

Identity and Social Cohesion in Kenya: Linkages and Correlates

Boaz Munga, Germano Mwabu and Isabella Kiplagat
Social Sector Division
Kenya Institute for Public Policy
Research and Analysis

KIPPRA Discussion Paper No. 160
2014

KIPPRA in Brief

The Kenya Institute for Public Policy Research and Analysis (KIPPRA) is an autonomous institute whose primary mission is to conduct public policy research leading to policy advice. KIPPRA's mission is to produce consistently high-quality analysis of key issues of public policy and to contribute to the achievement of national long-term development objectives by positively influencing the decision-making process. These goals are met through effective dissemination of recommendations resulting from analysis and by training policy analysts in the public sector. KIPPRA therefore produces a body of well-researched and documented information on public policy, and in the process assists in formulating long-term strategic perspectives. KIPPRA serves as a centralized source from which the Government and the private sector may obtain information and advice on public policy issues.

Published 2014

© Kenya Institute for Public Policy Research and Analysis

Bishops Garden Towers, Bishops Road

PO Box 56445-00200 Nairobi, Kenya

tel: +254 20 2719933/4; fax: +254 20 2719951

email: admin@kippra.or.ke

website: <http://www.kippra.org>

ISBN 9966 058 28 7

The Discussion Paper Series disseminates results and reflections from ongoing research activities of the Institute's programmes. The papers are internally refereed and are disseminated to inform and invoke debate on policy issues. Opinions expressed in the papers are entirely those of the authors and do not necessarily reflect the views of the Institute.

KIPPRA acknowledges generous support from the Government of Kenya, African Capacity Building Foundation (ACBF), and the Think Tank Initiative of IDRC.



Abstract

Strategies aimed at enhancing social cohesion and reinforcing a “national identity” need to identify factors that are strongly associated with social cohesion as well as national attachment, and which are amenable to policy. This study uses household level data to examine and discuss the nexus between national and ethnic identities and social cohesion. It sheds light on the widely held assumption that widespread preference for a national identity over an ethnic identity promotes social cohesion. Focusing on the role of education, age, location, gender and ethnic diversity, it examines the determinants of social identity. The most robust factor seems to be education, which is positively correlated with a preference for a national identity. Three components of social cohesion are examined, namely trust, identity and inequality, and how their scores differ across social identity groups. Results indicate that the level of trust as well as pride by ‘individuals who prefer a national identity’ and those ‘who prefer an ethnic identity’ is not statistically different. These two groups, however, seem to differ on their perceptions of inequality. Specifically, the results suggest that perceptions of higher inequality may shift preferences from the ‘national identity’ towards an ‘ethnic identity.’ The study concludes that addressing inequality and perceptions of inequality may form a robust approach to promoting widespread preference for a national identity. Socio-economic factors influencing social cohesion and its components are also discussed. Education remains a robust factor. Ethnic diversity appears not to be bad for cohesion as is commonly found in the literature, while regions with higher poverty rates have lower social cohesion index scores.

Table of Contents

Abstract	iii
1. Introduction	1
1.1 Background and Context.....	1
1.2 Research Problem and Policy Concerns.....	3
1.3 Research Objectives.....	3
1.4 Conceptual Framework	4
2. Literature Review	6
2.1 Ethnic Identity versus National Identity	6
2.2 Social Identities and Social Cohesion	8
2.3 Components of Social Cohesion	8
3. Methodology.....	10
3.1 Measurement Framework	10
3.2 Modelling Approach.....	10
3.3 Data	11
4. Results and Discussion	14
4.1 Descriptive Statistics	14
4.2 Measuring and Explaining Social Cohesion	14
4.3 Identity and Components of Social Cohesion	16
4.4 Socio-Economic Determinants of Self-Categorization.....	18
4.5 Correlates of the Components of Social Cohesion.....	19
5. Summary, Conclusions and Policy Implications	25
5.1 Summary and Conclusions	25
5.2 Policy Implications and Unresolved Questions.....	27
References	29
Appendix	31

