose our identity? A revealed preference approach using food consumption. *Journal of Political Economy* 129(4), 1193–1251.
- Bagachwa, M. S. and A. Naho (1995). Estimating the second economy in tanzania. *World Development* 23(8), 1387–1399.
- Bandiera, O., M. Mohnen, I. Rasul, and M. Viarengo (2018). Nation-building through compulsory schooling during the age of mass migration. *The Economic Journal* 129(617), 62–109.
- Bandyopadhyay, S. and E. Green (2013). Nation-building and conflict in modern africa. *World Development* 45, 108–118.
- Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2019). Unity in diversity? How intergroup contact can foster nation building. *American Economic Review* 109(11), 3978–4025.
- Bazzi, S., M. Hilmy, and B. Marx (2020). Islam and the state: Religious education in the age of mass schooling. *NBER Working Paper*.
- BenYishay, A., R. Rotberg, J. Wells, Z. Lv, S. Goodman, L. Kovacevic, and D. Runfola (2017). *Geocoding Afrobarometer Rounds 1 - 6: Methodology & Data Quality*. Williamsburg, VA: AidData.
- Besley, T. (2020). State capacity, reciprocity, and the social contract. *Econometrica* 88(4), 1307–1335.
- Besley, T. and T. Persson (2011). *Pillars of Prosperity*. Princeton University Press.
- Besley, T. and M. Reynal-Querol (2014). The legacy of historical conflict: Evidence from Africa. *American Political Science Review* 108(2), 319–336.
- Beswick, D. (2010). Managing dissent in a post-genocide environment: the challenge of political space in Rwanda. *Development and Change* 41(2), 225–251.
- Blouin, A. and S. W. Mukand (2019). Erasing ethnicity? Propaganda, nation building, and identity in Rwanda. *Journal of Political Economy* 127(3), 1008–1062.
- Botticini, M. and Z. Eckstein (2012). *The Chosen Few: How Education Shaped Jewish History, 70-1492*. Princeton University Press.
- Breton, A. (1964). The economics of nationalism. *Journal of Political Economy* 72(4), 376–386.
- Bryceson, D. F. (1982). Peasant commodity production in post-colonial tanzania. *African Affairs* 81(325), 547–567.

Bureau of Statistics, Ministry of Planning and Economic Affairs (1981). *1978 Population Census*. Dar es Salaam: United Republic of Tanzania.

Bureau of Statistics/Tanzania and Macro International (1997). *Tanzania Demographic and Health Survey 1996*. Calverton, Maryland: Bureau of Statistics/Tanzania and Macro International.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics* 225(2), 200–230.

Cantoni, D., Y. Chen, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2017). Curriculum and ideology. *Journal of Political Economy* 125(2), 338–392.

Cantoni, D. and N. Yuchtman (2013). The political economy of educational content and development: Lessons from history. *Journal of Development Economics* 104, 233–244.

Cassan, G. (2015). Identity-based policies and identity manipulation: Evidence from Colonial Punjab. *American Economic Journal: Economic Policy* 7(4), 103–31.

Cassan, G., D. Keniston, and T. Kleineberg (2021). A division of laborers: Identity and efficiency in India. *NBER Working Paper* 28462.

Clingingsmith, D., A. I. Khwaja, and M. Kremer (2009). Estimating the impact of the Hajj: Religion and tolerance in Islam's global gathering. *The Quarterly Journal of Economics* 124(3), 1133–1170.

Clots-Figueras, I. and P. Masella (2013). Education, language and identity. *The Economic Journal* 123(570), F332–F357.

Collier, D., P. Collier, S. Radān, S. Rawān, S. Wangwe, S. Radwan, A. Wagner, P. mondial de l'emploi, I. L. Office, and W. E. Programme (1986). *Labour and Poverty in Rural Tanzania: Ujamaa and Rural Development in the United Republic of Tanzania*. Clarendon Press.

Cornwell, C., I. M. Rivera, and I. Schmutte (2014). Wage discrimination when identity is subjective: Evidence from changes in employer-reported race. *Working Paper, University of Georgia*.

Court, D. and K. Kinyanjui (1980). Development policy and educational opportunity: The experience of Kenya and Tanzania. *Institute for Development Studies, University of Nairobi*.

Croke, K. (2012). The political economy of child mortality decline in Tanzania and Uganda, 1995–2007. *Studies in Comparative International Development* 47(4), 441–463.

Dahis, R., E. Nix, and N. Qian (2019). Choosing racial identity in the United States, 1880–1940. *NBER Working Paper* 26465.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9), 2964–96.

Dehdari, S. H. and K. Gehring (2022). The origins of common identity: Evidence from Alsace-Lorraine. *American Economic Journal: Applied Economics* 14(1), 261–92.

- Dell, M. and P. Querubin (2018). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics* 133(2), 701–764.
- Depetris-Chauvin, E., R. Durante, and F. Campante (2020). Building nations through shared experiences: Evidence from African football. *American Economic Review* 110(54), 1572–1602.
- Dippel, C. (2014). Forced coexistence and economic development: Evidence from Native American Reservations. *Econometrica* 82(6), 2131–2165.
- Doorenspleet, R. (2012). Critical citizens, democratic support and satisfaction in African democracies. *International Political Science Review* 33(3), 279–300.
- Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review* 91(4), 795–813.
- Duggan, C. (2008). *The Force of Destiny: A History of Italy Since 1796*. Houghton Mifflin Harcourt.
- Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics* 112(4), 1203–1250.
- Enos, R. D. (2014). Causal effect of intergroup contact on exclusionary attitudes. *Proceedings of the National Academy of Sciences* 111(10), 3699–3704.
- Ethnologue (2015). *Languages of the World, 18th edition*. United States.
- Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth* 8(2), 195–222.
- Friedman, W., M. Kremer, E. Miguel, and R. Thornton (2016). Education as liberation? *Economica* 83(329), 1–30.
- Gennaioli, N. and I. Rainer (2007). The modern impact of precolonial centralization in Africa. *Journal of Economic Growth* 12(3), 185–234.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics* 225(2), 254–277.
- Gordon, D. C. (2015). *The French Language and National Identity (1930–1975)*, Volume 22. Walter de Gruyter GmbH & Co KG.
- Government of Tanzania (1984). Universal primary education and village based teacher training programme in Tanzania. Ministry of Education.
- Guiso, L., P. Sapienza, and L. Zingales (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives* 20(2), 23–48.
- Habermas, J. (1992). Citizenship and national identity: Some reflections on the future. *Praxis International* 12(1), 1–19.

- Herbst, J. (2014). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton University Press.
- Hewstone, M., S. Lolliot, H. Swart, E. Myers, A. Voci, A. Al Ramiah, and E. Cairns (2014). Intergroup contact and intergroup conflict. *Peace and Conflict: Journal of Peace Psychology* 20(1), 39.
- Hobsbawm, E. J. (1992). *Nations and Nationalism since 1780: Programme, Myth, Reality* (2 ed.). Canto. Cambridge University Press.
- Holsti, K. J. (1996). *The State, War, and the State of War*. Cambridge University Press.
- Hudson, M. C. and C. L. Taylor (1972). *World Handbook of Political and Social Indicators*. Yale University Press.
- Hyde, S. D. (2020). Democracy's backsliding in the international environment. *Science* 369(6508), 1192–1196.
- Hydén, G. (1980). *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry*. University of California Press.
- Ibhawoh, B. and J. I. Dibua (2003). Deconstructing Ujamaa: The legacy of Julius Nyerere in the quest for social and economic development in Africa. *African Journal of Political Science* 8(1), 59–83.
- Inglehart, R. (1997). *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton University Press.
- Jensen (1968). *Government of Tanzania District Statistics*. Dar es Salaam: Government Printers.
- Jia, R. and T. Persson (2013). Ethnicity in children and mixed marriages: Theory and evidence from China. *Working Paper, UC San Diego*.
- Kikula, I. S. (1997). *Policy Implications on Environment: The Case of Villagisation in Tanzania*. Nordic Africa Institute.
- Komba, W. L. M. (1996). *Changing Politics and Political Culture in Tanzania: The Impact on Political Education and Civics Curricula 1967-1994*. Ph. D. thesis, Institute of Education, University of London.
- Lofchie, M. F. (2014). *The Political Economy of Tanzania: Decline and Recovery*. University of Pennsylvania Press.
- Lorgen, C. C. (2000). Villagisation in Ethiopia, Mozambique, and Tanzania. *Social Dynamics* 26(2), 171–198.
- Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review* 111(6), 1807–44.
- Martin, T. D. (2001). *The Affirmative Action Empire: Nations and Nationalism in the Soviet Union, 1923-1939*. Cornell University Press.

- Mbuyi, D. M. (1987). *Beyond Policy and Language Choice: An Analysis of Texts in Four Instructional Contexts in East Africa*. Special Studies in Comparative Education, Number Eighteen. ERIC.
- McHenry, D. E. (1979). *Tanzania's Ujamaa Villages. The Implementation of a Rural Development Strategy*. Institute of International Studies, University of California, Berkeley.
- Michalopoulos, S. and E. Papaioannou (2013). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics* 129(1), 151–213.
- Michalopoulos, S. and E. Papaioannou (2016). The long-run effects of the Scramble for Africa. *American Economic Review* 106(7), 1802–48.
- Miguel, E. (2004). Tribe or nation? Nation building and public goods in Kenya versus Tanzania. *World Politics* 56(3), 327–362.
- Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics* 89(11-12), 2325–2368.
- Mohiddin, A. (1968). Ujamaa: A commentary on President Nyerere's vision of Tanzanian Society. *African Affairs* 67(267), 130–143.
- Müller-Crepon, C., Y. Pengl, N.-C. Bormann, et al. (2022). Linking Ethnic Data from Africa (LEDA). *Journal of Peace Research* 59(3), 425–435.
- National Bureau of Statistics (2010). *Tanzania National Panel Survey - Wave 1, 2008*. Dar es Salaam, Tanzania: NBS.
- Ngallaba, S., I. R. Saidi Hussein Kapiga, and J. T. Boerma (1993). *Tanzania Demographic and Health Survey 1991/1992*. Dar es Salaam, Tanzania: Bureau of Statistics/Tanzania and Macro International.
- Nyerere, J. (1963). *Democracy and the Party System*. Tanganyika Standard.
- Nyerere, J. K. (1969a). *Freedom and Socialism/Uhuru na Ujamaa: A Selection from Writings and Speeches, 1965-1967*. Oxford University Press.
- Nyerere, J. K. (1969b). *Ujamaa - Essays on Socialism*. Oxford University Press.
- Nyerere, J. K. (1982). *Five Years of CCM Government*. Dar es Salaam, Tanzania: CCM.
- Okunogbe, O. M. (2021). Does exposure to other ethnic regions promote national integration? Evidence from Nigeria. *World Bank Policy Research Working Paper*.
- Omari, C. K. (1995). The management of tribal and religious diversity. In *Mwalimu: Influence of Nyerere*. London: James Currey.
- Osafu-Kwaako, P. (2012). *Essays in Economic History and Development*. Ph. D. thesis, Harvard University.
- Paget, D. (2021). Tanzania: the authoritarian landslide. *Journal of Democracy* 32(2), 61–76.

- Paluck, E. L. and D. P. Green (2009). Deference, dissent, and dispute resolution: An experimental intervention using mass media to change norms and behavior in Rwanda. *American Political Science Review* 103(4), 622–644.
- Peeples, A. (2018). Relearning Ujamaa: Education in Tanzania from 1954 to 2002. *Department of History, UNC Thesis*.
- Pisani, E. (2014). *Indonesia, etc.: Exploring the Improbable Nation*. WW Norton & Company.
- Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian Political Studies* 30(2), 137–174.
- Puttermann, L. (1986). *Peasants, Collectives, and Choice: Economic Theory and Tanzania's Villages*. Greenwich CT: Jai Press Inc.
- Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review* 109(3), 774–809.
- Ray, D. and J. Esteban (2017). Conflict and development. *Annual Review of Economics* 9, 263–293.
- Samoff, J. (1991). Local initiatives and national policies: The politics of private schooling in Tanzania. *International Journal of Educational Research* 15(5), 377–391.
- Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.
- Shao, I. F. (1982). A neo-colony and its problems during the process of attempting to bring about socialist rural transformation: the case of Tanzania. In *Taamuli: A Political Science Forum*, Volume 12, pp. 29–46.
- Shengena, J. (1979). Workers' education. In *The Tanzanian Experience: Education for Liberation and Development*, Volume 16. Unesco Institute for Education.
- Simpson, A. (2008). *Language and National Identity in Africa*. Oxford University Press.
- Smith, A. D. (1986). State-making and nation-building. *States in History* 15, 228–263.
- Straus, S. and L. Waldorf (2011). *Remaking Rwanda: State Building and Human Rights after Mass Violence*. Critical Human Rights. University of Wisconsin Press.
- Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225(2), 175–199.
- Thomas, I. (1982). *Development Projects and Aspects of Population Redistribution in Tanzania*. Number 118. School of Development Studies, University of East Anglia.
- Thomson, S. (2013). *Whispering Truth to Power: Everyday Resistance to Reconciliation in Postgenocide Rwanda*. Africa and the Diaspora. University of Wisconsin Press.
- Tilly, C. and G. Ardant (1975). *The Formation of National States in Western Europe*. Princeton University Press.

- Tordoff, W. (1967). Tanzania: democracy and the one-party state. *Government and Opposition* 2(4), 599–614.
- Tripp, A. M. (1999). The political mediation of ethnic and religious diversity in tanzania. In *The Accommodation of Cultural Diversity*, pp. 37–71. Springer.
- V-Dem Institute (2022). Autocratization changing nature? Democracy Report 2022.
- Von Freyhold, M. (1979). *Ujamaa Villages in Tanzania: Analysis of a Social Experiment*. Heinemann.
- Weber, E. (1976). *Peasants into Frenchmen: the Modernization of Rural France, 1870-1914*. Stanford University Press.
- Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology*, Volume 2. University of California Press.
- Wenban-Smith, H. (2014). *Urbanisation in Tanzania: Population Growth, Internal Migration and Urbanisation in Tanzania 1967–2012: A Census Based Regional Analysis*. International Growth Centre.
- Wimmer, A. (2018). *Nation Building: Why Some Countries Come Together While Others Fall Apart*. Princeton University Press.

# TABLES AND FIGURES

Table 1: Descriptive Statistics

Variable	Data Source	Mean	Std.Dev.	Min.	Max.	Obs.
<b>Panel A. District level (baseline sample)</b>						
Villagization (share of rural population in official villages in 1978)	Population Census	0.93	0.08	0.50	1	54
Share urban in 1967 (% of population)	Population Census	0.03	0.05	0	0.24	54
Primary school enrollment rate in 1967 (per '000 inhabitants)	Population Census	72.27	23.81	32.70	155.80	54
District revenue in 1967 ('000 shs)	Jensen (1968)	1922.97	1525.97	220.60	8219.90	54
Ethnolinguistic fractionalization in 1967	Population Census	0.57	0.24	0.08	0.91	54
Distance to Dar-Es-Salaam in kilometers	ESRI	516.29	267.75	6.37	1021.23	54
Centroid latitude	ESRI	-5.93	2.88	-10.96	-1.34	54
Centroid longitude	ESRI	34.95	2.71	30.20	40.01	54
Average altitude in meters	ESRI	1051.45	425.69	110.95	1729.06	54
Average slope	ESRI	25.51	14.92	5.13	63.72	54
Hospital beds in 1968 (per '000.000 inhabitants)	Jensen (1968)	227.00	176.82	0	768.00	54
Drought in 1974 (truncated z-score)	Tanzania Meteo	-0.03	0.10	-0.52	0	54
Drought in 1975 (truncated z-score)	Tanzania Meteo	-0.55	0.44	-1.53	0	54
Drought in 1976 (truncated z-score)	Tanzania Meteo	-0.48	0.40	-1.44	0	54
Average linguistic distance to Swahili	LEDA	0.21	0.17	0.12	1	54
In-migration rate in 2008 (% ever born in different district)	TNPS	0.22	0.16	0	0.67	54
Out-migration rate in 2008 (% ever born in different district)	TNPS	0.27	0.16	0	0.62	53
<b>Panel B. Individual level (baseline sample)</b>						
National identity, five-level Likert scale (recoded)	Afrobarometer	0.89	0.23	0	1	862
Birth year	Afrobarometer	1962	6.50	1948	1971	862
Completed primary school	Afrobarometer	0.89	0.32	0	1	733
Respect authority	Afrobarometer	0.21	0.31	0	1	848
Support one-party rule	Afrobarometer	0.44	0.39	0	1	841
See government as parent	Afrobarometer	0.58	0.41	0	1	362
Took action to hold government accountable	Afrobarometer	0.77	0.42	0	1	857
Trust government newspapers	Afrobarometer	0.84	0.23	0	1	421
Trust government TV/radio	Afrobarometer	0.87	0.22	0	1	446
Trust independent newspapers	Afrobarometer	0.77	0.26	0	1	417
Trust independent TV/radio	Afrobarometer	0.78	0.25	0	1	427
Main occupation: employed in government (%)	TNPS	0.05	0.22	0	1	1190
Main occupation: employed in private sector (%)	TNPS	0.03	0.17	0	1	1190
Main occupation: agriculture (%)	TNPS	0.78	0.41	0	1	1190
Main occupation: self-employed (%)	TNPS	0.10	0.30	0	1	1190
Married within same ethnic group (if married, husband)	DHS	0.67	0.47	0	1	1,192

*Note:* Data includes observations from cohorts used in the main analysis (born 1948-1971) presented in Table 2. The capital Dar es Salaam and the islands Zanzibar, Pemba and Mafia are excluded from all analyses. Data construction and sources are described in detail in Section 3, Table A.2 and Table A.3.

Table 2: The Effect of *Ujamaa* on National Identity, Main Results

	Dependent Variable: National vs. Ethnic Identity							
	controlling for treated cohort dummy interacted with							
	no zone-cohort FE	baseline	distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Treated Cohort	0.181** (0.086)	0.243** (0.114)	0.230* (0.117)	0.251** (0.113)	0.197* (0.110)	0.179** (0.087)	0.240** (0.116)	0.204** (0.098)
Number of Observations	862	862	862	862	862	862	862	862
Number of Districts	54	54	54	54	54	54	54	54
R-squared	0.102	0.118	0.120	0.119	0.120	0.130	0.118	0.125
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

*Note:* The unit of observation is an individual Afrobarometer respondent in district  $d$  belonging to school cohort  $t$ . All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in  $1/4$  increments ( $1 =$  respondent identifies only with the nation as a whole;  $0 =$  respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. All district-level variables and the outcome variable are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by the Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 3: The Effect of *Ujamaa* on National Identity, Robustness

	Dependent Variable: National vs. Ethnic Identity						
	individual level						
	baseline	unweighted	controlling for respondent beliefs	rural sample only	dropping districts with highest migration rates	migration bounding exercise, bounding exercise, extreme case	migration alternative case
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Villagization × Treated Cohort	0.243** (0.114)	0.221** (0.089)	0.238** (0.114)	0.401** (0.173)	0.206* (0.115)	0.145 (0.101)	0.207** (0.101)
Number of Observations	862	862	862	714	725	862	862
Number of Districts	54	54	54	49	43	54	54
R-squared	0.118	0.110	0.122	0.147	0.101	0.115	0.118
District FE	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓	✓	✓	✓

*Note:* The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to school cohort  $t$ . All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the respondent is in the treated cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recorded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). All district-level variables and the outcome variable are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer, except column (2) which does not apply the survey weights. Column (3) controls for a dummy variable that indicates whether the respondent erroneously believes that the Afrobarometer surveyor was sent by a Tanzanian government entity. Column (4) includes only rural respondents. Column (5) drops all districts with either an in- or an out-migration rate in the highest deciles. Columns (6) and (7) are explained in [Appendix B](#). Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 4: The Effect of *Ujamaa* on Intermarriage

	Dependent Variable: Married Same Ethnicity							
	controlling for treated cohort dummy interacted with							
no zone-cohort FE	baseline	distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks	(8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Treated Cohort	-0.055* (0.032)	-0.073** (0.036)	-0.087*** (0.032)	-0.068* (0.036)	-0.061* (0.036)	-0.078** (0.034)	-0.074** (0.036)	-0.085** (0.038)
Number of Observations	1,192	1,192	1,192	1,192	1,192	1,192	1,192	1,192
Number of Districts	57	57	57	57	57	57	57	57
R-squared	0.109	0.111	0.114	0.111	0.113	0.115	0.111	0.113
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

*Note:* The unit of observation is a married, male DHS respondent  $i$  in district  $d$  belonging to school cohort  $t$ . All columns regress a dummy for whether the respondent's spouse is of the same ethnic group as the respondent on the interaction between the respondent's district-level measure of historical villagization and a dummy that indicates whether the respondent is in the treated cohort, controlling for the 1967 district primary schooling enrollment share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment share in rural areas interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from DHS 1991 and 1996 (later rounds did not report the respondent's ethnicity). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. The villagization variable is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 5: The Effect of *Ujamaa* on National Identity, By Primary Schooling

	no zone-cohort FE	baseline	Dependent Variable: National vs. Ethnic Identity						(8)	
			controlling for treated cohort dummy interacted with							
			distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)			
Completed School	-0.118 (0.235)	-0.100 (0.230)	-0.075 (0.232)	-0.106 (0.233)	-0.074 (0.228)	-0.051 (0.230)	-0.101 (0.231)	-0.100 (0.232)		
Villagization × Treated Cohort	-0.057 (0.163)	-0.085 (0.174)	-0.080 (0.180)	-0.076 (0.179)	-0.166 (0.171)	-0.145 (0.167)	-0.092 (0.171)	-0.087 (0.172)		
Villagization × Completed School	-0.238* (0.129)	-0.206 (0.126)	-0.201 (0.127)	-0.201 (0.125)	-0.224* (0.131)	-0.228* (0.131)	-0.208 (0.126)	-0.172 (0.119)		
Treated Cohort × Completed School	0.320 (0.296)	0.296 (0.286)	0.269 (0.280)	0.300 (0.287)	0.279 (0.281)	0.261 (0.282)	0.298 (0.287)	0.282 (0.288)		
Villagization × Treated Cohort × Completed School	0.321 (0.291)	0.361* (0.190)	0.344* (0.192)	0.355* (0.191)	0.377* (0.190)	0.354* (0.193)	0.362* (0.190)	0.333* (0.190)		
Number of Observations	733	733	733	733	733	733	733	733		
Number of Districts	54	54	54	54	54	54	54	54		
R-squared	0.136	0.155	0.157	0.155	0.158	0.163	0.155	0.157		
District FE	✓	✓	✓	✓	✓	✓	✓	✓		
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓		
Zone-Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓		