1. Introduction

1.1 Background and Context

The reinforcement of a “national identity” has long been a concern of governments across the globe (United Nations, 2007). National identity is usually used to mean preference for identification with the nation when an individual is presented with other competing self-identification entities, which may include an ethnic group. It is generally accepted that building a national identity that supersedes sub-national identities as a stable form of identity results in social cooperation and prosperity (Robinson, 2009). The prevalence of a national identity is presumed to be good for social cohesion. It is thought to be correlated with economic outcomes of interest, such as provision and financing of public goods, and with sustainable development. Social cohesion in this case is characterized by elements such as generalized trust, national pride and social equity. Given Kenya’s diversity, particularly ethnic diversity, the building of a national identity or a Kenyan identity has long been a concern of the government. The government has put in place interventions to promote harmonious coexistence among the various social groups. These interventions include establishment of institutions and enactment of laws for promoting peace. One such institution is the National Cohesion and Integration Commission (NCIC), which was created through the National Cohesion and Integration Act of 2008. Other recent government attempts include the “*Najivunia Kuwa Mkenya*” or the “Proud to be Kenyan” campaign, as well as the Brand Kenya campaign, which aims to “to build a strong country brand that fosters national pride patriotism and earns global recognition and preferences”.

The concept of identity has been studied extensively in other disciplines, including political science, sociology and philosophy (Fearon, 1999; Horowitz, 1985). The concept has been introduced and studied in economic analysis by several authors including Akerlof and Kranton (2000) and Davis (2006). Identity is usually defined as a person’s sense of self, as well as his/her assigned categories. The concept can resolve issues that hitherto could be confounding and allows a new view of possibilities for solving many economic problems (Akerlof and Kranton, 2000).

Personal and national identities

There are many definitions provided across various disciplines on the meaning of the related concepts of personal identity and social identity. To begin with, there are various approaches to understanding self-identification, including sociological and psychological approaches. This study adopts meanings that are most appropriate to its contexts. Personal identity is intricately linked to a social category: individuals usually understand their personal identity in terms of

membership in particular social entities such as ethnic groups (Tajfel, 1972 as cited by Davis, 2006). Individuals have multiple identities at any given time, in addition to which they are known to express preferences over allied or competing identities such as national identity and/versus ethnic identity. Individuals' preferences of one form of identity over another provide an opportunity to examine the reasons for such differences. The preferences can be expressed through self-categorization when individuals are offered hypothetical or realizable opportunities to belong to different social groups.

Social equity

Social equity refers to fairness in the distribution of public goods. In this study, it is measured by the perception of individuals on the fair distribution of these goods. It is one of the pillars of achieving sustainable development.

Generalized trust

This refers to individual trust in other people, and extends beyond one's ethnic group. Trust is a belief that people will meet their obligations or do as expected without evidence. In social interactions, generalized trust avoids the cost of collecting evidence. Thus, trust reduces the transactions costs in economic activities and improves efficiency in operations of organizations and governments. Abundance of trust may therefore drive or stimulate economic success (Alesina and Ferrara, 2000).

Social cohesion

The Government of Kenya (2011) in its Sessional Paper No. 8 of 2012 conceptualizes social cohesion as both a process and an outcome of social interactions, some of whose elements include unity, equity, freedom, democracy, just peace, absence of war or conflict, social justice, and rule of law. KIPPRA (2013) considers that social cohesion can be viewed as a sort of stable equilibrium (implying that communities are able to get back to a non-conflict state within a short duration after a disturbance). In Langer and Stewart (forthcoming), social cohesion is conceptualized as a triangle with three vertices, namely horizontal inequality, generalized trust, and national identity. A decline in horizontal inequality, increase in generalized trust and increase in levels of national attachment by diverse groups are expected to lead to an improvement in social cohesion. In this and other studies, "identification with the nation" or national identity is assumed to be good for social cohesion. If a larger proportion of individuals in a diverse nation identify with the state, then social cohesion is presumed to be enhanced. An implication of defining identity in this manner is that preference for identification with an ethnic group is undesirable and is negatively associated with social cohesion.

1.2 Research Problem and Policy Concerns

The presumption that widespread preference for a national identity over ethnic identity could be beneficial for social cohesion is occasionally taken as a given fact, yet there is no obvious reason that prevents an individual from considering themselves, first and foremost, as belonging to a particular ethnic group and also “trust other ethnic groups” just as much as other individuals would identify themselves preferentially with a national identity such as “Kenyan.” Preferentially categorizing oneself with an ethnic group may not be prejudicial to social cohesion; in fact, if individuals respect diversity in its various forms, it would be expected that individuals who prefer, first and foremost, to be identified by an ethnic group to “trust other individuals and groups” in a similar manner as those who prefer the national label. In a similar vein, there should be no reason for “ethnics” to be less “Kenyan” relative to those who show a preference for a national identity.