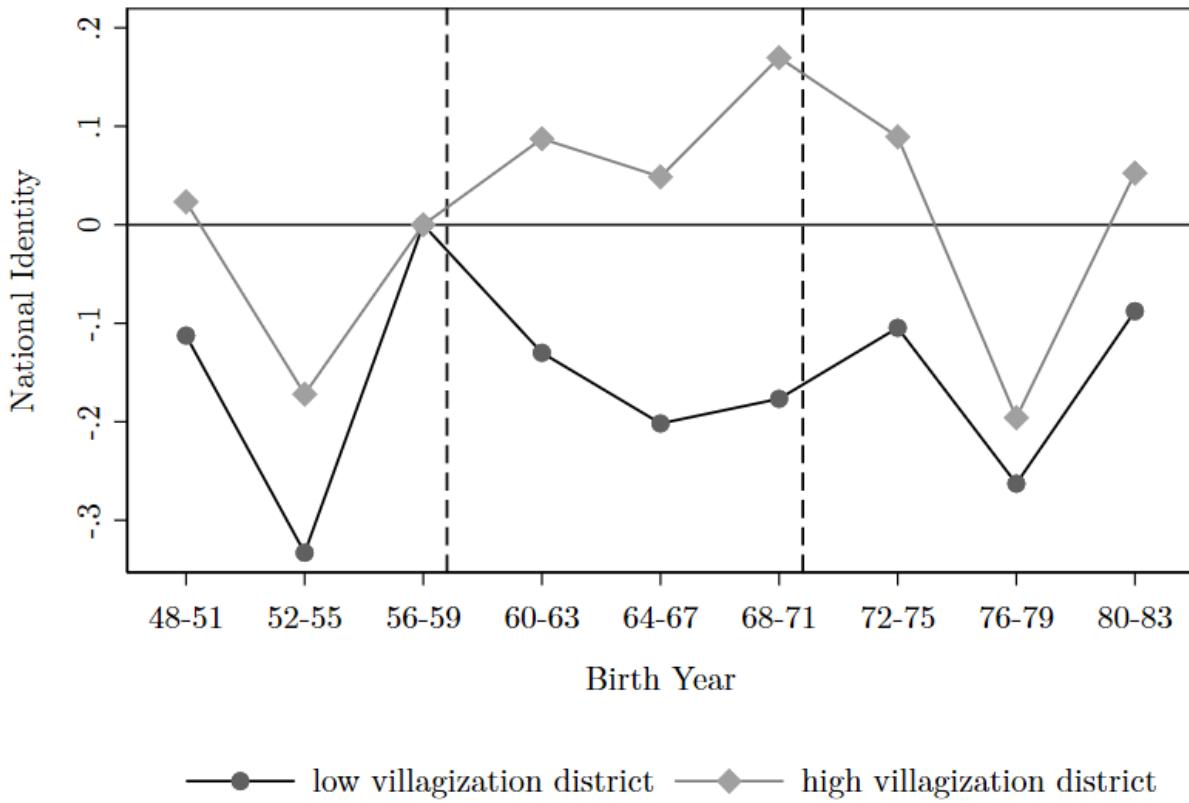
*Note:* The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to school cohort  $t$ . All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the respondent is in the treated cohort, a dummy that indicates whether the respondent ever completed some formal primary schooling, all double and triple interactions between villagization, cohort and schooling (that are not included in the fixed effects), controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The variables varying at the district- or cohort-level only are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in  $1/4$  increments ( $1 =$  respondent identifies only with the nation as a whole;  $0 =$  respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. All district-level variables and the outcome variable are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 6: The Effects of *Ujamaa* on State Legitimacy and Accountability

	(1)	(2)	(3)	(4)
	Panel A.			
	Dependent Variable:			
	respect authority	support one-party rule	see government as parent	took action to hold government accountable
Villagization × Treated Cohort	0.223** (0.088)	0.212*** (0.073)	0.435*** (0.134)	-0.096 (0.089)
Number of Observations	866	858	363	874
Number of Districts	54	54	49	54
R-squared	0.085	0.090	0.183	0.137
	Panel B.			
	Dependent Variable:			
	trust government newspapers	trust government TV / radio	trust independent newspapers	trust independent TV / radio
Villagization × Treated Cohort	0.122 (0.134)	0.253* (0.148)	0.038 (0.132)	0.067 (0.154)
Number of Observations	433	460	429	439
Number of Districts	47	47	47	47
R-squared	0.208	0.187	0.191	0.218
District FE	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓

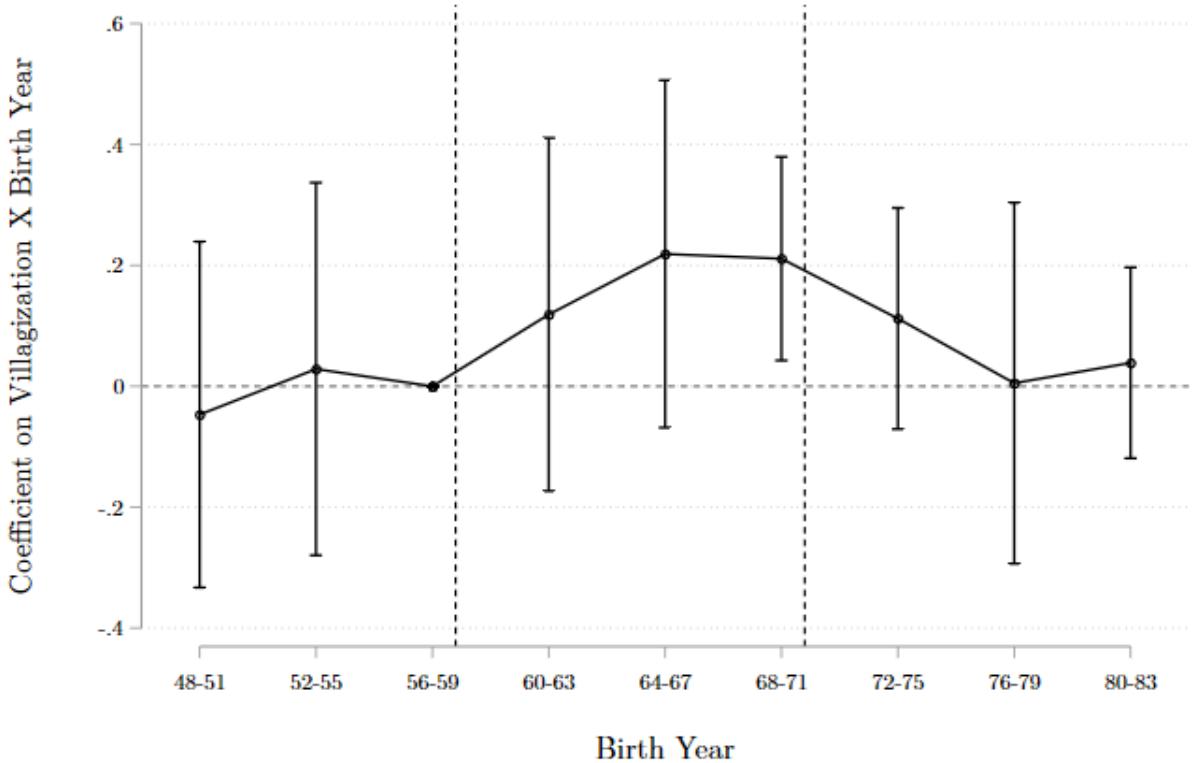
Note: The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to school cohort  $t$ . All columns regress the outcome stated in the column heads on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether that individual is in the treated cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variables are from Afrobarometer rounds 3 and 4 (2005-2008) and are as follows (recoded from Likert scale). Panel A: (1) "Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: Citizens should be more active in questioning the actions of leaders. Statement 2: In our country, citizens should show more respect for authority." The dependent variable registers agreement with Statement 2.; (2) "There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office." The dependent variable registers approval with this option.; (3) "Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: People are like children; the government should take care of them like a parent. Statement 2: Government is like an employee; the people should be the bosses who control the government." The dependent variable registers agreement with Statement 1.; (4) "Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year" Got together with others to raise an issue"; Panel B: (1) "How much do you trust each of the following, or haven't you heard enough about them to say: Government newspapers?"; (2) "Government broadcasting service (TV / radio)?"; (3) "Independent newspapers?"; (4) "Independent broadcasting services (TV / radio)?". Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All district-level variables and the outcome variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Figure 1: National Identity by Cohort and Villagization (Descriptive Evidence)



*Note:* The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to birth cohort  $t$ . Each birth cohort includes four years between 1948 and 1991 (1948-1951, 1952-1955, ...). The dashed lines indicate the first and last birth cohorts that were of 4<sup>th</sup> grade age (10 years old) during the villagization period (1970-1981). The black line plots the coefficients from a regression of the measure of an individual's current national identity on birth cohort dummies among high (low) villagization districts (high = above sample mean, low = below sample mean). The two regressions include no fixed effects or controls. The estimates are relative to the 1956-1959 birth cohort, which is the omitted category. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole, 0 = respondent identifies only with her ethnic group). The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. The outcome variable is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. The regressions are weighted using the survey weights provided by Afrobarometer.

Figure 2: The Effect of *Ujamaa* on National Identity, All Cohorts



*Note:* The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to birth cohort  $t$ . Each birth cohort includes four years between 1948 and 1991 (1948-1951, 1952-1955, ...). The dashed lines indicate the first and last birth cohorts that were of 4th grade age (10 years old) during the villagization period (1970-1981). The thick line plots the coefficients from a regression of the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and dummies that indicate whether the individual is in each of the cohorts shown. The vertical solid lines show 95% confidence intervals based on robust standard errors clustered at the district level. The estimates are relative to the 1956-1959 birth cohort, which is the omitted category. District and cohort fixed effects as well as all baseline controls (1967 district share urban interacted with cohort fixed effects, 1967 district primary school enrollment rate interacted with cohort fixed effects, zone-cohort fixed effects) are included. The un-interacted variables are included in the fixed effects. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole, 0 = respondent identifies only with her ethnic group). The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort fixed effects, are at the district level and based on the 1967 Census (except weather shocks and geographical features). All variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. The regression is weighted using the survey weights provided by Afrobarometer.

# ONLINE APPENDIX

## Appendix A. Correlates with Villagization

Here we examine potential correlates with the observed variation in villagization across districts. Table A.11 documents the pairwise correlations of these variables with villagization (conditional on zone fixed effects, as in our baseline specification). We use newly digitized district-level variables from the 1967 population census, the last census before the villagization policy was implemented, as well as other official government statistics (Jensen, 1968).

Our measure of villagization (share of the rural district population living in planned villages in 1978) is weakly positively correlated with the pre-*Ujamaa* degree of urbanization (1 - share of the rural district population in 1967). While the policy was rural in nature, the government may have been more successful at consolidating the population in districts with existing settlements. Furthermore, we might expect the government to target areas with low levels of previous access to schooling given the policy's goal to build national identity through public education. However, the data shows a weakly positive correlation between villagization and the pre-policy primary school enrollment rate. This could be explained by the fact that the government was more successful in implementing the policy in areas that already had existing (schooling) infrastructure. In any case, both urbanization (3% of the country's population according to the 1967 population census) and primary school enrollment (7%) levels were very low prior to villagization, and we account for both in our preferred specification.

Since the policy was implemented by district officials, we might expect districts with higher local state capacity to villagize more intensely. We proxy for local state capacity using district government revenue per capita. As expected, we find a positive, albeit small, correlation between pre-*Ujamaa* government revenue and villagization. We next consider the correlation between the distance to the capital city, as a measure of central government influence, and the intensity of villagization. Given limited control of the hinterland by the state (Herbst, 2014), we might expect that the policy was particularly targeted to the remote parts of the country to bring those areas into the state's orbit. Indeed, we find a positive relationship between distance and our measure.

We next look at the relationship between villagization and ethnolinguistic fractionalization (ELF) in 1967. McHenry (1979) argues that ethnic diversity hindered the population's willingness to comply with the villagization policy, which involved potentially co-inhabiting villages with members of different ethnic groups. Consistent with this argument, we find that the intensity of villagization was lower in districts with high pre-*Ujamaa* ELF.

Geographic characteristics such as topology or climatic zones may have also influenced the degree of villagization. For example, more rugged or remote terrain may have been

harder to consolidate. However, we find no significant correlations between several geographic characteristics (altitude, slope, latitude and longitude) and villagization.

The government also used access to public infrastructure and services as an inducement to encourage people to move to villages. We examine the implications of this in two ways. First, we test whether there is a relationship between the intensity of villagization and existing public health infrastructure, proxied by the number of hospital beds per hundred thousand inhabitants in 1967. We find no significant correlation. Second, we check whether unexpected weather shocks in the form of droughts affected villagization. Several regions in Tanzania experienced severe droughts during the height of the resettlement period (1974–1976). [Hydén \(1980\)](#) argues that the government used these droughts to induce the population to comply with the policy by conditioning drought relief on living in registered villages. Indeed, we find a positive correlation between drought occurrence in 1976 and villagization.<sup>50</sup>

Lastly, in Table A.11 and Figure A.8, we report the association between villagization in 1978 and feelings of national identity in 1967 (i.e., prior to the *Ujamaa* policy) at the region level.<sup>51</sup> The data on national identity in 1967 is based on a nationwide survey of 3,000 secondary students ([Prewitt et al., 1970](#)) and is almost identical in wording to the measure of national identity from the contemporary Afrobarometer data used in the paper. We find a negative relationship between villagization and pre-*Ujamaa* feelings of national identity. This suggests that, in line with the government’s stated intention of building national identity using villagization, the policy was aimed at regions that historically had lower levels of identification with the nation.

---

<sup>50</sup> We measure droughts using the standard deviation in yearly rainfall from a district’s long-term mean rainfall, truncated at 0 (following [Dell et al. \(2014\)](#)). Rainfall is calculated based on readings from 107 weather stations across Tanzania and aggregated to the district level using Kriging (spatial interpolation).

<sup>51</sup> The 1967 survey by [Prewitt et al. \(1970\)](#) does not contain district-level identifiers. We thank Ted Miguel and Tina Green for help with accessing the data.

## Appendix B. Migration

Here we conduct two exercises to show that selective migration is unlikely to explain a large part of the positive relationship between the *Ujamaa* policy and national identity we observe. Recall that only inter-district migration potentially threatens the validity of our estimates.

The Afrobarometer data do not contain information on respondents' migration. However, we can calculate inter-district migration rates, during the period after the *Ujamaa* policy and before the Afrobarometer survey took place, by district and cohort using the Tanzania National Panel Survey 2008 data.<sup>52</sup> That is, we can calculate the in- and out-migration rate for each cohort- and origin-/destination-district-combination. We use the migration matrices constructed this way for the robustness checks related to selective migration.

As a first exercise, in Table 3 column (5) and Table A.12 we drop 11 out of 54 districts that are in the top deciles in terms of either in- or out-migration rates. If (selective) migration were driving our results, we should see a clear drop in our main coefficient when we exclude such high-migration districts. Reassuringly, even though we lose some statistical power due to the reduced sample size, the magnitude of the coefficient with this reduced sample is similar to the coefficient for our baseline specification with the full sample.

As a second exercise, we quantify the potential measurement error stemming from selective migration to provide bounds on our main coefficient under different assumptions about the extent of selective migration. We begin by noting that migration may or may not lead to measurement error in a respondent's villagization measure: for example, if a respondent in the treated cohort is at the time of survey (2005–2008) in a district with a villagization measure of 0.5, but was in a district with a villagization measure of 1 during the villagization period (1970–1981), then the measurement error in her treatment amounts to -0.5. However, if she migrated from a district with a villagization measure of 1 to another district with a villagization measure of 1, the measurement error is 0 and does not bias our estimates.

The total measurement error in the villagization measure among respondents in a district thus depends on two factors: 1) the difference in villagization between the receiving district and the sending district, and 2) the migration rates between each district-pair. To estimate the measurement error in our sample, we proceed as follows. We first calculate the measurement error for each district-pair (receiving and sending) as the product of the difference

---

<sup>52</sup> The Tanzania National Panel Survey data contains variables on each respondent's birth year, current district, previous district, and year of migration (if migrated). We assume that if the respondent migrated between 1982 and 2008, the origin district is the district where she resided during the villagization period.

between the two districts' villagization measures and the rate of migration from the sending to the receiving district:<sup>53</sup>

$$error_{ds} = (V_d - V_s) \cdot m_{ds} \quad (\text{A1})$$

where  $V_d$  and  $V_s$  denote villagization in receiving and sending districts  $d$  and  $s$ , and  $m_{d,s}$  denotes the migration rate from district  $s$  to district  $d$ . Migration from districts with higher villagization thus leads to negative measurement error in the villagization measure, while migration from districts with lower villagization leads to positive measurement error. The larger the migration rate from a district, the larger the potential measurement error due to migration from that district.

Next, we make the extreme “worst case” assumption that would lead to selective migration working against our main finding: assume that all residents in the current district that came from districts with lower villagization identify only with the nation, and that all residents in the current district that came from districts with higher villagization identify only with their ethnic group. We adjust the villagization measures of the respondents in our main sample by these “worst case” measurement errors:

$$\tilde{V}_{id} = V_d - \sum_{s=1}^S error_{ds} \cdot \mathbb{1}[error_{ds} > 0] \cdot \mathbb{1}[y_{id} = 1] - \sum_{s=1}^S error_{ds} \cdot \mathbb{1}[error_{ds} < 0] \cdot \mathbb{1}[y_{id} = 0] \quad (\text{A2})$$

where  $\tilde{V}_{id}$  denotes individual  $i$  in district  $d$ 's “error-adjusted” villagization measure and  $y_{id}$  denotes national identity (0 = identifies only with ethnic group, 1 = identifies only with nation as a whole).

Finally, we re-estimate our main specification using this adjusted villagization measure:

$$y_{idzt} = \gamma(\tilde{V}_{id} \cdot treatedcohort_t) + (\mathbf{X}'_{dz} \cdot treatedcohort_t)\boldsymbol{\Gamma} + \alpha_{dz} + \delta_{zt} + \epsilon_{idzt} \quad (\text{A3})$$

The coefficient estimate of  $\gamma$  then provides a lower bound to our baseline coefficient estimate of  $\beta$  in the presence of most extreme selective migration. Table 3 column (6) displays the result: the estimated coefficient equals 0.145. As a reminder, our baseline coefficient estimate equals 0.243. That is, at the very most 40% of our baseline coefficient estimate can be explained by selective migration. However, this extreme worst case is very unlikely. It assumes that all individuals in the treated cohort that migrated to low-villagization districts after villagization ended identify only with their ethnic group,

---

<sup>53</sup> For example, assume that district  $A$  has a villagization measure of 0.8 and district  $B$  has a villagization measure of 0.6. Further assume that 5% of current residents in district  $A$  migrated from district  $B$ . Then, the measurement error in the villagization measure among the current residents of district  $A$  stemming from migration from district  $B$  equals  $(0.8 - 0.6) \cdot 0.05 = 0.01$ .

and that all individuals in the treated cohort that migrated to high-villagization districts after villagization ended identify only with Tanzania as a whole. Alternatively, assume that only those with low national identity migrated to low-villagization districts after the end of the *Ujamaa* period, whereas there was no selective migration among those with high national identity (that is, they did not systematically go to districts with higher villagization after villagization ended). In this case, the adjusted coefficient estimate, reported in Table 3 column (7), equals 0.207 and, like the baseline coefficient, is statistically significant at the 5% level. That is, under this more realistic assumption, at most 15% of our main result can be explained by selective migration. In sum, under reasonable assumptions, it is unlikely that selective migration explains a large fraction of our main result.

## Appendix C. Additional Tables

Table A.1: Timeline

Year	Event	Details
1961	Independence	Tanganyika (Tanzania Mainland) becomes independent with Julius Nyerere as prime minister.
1963	One-party state declared	Nyerere makes TANU the sole national party in the interest of the whole nation and economic development ( <a href="#">Nyerere, 1963</a> ).
1967	National vision outlined	<i>Arusha Declaration</i> as basis for national ideology; statements by party and prime minister announcing plans for nation-building policy and <i>Ujamaa</i> villages.
1967	<i>Education for Self-Reliance</i>	Formed basis of <i>Ujamaa</i> education reforms; Swahili made language of instruction in public primary schools.
1967	1967 Population Census	First population census after independence conducted; administrative districts defined.
1968	Roll-out of new education policies	Political education added as a new subject to primary school curriculum (grades 4 to 7); new syllabi and textbooks emphasizing national identity introduced in social studies, history, geography and civics.
1969	<b>Presidential Circular No. 1</b>	<b>Mandatory villagization begins.</b>
1975	<i>Villages &amp; Ujamaa Villages Act</i>	Villagization formalized in the law.
1978	1978 Population Census	First census after villagization was declared concluded.
1982	<b>Repeal of <i>Villages &amp; Ujamaa Villages Act</i></b>	<b>Villagization period ends.</b>
1985	Nyerere steps down	Ali Hassan Mwinyi replaces Nyerere as prime minister, abandons many of Nyerere's policies and enacts reforms.

*Note:* This table provides a timeline of activities by Nyerere and the Tanzanian state after Independence and during the *Ujamaa* period. See sources described in Section 2 for additional details.

Table A.2: District-level Variables: Sources and Definitions

Variable	Source	Details and Coding
Villagization 1978	Population Census	Share of total rural district population living in registered villages (as opposed to unregistered villages, scattered and migratory population), as classified by the 1978 population census.
Share urban 1967	Population Census	Share of total district population living in urban wards, as classified by the 1967 population census.
Primary school enrollment rate in 1967	Jensen (1968)	Number of children enrolled in primary schools per one thousand inhabitants according to official government statistics (Jensen, 1968).
District revenue in 1967	Jensen (1968)	Total district revenue in thousands of Tanzanian Shillings according to official government statistics (Jensen, 1968).
Ethnolinguistic fractionalization in 1967	Population Census	Calculated as 1 minus the Herfindahl concentration index of ethnolinguistic group shares in each district given by the 1967 population census data.
Distance to Dar es Salaam	ESRI	Distance of the district centroid to Tanzania's capital city in kilometers.
Centroid latitude / longitude	ESRI	Latitude / longitude of the district centroid in degrees.
Average altitude	ESRI	Mean district altitude in meters, calculated from gridded topology data.
Average slope	ESRI	Mean slope, calculated from gridded topology data.
Hospital beds in 1968	Jensen (1968)	Number of hospital beds per 100,000 inhabitants according to official government statistics (Jensen, 1968).
Drought in 1974-1976	Tanzania Meteo	Average amount of yearly rainfall per district, in terms of standard deviations from the long-run district mean (1960-2010), truncated at 0. Data constructed from readings of 107 weather stations using Kriging (spatial interpolation).
National identity in 1967	Prewitt et al. (1970)	Based on a nationwide survey of almost 3,000 Tanzanian secondary school students in 1967 (Prewitt et al., 1970). The original variable is coded as 1 if a respondent feels that the nation is more important than the tribe and 0 otherwise.
Linguistic distance	LEDA	Average linguistic distance from each Afrobarometer respondent's ethnic group to Swahili. Extracted from <i>Linking Ethnic Data from Africa</i> (LEDA) project database. See <a href="https://github.com/carl-mc/LEDA">https://github.com/carl-mc/LEDA</a> for a description of the data.
In- / out-migration rate	TNPS	Fraction of district population that was born in a different district than the current district / was born in a district but lived in a different district at the time of the survey. Calculated from individual-level data on household heads in the TNPS Round 1 (2008).

Note: This table provides the key district-level variables, their sources and coding definitions used in our empirical analysis. There are 54 districts in our main sample. Section 3 provides additional discussion.