In addition to discussing or clarifying these presumptions, it is useful to examine if there are general factors that can be associated with attachment to the nation. In particular, the factors that determine why an individual prefers an ethnic identity over a national identity could deepen the understanding of the intricate relationships between self-categorization and social cohesion. Some of the socio-economic factors that may be important for self-categorization include income, education, place of residence, ethnicity of parents, gender, age, actual and perceptual inequality, discrimination, and access to social amenities. Understanding the broad relationship between these factors and social cohesion is important in the design and implementation of policies for transforming countries into stable and prosperous economies.

1.3 Research Objectives

The overall objective of this study is to examine and discuss the nexus between national and ethnic identities and social cohesion. The study sheds light on the widely held assumption that widespread preference for a national identity over an ethnic identity promotes social cohesion. It begins by measuring social cohesion.

The study seeks to answer the following research questions:

- (i) With respect to social cohesion, how do individuals who categorize themselves as Kenyans differ from those who prefer other identities?
- (ii) What socio-economic factors explain self-categorization into various social identities?
- (iii) What socio-economic factors influence social cohesion and its components?

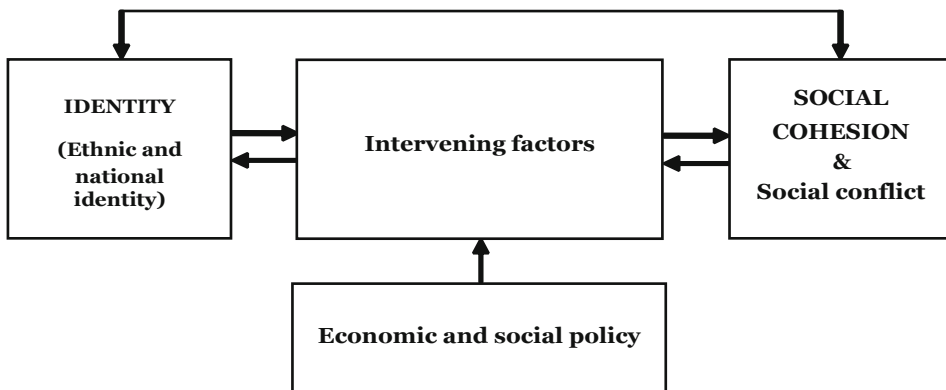
1.4 Conceptual Framework

Figure 1.1 attempts to illustrate the hypothesized relationship between identity and social cohesion. The two phenomena are portrayed as having an indirect relationship through “intervening factors”, but also providing for a direct relationship.

In this context, identity is restricted to whether an individual identifies preferentially with an “ethnic group” or the “nation”. It is a natural outcome of diversity—one form of which is ethnic diversity. The “intervening factors” are numerous and examples include the criteria designed by the state to share or access scarce resources and opportunities, such as land, social infrastructure or employment opportunities. Other factors are discrimination and education (or schooling). A key point to note about the intervening factors is that they can be favourable (enhancing social cohesion) or unfavourable (tending to arouse conflict).

The basic intuition guiding this study is that ethnic identity and national identity could be viewed as a natural state of affairs that has no direct connection with observed society-level conditions such as social cohesion or harmony. The observed relationship between identity (ethnic versus national) and society level conditions (social conflict or cohesion) is indirect. Identity affects these relationships when it interacts with adverse intervening factors (such as poorly designed or implemented policies). It is the adverse intervening factors (rather than identity) that are responsible for social conflict/discord. As an illustration, individuals sometimes identify with the ethnic group preferentially and shun the “national identity” perhaps because they are discriminated against (as a group). This does not make identity the cause of any potential problem, but the discrimination (an unfavourable intervening factor).

Figure 1.1: The relationships between identity and social cohesion



Source: Authors' conceptualization

Many a time, real world societies are characterized by adverse intervening factors that make identity—and in particular ethnic identity—look impure and something to be done away with. However, it appears that the pragmatic route would be to address the intervening factors. Irrespective of whether there is diversity or not (i.e. one or numerous identities), social cohesion will be elusive as long as the intervening factors are not favourable. As long as intervening factors are unfavourable, social conflict may still abound even if individuals ascribe to one identity (ethnic or national).