Table A.3: Individual-level Variables: Sources and Definitions

Variable	Source	Details and Coding
National identity	Afrobarometer	"Let us suppose that you had to choose between being a [nationality] and being a [respondent's identity group]. Which of the following statements best expresses your feelings?" 0 = ethnic only, 1/4 = ethnic more than national, 1/2 = ethnic and national equal, 3/4 = national more than ethnic, 1 = national only.
School cohort	Afrobarometer	Constructed from birth year of respondent plus 10 years (typical age of 4th grader). "Treated cohort" = born 1960-1971, "control cohort" = born 1948-1959.
Completed school	Afrobarometer	"What is the highest level of education you have completed?" 1 = completed at least some formal primary schooling, 0 = completed no formal schooling.
Respect authority	Afrobarometer	"Let's talk for a moment about the kind of society we would like to have in this country. Which of the following statements is closest to your view? Choose Statement A or Statement B. A: As citizens, we should be more active in questioning the actions of our leaders. B: In our country these days, we should show more respect for authority." 0 = agree very strongly with A, 1/4 = agree with A, 1/2 = agree with neither, 3/4 = agree with B, 1 = agree very strongly with B.
Support one-party rule	Afrobarometer	"There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office." 0 = strongly disapprove, 1/4 = disapprove, neither approve nor disapprove, 3/4 = approve, 1 = strongly approve.
See government as parent	Afrobarometer	"Let's talk for a moment about the kind of society you would like to have in this country. Which of the following statements is closest to your view? Choose Statement 1 or Statement 2. Do you agree or agree very strongly? Statement 1: People are like children; the government should take care of them like a parent. Statement 2: Government is like an employee; the people should be the bosses who control the government." 0 = agree very strongly with 2, 1/4 = agree with 2, 1/2 = agree with neither, 3/4 = agree with 1, 1 = agree very strongly with 1.
Took action to hold government accountable	Afrobarometer	"Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year. If not, would you do this if you had the chance: Got together with others to raise an issue? Attended a demonstration or protest march?" 0 = if has never done any of these things, 1 = has done at least one of these things at least once.
Trust government / independent newspapers / TV / radio	Afrobarometer	"How much do you trust each of the following, or haven't you heard enough about them to say: Government / independent newspapers / TV / radio?" 0 = not at all, 1/3 = just a little, 2/3 = somewhat, 1 = a lot.
Main occupation	TNPS	Dummies for different occupations that equal 1 if household head's occupation is in government / private sector / agriculture / self-employment, and 0 otherwise.
Married within ethnic group	DHS	= 1 if respondent's spouse shares self-reported ethnic group, 0 otherwise. Includes married couples only.

Note: This table provides the key individual-level variables, their sources and coding definitions used in our empirical analysis. Section 3 provides additional discussion.

Table A.4: The Effect of *Ujamaa* on National Identity, All Cohorts

	Dependent Variable: National vs. Ethnic Identity							
no zone-cohort FE	baseline	controlling for treated cohort dummy interacted with						weather shocks
	(1)	(2)	(3)	(4)	ethnic fractionalization	geographical features	hospital beds	
Villagization × Birth Year 1948-1951	0.047 (0.096)	-0.047 (0.143)	-0.034 (0.154)	-0.016 (0.140)	0.036 (0.150)	0.065 (0.127)	-0.046 (0.138)	-0.031 (0.129)
Villagization × Birth Year 1952-1955	-0.042 (0.131)	0.029 (0.154)	0.034 (0.160)	0.027 (0.132)	0.085 (0.152)	0.121 (0.190)	-0.201 (0.127)	0.065 (0.131)
Villagization × Birth Year 1960-1963	0.130 (0.116)	0.119 (0.146)	0.119 (0.143)	0.152 (0.153)	0.062 (0.157)	0.163 (0.130)	0.095 (0.153)	0.115 (0.115)
Villagization × Birth Year 1964-1967	0.132 (0.102)	0.219 (0.143)	0.190 (0.148)	0.265* (0.134)	0.206 (0.152)	0.171 (0.126)	0.182 (0.113)	0.217 (0.149)
Villagization × Birth Year 1968-1971	0.157* (0.090)	0.211** (0.084)	0.218** (0.087)	0.234*** (0.083)	0.184** (0.089)	0.253*** (0.092)	0.195** (0.078)	0.221*** (0.093)
Villagization × Birth Year 1972-1975	0.085 (0.085)	0.112 (0.091)	0.101 (0.100)	0.160* (0.095)	0.087 (0.089)	0.099 (0.106)	0.093 (0.085)	0.152 (0.106)
Villagization × Birth Year 1976-1979	0.042 (0.093)	0.005 (0.149)	-0.006 (0.151)	0.022 (0.151)	-0.058 (0.150)	0.068 (0.167)	0.055 (0.143)	0.010 (0.163)
Villagization × Birth Year 1980-1983	0.045 (0.071)	0.039 (0.079)	0.053 (0.084)	0.083 (0.068)	0.048 (0.076)	0.115* (0.067)	-0.002 (0.064)	0.051 (0.081)
Number of Observations	1,644	1,644	1,644	1,644	1,644	1,644	1,644	1,644
Number of Districts	54	54	54	54	54	54	54	54
R-squared	0.090	0.118	0.122	0.124	0.125	0.144	0.129	0.129
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

*Note:* The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to birth cohort  $t$ . All columns regress the measure of an individual's current national identity on the interactions between that individual's district-level measure of historical villagization and dummies that indicate the individual's cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummies, the 1967 district primary schooling enrollment rate interacted with the cohort dummies, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recorded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). The reference group is the cohort born in 1956-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. All district-level variables and the outcome variable are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.5: The Effect of *Ujamaa* on National Identity, Unweighted

	Dependent Variable: National vs. Ethnic Identity							
	controlling for treated cohort dummy interacted with							
	no zone-cohort FE	baseline	distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Treated Cohort	0.157** (0.073)	0.221** (0.089)	0.211** (0.091)	0.226** (0.087)	0.186** (0.089)	0.180** (0.069)	0.224** (0.093)	0.199** (0.077)
Number of Observations	862	862	862	862	862	862	862	862
Number of Districts	54	54	54	54	54	54	54	54
R-squared	0.098	0.110	0.111	0.110	0.111	0.121	0.110	0.115
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

*Note:* The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to birth cohort  $t$ . All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in  $1/4$  increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. All district-level variables and the outcome variable are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.6: The Effect of *Ujamaa* on National Identity, Controlling for Whether Respondent Believes Government Sent Surveyor

	Dependent Variable: National vs. Ethnic Identity							
	controlling for treated cohort dummy interacted with							
	no zone-cohort FE	baseline	distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Treated Cohort	0.181** (0.085)	0.238** (0.114)	0.223* (0.117)	0.246** (0.114)	0.191* (0.110)	0.166* (0.087)	0.236** (0.117)	0.197** (0.097)
Number of Observations	862	862	862	862	862	862	862	862
Number of Districts	54	54	54	54	54	54	54	54
R-squared	0.106	0.122	0.125	0.123	0.124	0.135	0.122	0.130
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE		✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE			✓	✓	✓	✓	✓	✓

*Note:* The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to birth cohort  $t$ . All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. All regressions control for an individual-level dummy variable that equals 1 if the respondent erroneously believes that a representative of a Tanzanian government entity sent the Afrobarometer survey. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.7: The Effect of *Ujamaa* on National Identity, Rural Sample Only

	Dependent Variable: National vs. Ethnic Identity							
	controlling for treated cohort dummy interacted with							
	no zone-cohort FE	baseline	distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Treated Cohort	0.239** (0.105)	0.393** (0.170)	0.389** (0.172)	0.409** (0.163)	0.333* (0.176)	0.273** (0.135)	0.342* (0.185)	0.342** (0.138)
Number of Observations	714	714	714	714	714	714	714	714
Number of Districts	49	49	49	49	49	49	49	49
R-squared	0.126	0.147	0.149	0.150	0.148	0.164	0.148	0.156
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

*Note:* The sample includes only Afrobarometer respondents who resided in rural areas during the survey. The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to birth cohort  $t$ . All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummy; the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. All district-level variables and the outcome variable are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.8: The Effect of *Ujamaa* on National Identity, By Ethnolinguistic Fractionalization

	Dependent Variable: National vs. Ethnic Identity								
	controlling for treated cohort dummy interacted with								
	no zone-cohort FE		baseline		geographical features			hospital beds	weather shocks
	(1)	(2)			(4)	(5)	(6)		(7)
				distance to capital	district revenue	geographical features	hospital beds		weather shocks
Villagization × Treated Cohort	0.130 (0.471)	0.123 (0.526)		0.127 (0.558)	0.085 (0.497)	0.683 (0.483)	0.391 (0.564)	0.071 (0.563)	
Treated Cohort × ELF	1.232 (1.217)	1.342 (1.169)		3.384** (1.349)	1.766 (1.383)	4.945 (14.01)	2.059* (1.101)	-0.866 (1.843)	
Villagization × Treated Cohort × ELF	0.060 (0.699)	0.014 (0.699)		0.062 (0.767)	0.0315 (0.684)	-0.743 (0.728)	-0.480 (0.766)	0.067 (0.846)	
Number of Observations	862	862		862	862	862	862	862	
Number of Districts	54	54		54	54	54	54	54	
R-squared	0.113	0.125		0.130	0.129	0.150	0.127	0.134	
District FE		✓		✓	✓	✓	✓	✓	
Cohort FE		✓		✓	✓	✓	✓	✓	
Zone-Cohort FE		✓		✓	✓	✓	✓	✓	

*Note:* The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to school cohort  $t$ . All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, the individual's district-level measure of historical ethnolinguistic fractionalization (ELF), all double and triple interactions of these three variables, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The variables varying at the district- or cohort-level only are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole, 0 = respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. All district-level variables and the outcome variable are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.9: The Effect of *Ujamaa* on Intermarriage, By Ethnolinguistic Fractionalization

	Dependent Variable: Married Same Ethnicity						
	no zone-cohort FE		baseline		controlling for treated cohort dummy interacted with		
	(1)	(2)	distance to capital (3)	district revenue (4)	geographical features (5)	hospital beds (6)	weather shocks (7)
Villagization × Treated Cohort	-0.069 (0.042)	-0.104** (0.032)	-0.106** (0.044)	-0.106** (0.048)	-0.128*** (0.045)	-0.116** (0.049)	-0.115** (0.048)
Treated Cohort × ELF	0.064 (0.072)	0.063 (0.077)	0.160 (0.099)	0.123 (0.097)	1.815** (0.695)	0.113 (0.079)	-0.053 (0.103)
Villagization × Treated Cohort × ELF	0.058 (0.045)	0.075 (0.056)	0.069 (0.049)	0.066 (0.057)	0.126** (0.048)	0.054 (0.056)	0.099 (0.060)
Number of Observations	1,192	1,192	1,192	1,192	1,192	1,192	1,192
Number of Districts	57	57	57	57	57	57	57
R-squared	0.114	0.117	0.120	0.118	0.123	0.119	0.120
District FE	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE			✓	✓	✓	✓	✓

*Note:* The unit of observation is a married, male DHS respondent  $i$  in district  $d$  belonging to school cohort  $t$ . All columns regress a dummy for whether a respondent's spouse is of the same ethnic group as the respondent on the interaction between the respondent's district-level measure of historical villagization and a dummy that indicates whether the respondent is in the treated cohort, the respondent's district-level measure of historical ethnolinguistic fractionalization (ELF), all double and triple interactions of these three variables, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The variables varying at the district- or cohort-level only are included in the fixed effects in all columns. The dependent variable is from DHS 1991 and 1996 (later rounds did not report respondents' ethnicity). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. The villagization variable is standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.10: The Effect of *Ujamaa* on Occupation

	Dependent Variable: Occupation			
	government	private sector	agriculture	self-employed
	(1)	(2)	(3)	(4)
Villagization × Treated Cohort	-0.065 (0.047)	0.023 (0.064)	0.043 (0.068)	-0.044 (0.054)
Number of Observations	1,190	1,190	1,190	1,190
Number of Districts	59	59	59	59
R-squared	0.087	0.185	0.064	0.133
District FE	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓
Zone-Cohort FE	✓	✓	✓	✓

Note: The unit of observation is an individual Tanzania National Panel Survey Round 1 (2008) respondent  $i$  in district  $d$  belonging to school cohort  $t$ . All columns regress a dummy for the occupation listed in the column head on the interaction between an individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). All variables are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the household survey weights provided by the data. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.11: Correlations Between Villagization and District Characteristics in 1967

1967 district characteristic	correlation	p-value
Share urban in 1967 (% of population)	0.14	0.32
Primary school enrollment rate in 1967 (per '000 inhabitants)	0.06	0.74
District revenue in 1967 ('000 shs)	0.07	0.60
Ethnolinguistic fractionalization in 1967	-0.17	0.08
Distance to Dar-Es-Salaam in kilometers	0.29	0.09
Centroid latitude	0.13	0.69
Centroid longitude	0.09	0.60
Average altitude in meters	-0.09	0.69
Average slope	0.17	0.26
Hospital beds in 1968 (per '000.000 inhabitants)	-0.16	0.17
Drought in 1974 (truncated z-score)	-0.07	0.72
Drought in 1975 (truncated z-score)	0.13	0.55
Drought in 1976 (truncated z-score)	0.32	0.02
National identity in 1967 (region-level)	-0.55	0.04

*Note:* This table shows pairwise correlations between villagization (share of total rural district population living in registered villages in 1978) and the variables indicated in the table, conditional on zone fixed effects. The reported p-values are robust to hetero-skedasticity. The data includes the districts in the main sample as in Table 2. The variables and their sources are described in detail in Section 3 and Appendix A.

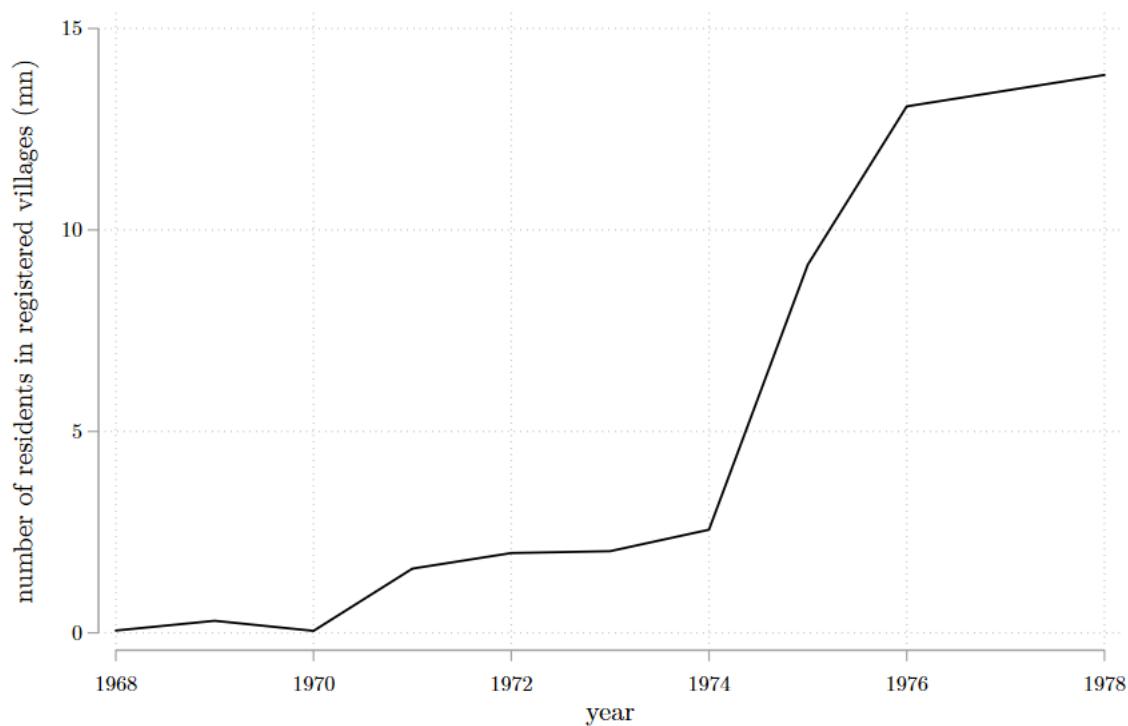
Table A.12: The Effect of *Ujamaa* on National Identity, Dropping Districts with Highest Migration Rates

	Dependent Variable: National vs. Ethnic Identity							
	controlling for treated cohort dummy interacted with							
	no zone-cohort FE	baseline	distance to capital	district revenue	ethnolinguistic fractionalization	geographical features	hospital beds	weather shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Villagization × Treated Cohort	0.135 (0.081)	0.206* (0.115)	0.186* (0.107)	0.218* (0.120)	0.189 (0.125)	0.177** (0.077)	0.196 (0.118)	0.181** (0.082)
Number of Observations	725	725	725	725	725	725	725	725
Number of Districts	43	43	43	43	43	43	43	43
R-squared	0.091	0.101	0.109	0.101	0.101	0.122	0.101	0.114
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Zone-Cohort FE		✓	✓	✓	✓	✓	✓	✓

*Note:* All districts with an in- or out-migration rate in the highest deciles are dropped from the sample (migration rates are calculated from the Tanzania National Panel Survey 2008). The unit of observation is an individual Afrobarometer respondent  $i$  in district  $d$  belonging to birth cohort  $t$ . All columns regress the measure of an individual's current national identity on the interaction between that individual's district-level measure of historical villagization and a dummy that indicates whether the individual is in the treated cohort, controlling for the 1967 district population share in urban areas interacted with the cohort dummy, the 1967 district primary schooling enrollment rate interacted with the cohort dummy, and the other controls and fixed effects indicated in the table. The un-interacted variables are included in the fixed effects in all columns. The dependent variable is from Afrobarometer rounds 3 and 4 (2005-2008), and was originally on a five-level Likert scale; it has been recoded from 0 to 1 in 1/4 increments (1 = respondent identifies only with the nation as a whole; 0 = respondent identifies only with her ethnic group). Treated cohort is a dummy that equals 1 if the respondent was born in 1960-1971. The reference group is the cohort born in 1948-1959. The villagization variable is the share of the respondent's current district's rural population that lived in official government villages by 1978 according to the 1978 Tanzania Population Census. All controls, interacted with the individual-level cohort dummy, are at the district level and based on the 1967 Census (except weather shocks and geographical features). Geographical features include latitude, longitude, altitude and slope. Weather shocks are measured as z-scores of rainfall in 1974, 1975 and 1976 (main period of forced relocation), truncated at 0 to capture droughts. The weather data is from the Tanzania Meteorological Service and based on readings from 107 weather stations which are spatially interpolated at the district level using Kriging. All district-level variables and the outcome variable are standardized to have a mean of 0 and a standard deviation of 1 to facilitate interpretation. All regressions are weighted using the survey weights provided by Afrobarometer. Standard errors are clustered at the district level and reported in parentheses. Statistical significance is represented by \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

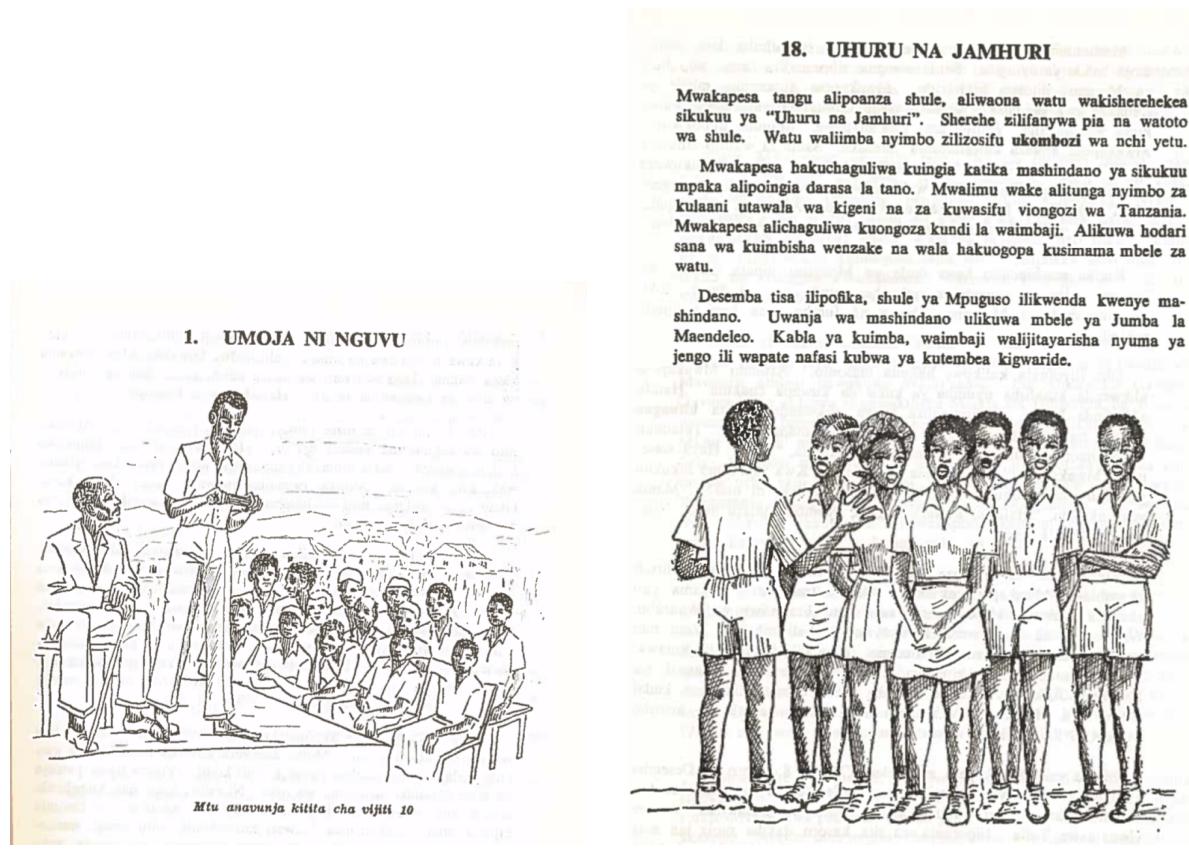
## Appendix D. Additional Figures

Figure A.1: Population in Registered Villages Over Time



Note: Authors' illustration based on data from [Shao \(1982\)](#). The implementation of the villagization program started after the Presidential Circular No. 1 in 1969, was accelerated in 1974 when villagization was declared mandatory, and formally ended with the repeal of the Villages and *Ujamaa* Villages Act in 1982.

Figure A.2: Excerpts from Swahili Textbook



Note: These images depict pages from Chapter 1 (left) on *Unity is Strength* and Chapter 18 (right) on *Independence and the Republic* of the Swahili textbook *Tujifunze Lughu Yetu: Kitabu Cha Saba* translated as *Let us learn our language: Class Seven* ([United Republic of Tanzania National Department of Education, 1971b](#)).

Figure A.3: Political Education Textbook for 4<sup>th</sup> Graders

<b>YALIYOMO</b>		<b>CONTENTS</b>
<b>Sura</b>		<b>Ukurasa</b>
<b>1</b>	Kazi ya Serikali ya TANU Mkoani . . . . .	1
<b>2</b>	Halmashauri ya Maendeleo Mkoani . . . . .	5
<b>3</b>	Mipango ya Maendeleo Mkoani . . . . .	8
<b>4</b>	Uchaguzi wa Mwenyekiti wa TANU wa Mkoa . . . . .	11
<b>5</b>	Madaraka ya Mwenyekiti wa TANU wa Mkoa . . . . .	13
<b>6</b>	Uwakilishaji wa Mkoa katika Kamati Kuu ya TANU . . . . .	14
<b>7</b>	Katibu wa TANU wa Mkoa . . . . .	16
<b>8</b>	Majimbo ya Uchaguzi wa Baraza la Taifa Mkoani . . . . .	18
<b>9</b>	Shughuli za Utamaduni Mkoani . . . . .	20
<b>10</b>	Vyama vya Ushirika Mkoani . . . . .	22
<b>11</b>	Alama za Taifa letu . . . . .	25
<b>12</b>	Sikukuu za Taifa . . . . .	30
<b>13</b>	Uhitaji wa kuwa na Serikali . . . . .	34
<b>14</b>	Wajibu wa Raia . . . . .	36
<b>15</b>	Historia ya TANU . . . . .	38
<b>16</b>	Historia ya Chama cha AFRO-SHIRAZ . . . . .	42
<b>17</b>	Uchaguzi wa Rais na Wabunge . . . . .	46
<b>18</b>	Kazi za Rais . . . . .	50
<b>19</b>	Serikali Kuu . . . . .	53
<b>20</b>	Kazi za Baraza la Taifa . . . . .	56
<b>21</b>	Serikali ya Mitaa na jinsi zinavyosaidia Serikali kuu . . . . .	59

1. The work of the TANU government in the regions
2. Regional development councils
3. Regional development plans
4. Election of TANU leaders in regions
5. Powers of TANU leaders in regions
6. Regional representatives in TANU Central Committee
7. TANU's provincial secretary
8. Constituencies for national elections
9. Services in the regions
10. Participatory committees in the regions
11. Symbols of our nation
12. National Holidays
13. The need to support the government
14. Civic duties
15. History of TANU
16. History of the AFRO-SHIRAZ party
17. Election of the President and Parliament
18. Work of the President
19. Central government
20. Work of the central council
21. Local governments and how they help the central government

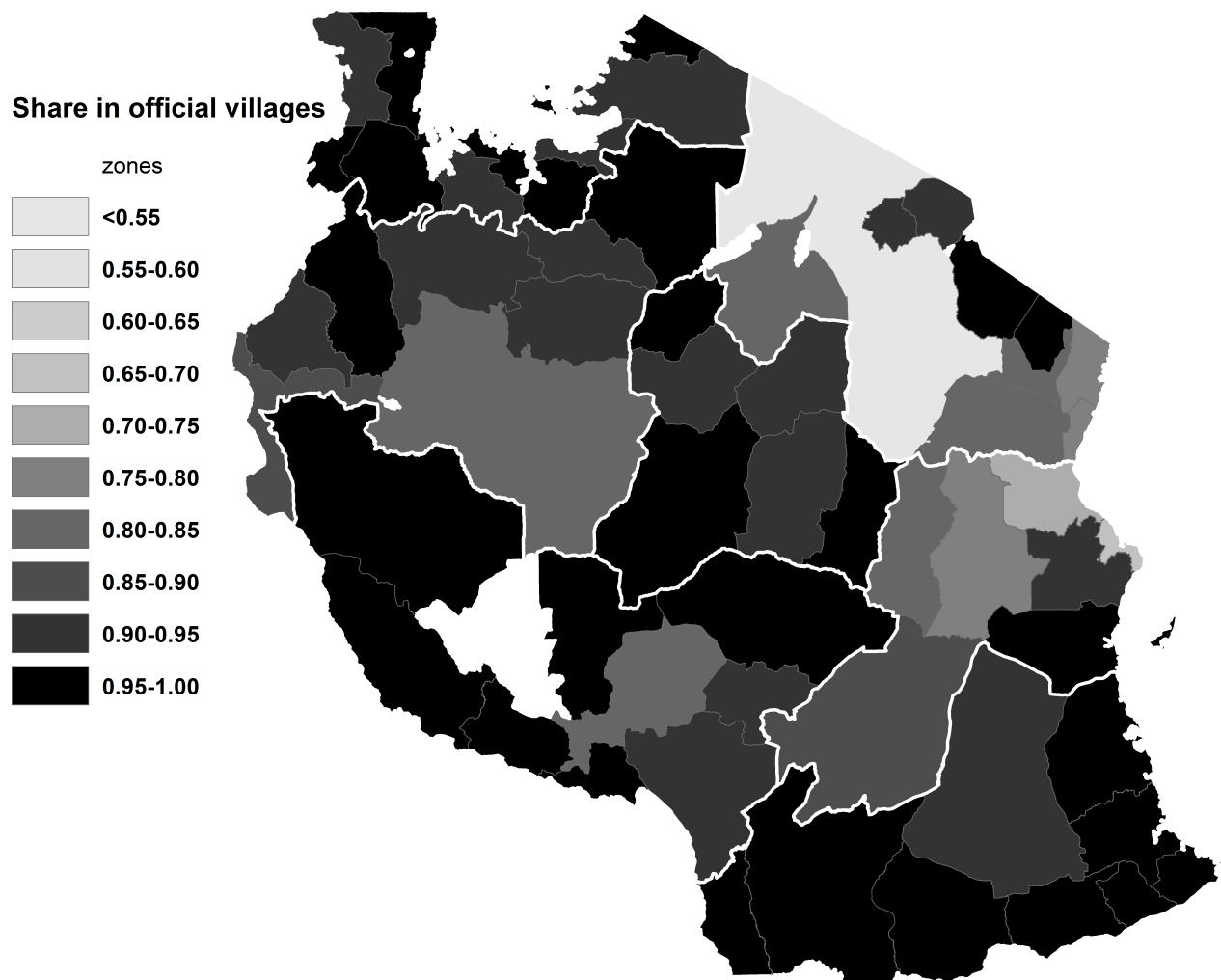
*Note:* This image depicts the contents page of a political education textbook for 4<sup>th</sup> graders. The left figure is the contents page in Swahili and the right figure is the equivalent English translation by the authors. *Sura* translates to chapter and *Ukurasa* to pages. Items underlined are the specific elements of nation-building. *Source:* Textbook titled *Elimu ya Siasa* ([United Republic of Tanzania National Department of Education, 1971a](#)).

Figure A.4: Population Census 1978, Share of Rural Population in Official Registered Villages

TABLE 15 DISTRIBUTION OF PRIVATE HOUSEHOLDERS BY LOCATION (IN VILLAGES/OUTSIDE VILLAGES/IN URBAN AREAS) AND CITIZENSHIP OF HEAD OF HOUSEHOLD FOR WARDS, DISTRICTS, AND REGIONS: 1978 CENSUS								PAGE 114			
REGION: 02 ARUSHA			L	O	C	A	T	I	O	N	C I T I Z E N S H I P
WARD CODE	WARD NAME		IN VILLAGES	OUTSIDE VILLAGES	URBAN AREAS	TOTAL	TANZANIAN CITIZENS		NON-TANZANIANS		
02 5 011	KIRU	(RURAL)	2876	146	0	3022	3008	14			
02 5 021	GALAPO	(RURAL)	2899	43	0	2942	2933	9			
02 5 031	MAMIRE	(RURAL)	1417	26	0	1443	1429	14			
02 5 041	BONGA	(RURAL)	1314	62	0	1376	1368	8			
02 5 051	GIDA	(RURAL)	1507	8	0	1515	1515	0			
02 5 061	RIRODA	(RURAL)	2989	223	0	3212	3212	0			
02 5 071	SINGE	(RURAL)	2103	0	0	2103	2081	22			
02 5 081	MASAKTA	(RURAL)	1280	85	0	1365	1364	1			
02 5 091	ENDASAK	(RURAL)	2321	4	0	2325	2321	6			
02 5 101	GIDAHABABIEK	(RURAL)	516	116	0	632	632	0			
02 5 111	SIROP	(RURAL)	786	235	0	1021	1021	0			
02 5 121	KATESH	(RURAL)	1923	329	0	2252	2248	4			
02 5 131	BALANGDALALU	(RURAL)	93	455	0	548	548	0			
02 5 141	GEHANDU	(RURAL)	495	321	0	816	815	1			
02 5 151	BASSOTU	(RURAL)	1451	655	0	2106	2080	26			
02 5 161	BASSODESH	(RURAL)	625	197	0	822	822	0			
02 5 171	GENDABI	(RURAL)	418	134	0	552	550	2			
02 5 181	GITTING	(RURAL)	800	26	0	826	826	0			
02 5 191	UFANA	(RURAL)	920	9	0	929	925	4			
02 5 201	MANDUNGA	(RURAL)	1246	3	0	1249	1245	4			
02 5 211	BASHANET	(RURAL)	1304	13	0	1317	1314	3			
02 5 221	DABIL	(RURAL)	1710	0	0	1710	1710	0			
02 5 231	DAREDA	(RURAL)	2567	0	0	2567	2561	6			
02 5 241	NKAIKI	(RURAL)	763	3	0	766	753	13			
02 5 251	MWADA	(RURAL)	1364	0	0	1364	1363	1			
02 5 261	MAGARA	(RURAL)	1028	53	0	1081	1051	30			
02 5 271	MAGUGU	(RURAL)	2605	126	0	2729	2709	20			
02 5 282	BABATI	(URBAN)	0	0	2081	2081	2070	11			
DISTRICT: 025		HANANG	39320	3270	2081	44671	44474	197			
		RURAL COMPONENT	39320	3270	0	42590	42404	186			
		URBAN COMPONENT	0	0	2081	2081	2070	11			
02 6 011	GEHANDU	(RURAL)	1241	0	0	1241	1241	0			
02 6 021	BARGISH	(RURAL)	1465	0	0	1465	1464	1			
02 6 031	HAREABI	(RURAL)	347	0	0	347	347	0			
02 6 041	DAUDI	(RURAL)	1227	0	0	1227	1227	0			
02 6 051	KANSAY	(RURAL)	1316	0	0	1316	1315	1			
02 6 061	DONGOBESH	(RURAL)	2074	3	0	2077	2075	2			
02 6 071	TUMATI	(RURAL)	1335	0	0	1335	1335	0			
02 6 081	MAGHANG	(RURAL)	1738	0	0	1738	1738	0			
02 6 091	HAYDOM	(RURAL)	1624	395	0	2019	2010	9			
02 6 101	YAEBA CHINI	(RURAL)	1145	379	0	1524	1524	0			
02 6 111	SONU	(RURAL)	1394	0	0	1394	1394	0			
02 6 121	TLAMI	(RURAL)	900	0	0	900	900	0			
02 6 131	KALINAM	(RURAL)	1095	4	0	1099	1099	0			
02 6 141	MURRAY	(RURAL)	1369	5	0	1374	1374	0			
02 6 152	MBULU MJINI	(URBAN)	0	0	736	736	733	3			
02 6 161	KARATU	(RURAL)	2716	691	0	3407	3396	11			
02 6 171	RHOFTA	(RURAL)	1449	212	0	1661	1661	0			
02 6 181	MBULU MBULU	(RURAL)	1528	2	0	1530	1530	0			
02 6 191	ENDABASH	(RURAL)	2500	50	0	2550	2546	4			
02 6 201	MANG'OLA	(RURAL)	1275	27	0	1302	1302	0			
02 6 211	OLDEANI	(RURAL)	1552	1158	0	2710	2699	11			
DISTRICT: 026		MBULU	29290	2926	736	32952	32910	42			
		RURAL COMPONENT	29290	2926	0	32216	32177	39			
		URBAN COMPONENT	0	0	736	736	733	3			
REGION: 02		ARUSHA	140491	16457	18877	175826	173999	1827			
		RURAL COMPONENT	140491	16457	0	156949	155632	1317			
		URBAN COMPONENT	0	0	18877	18877	18367	510			

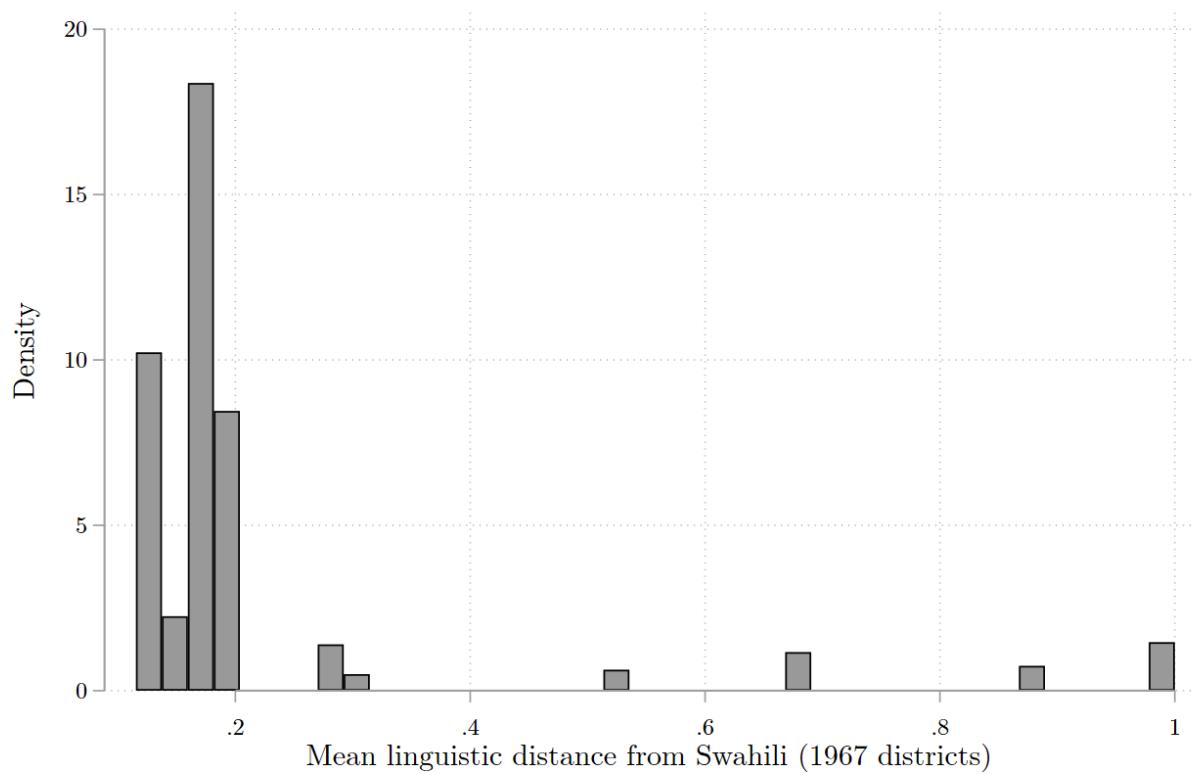
Note: This image shows an excerpt from the 1978 Tanzania Population Census, Volume 6 ([Bureau of Statistics, Ministry of Planning and Economic Affairs, 1981](#)), retrieved from the Herskovits Library of African Studies at Northwestern University. Each row shows the number of residents in and outside official villages (for rural areas) and in urban areas for various wards of Arusha region.

Figure A.5: Distribution of Villagization



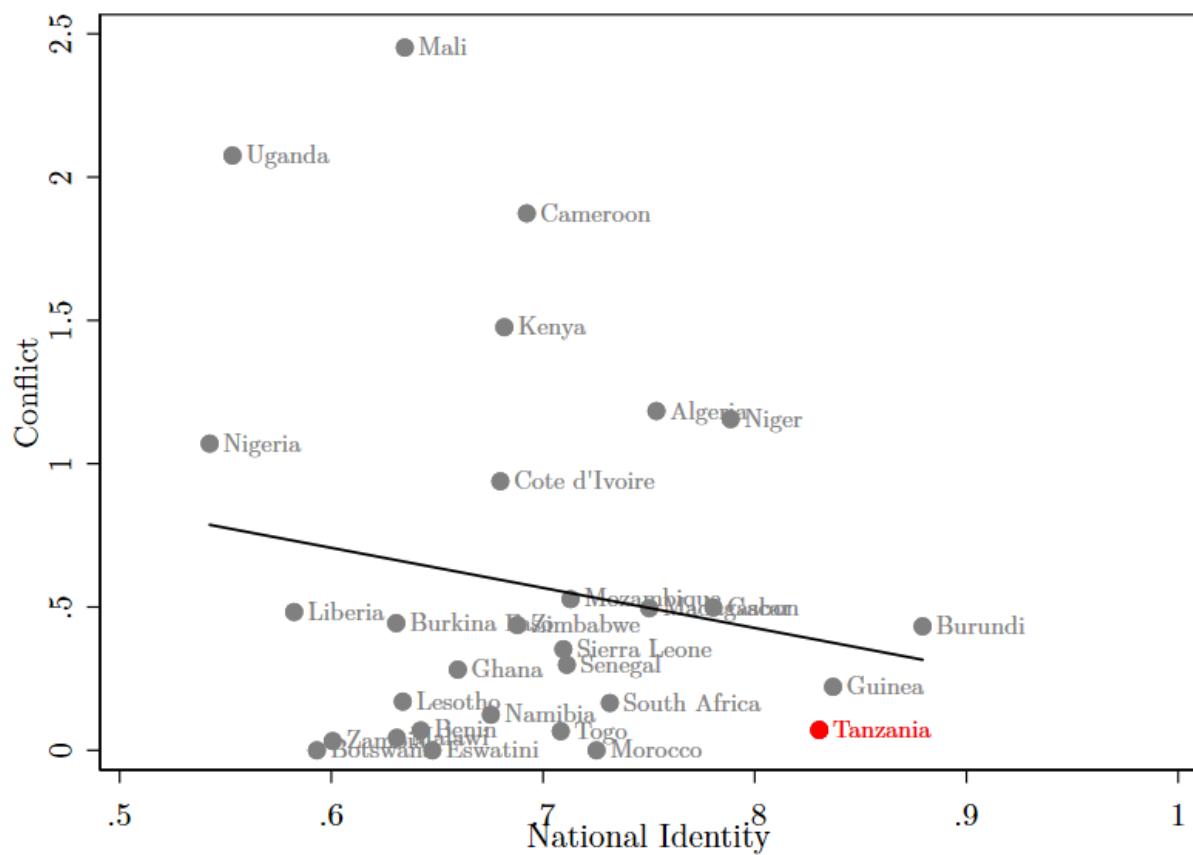
*Note:* This figure depicts our district-level measure of villagization: the share of the rural population living in official registered villages according to the 1978 Tanzania Population Census (at 1967 district borders). The population not living in official registered villages includes those in unregistered villages and those outside of villages (scattered and migratory population). Thick white lines denote zone boundaries. Data for Songwe district is missing. Dar es Salaam and islands (Zanzibar, Pemba, Mafia) are excluded from the sample. Sources: Authors' illustration. Base map from The Humanitarian Data Exchange (HDX) by United Nations OCHA (<https://data.humdata.org/>), accessed July, 2021).

Figure A.6: Linguistic Distance to Swahili in 1967



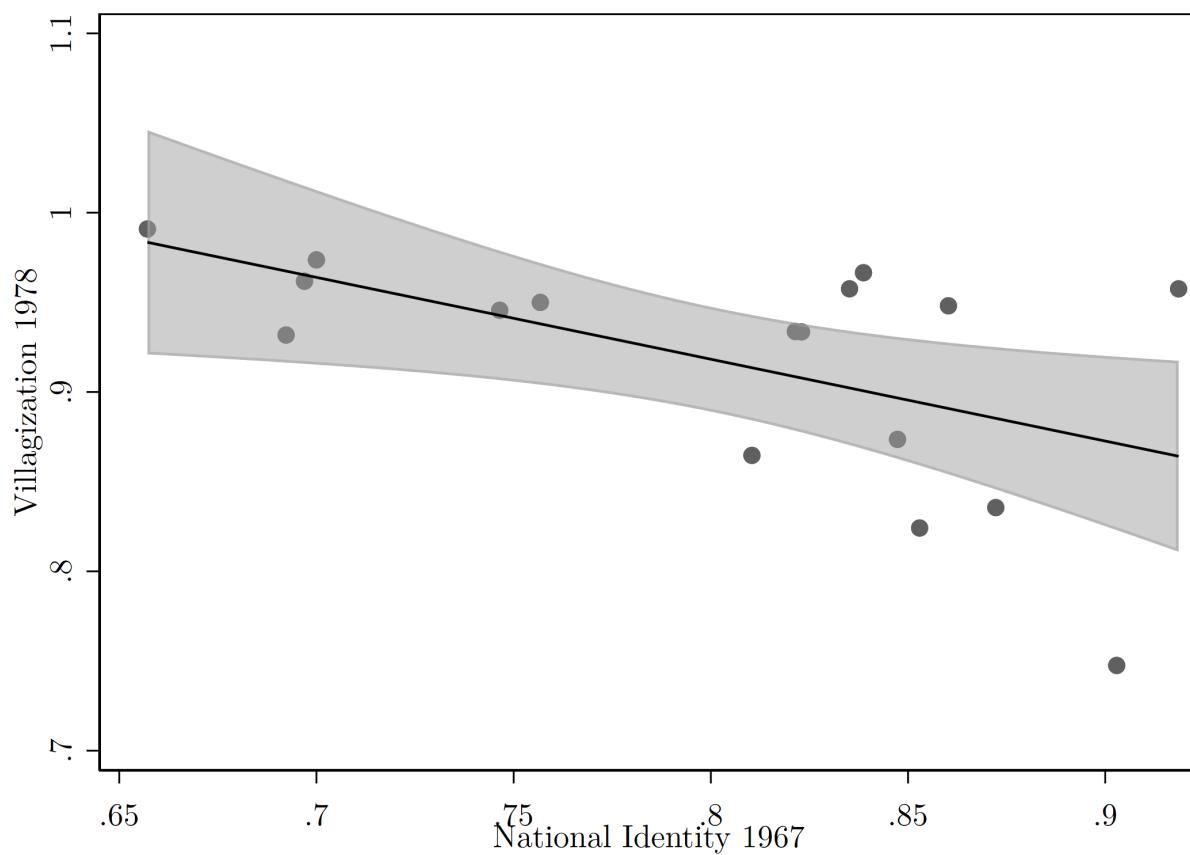
*Note:* This figure displays a histogram of weighted average linguistic distance to Swahili at the district level. Data on the distribution of language groups across districts is from the 1967 Tanzania population census. Data on linguistic distance between groups is from the LEDA project ([Müller-Crepon et al., 2022](#)).

Figure A.7: National Identity and Conflict in Africa



Note: This plot illustrates the correlation between the mean number of battles per 1 million inhabitants (defined as “conflict”) and average expressed feelings of national identity (as opposed to ethnic identity) for 33 African countries between 1997 and 2018. Conflict data from ACLED (Raleigh et al., 2010). Country average of national identity from Afrobarometer Rounds 3-7.

Figure A.8: National Identity in 1967 and Villagization in 1978



Note: This plot shows the correlation (with 95% confidence interval) between region-level villagization in 1978 and average feelings of national identity in 1967. The estimated coefficient is -0.55 and the p-value is 0.04 (standard errors are robust to hetero-skedasticity). Villagization is defined as the fraction of the rural region population living in official registered villages according to the 1978 Tanzania Population Census. The data on national identity is based on a nationwide survey of almost 3,000 Tanzanian secondary school students in 1967 by [Prewitt et al. \(1970\)](#). The original variable is coded as 1 if a respondent feels that the nation is more important than the tribe and 0 otherwise.

## Online Appendix References

- Bureau of Statistics, Ministry of Planning and Economic Affairs (1981). *1978 Population Census*. Dar es Salaam: United Republic of Tanzania.
- Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature* 52(3), 740–98.
- Herbst, J. (2014). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton University Press.
- Hydén, G. (1980). *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry*. University of California Press.
- Jensen (1968). *Government of Tanzania District Statistics*. Dar es Salaam: Government Printers.
- McHenry, D. E. (1979). *Tanzania's Ujamaa Villages. The Implementation of a Rural Development Strategy*. Institute of International Studies, University of California, Berkeley.
- Müller-Crepion, C., Y. Pengl, N.-C. Bormann, et al. (2022). Linking Ethnic Data from Africa (LEDA). *Journal of Peace Research* 59(3), 425–435.
- Prewitt, K., G. Von Der Muhll, and D. Court (1970). School experiences and political socialization: A study of tanzanian secondary school students. *Comparative Political Studies* 3(2), 203–225.
- Raleigh, C., A. Linke, H. Hegre, and J. Karlsen (2010). Introducing ACLED: an armed conflict location and event dataset: Special data feature. *Journal of Peace Research* 47(5), 651–660.
- Shao, I. F. (1982). A neo-colony and its problems during the process of attempting to bring about socialist rural transformation: the case of Tanzania. In *Taamuli: A Political Science Forum*, Volume 12, pp. 29–46.
- United Republic of Tanzania National Department of Education (1971a). *Elimu ya Siasa 6: Taifa na Serikali Wetu*. Tanzania Publishing House.
- United Republic of Tanzania National Department of Education (1971b). *Tujifunze lugha yetu : kitabu cha saba*. Tanzania Publishing House.