The intervening factors can be modified or brought into existence by economic and social policy; hence, there is a separation in Figure 1.1 above between economic and social policy, on the one hand, and intervening factors, on the other. The model implies that managing social cohesion would imply modifying or initiating intervening factors in a manner that allows achievement of a socially acceptable outcomes.

2. Literature Review

Economists only began considering the concept of identity in relation to national development in the 1990s. Akerlof and Kranton (2000) opened the possibilities for the individual agent who not only derives utility from consumption and reveals their tastes, but who can also draw boundaries between themselves and “others” as well. One’s identity is believed to shape the way individuals behave, their labourforce participation, and their well-being.

2.1 Ethnic Identity versus National Identity

Ethnic identity (attachment to an ethnic group) and national identity (identification with the state) are two among the multiple identities that individuals may hold. In theory, ethnic identity is, on some occasions, viewed as an alternative to national identity—and the overriding view is that “the increased salience of a national identity must produce a concurrent reduction of the relative salience of ethnic difference” (Robinson, 2009). The empirical literature therefore discusses the factors determining preference for either of the two social identities.

The important question is why individuals preferentially identify with the ethnic group and not the nation—or sometimes with the nation and not with the ethnic group. At the individual level, it is theorized that urbanization, education and formal sector employment are positively related to an individual’s tendency to identify with the nation. Urbanization is associated with greater intra-state migration and mass education, which leads to countrywide homogenization; both processes result in severed ties with communal or ethnic identities (Gellner, 1983; Weber, 1976). Thus, it could be reasoned that individuals with less education as well as rural residents should identify less with the state, all else equal (Hyden, 1983).

In addition, individuals who have membership in poorer ethnic groups will be less likely to identify with the nation. Groups that have benefited less from prosperity (or perceive greater inequality) are inclined to identify with their ethnic groups (Robinson, 2009). Ethnic demographics also play a role in how individuals preferentially identify themselves. Larger ethnic groups may be less likely to identify with the nation because, in doing so, they might have to sacrifice some of their power or resources. At the individual level, when a person shares a certain identity with a greater number of people, his commitment to that identity tends to be higher (Posner, 2004; 2005). Robinson (2009) also notes that when an ethnic group makes up nearly the total population, the association may vary. Specifically, the citizens may see the state and ethnic group as synonymous and in such a situation the groups may espouse the national identification.

The correlates to self-categorization have been investigated in a number of studies. Using data from the World Values Surveys (WVS), Masella (2011) investigated the main determinants of individual self-identification in national rather than ethnic terms in 21 countries. The WVS data had the question: Which of the following describes you? An example is provided for the United States of America where possible answers include: 1. Above all, I am Hispanic American; 2. Above all, I am a black American; ... and, 5. I am an American first and a member of some ethnic group second. They conclude that level of education and income increased an individual's sense of national identity. Individual identity, however, did not significantly vary with occupation and ethnic diversity as measured by the ethnic fractionalization index (ELF).

In a related study, Bossuroy (2011) examined the individual determinants of ethnic identification using sample surveys from main cities of seven West African countries (Benin, Burkina Faso, Cote d'Ivoire, Mali, Niger, Dakar and Togo). Identity was established by asking the question: "Which group (or community) are you the most proud to belong to?" The answer is either, one is proud of his/her ethnic group (1) or the nation (0). Using the logistic model on survey data from the different countries, the findings indicate that ethnic salience is generally high among individuals deprived of social and human capital; that is, the less schooled, the unemployed, informal sector employees and the immigrants. One key conclusion of the study is that ethnic salience may be a rational response of individuals facing economic constraints as the lack of education or the difficulty to get integrated in the job market.

Robinson (2009) sought to establish the correlates of national identity in African states using survey data from Afrobarometer studies using a logistic regression. The study concludes that at the individual level, national attachment is associated with the wealthy, more educated, urban residence and formally employed individuals.

Besides these individual level factors, there are a number of other factors that strengthen the competing or complementary identities. According to Eifert, Miguel and Posner (2010), ethnic identities are strengthened by political competition, as is observed in a study based on Afrobarometer surveys conducted in 10 African countries during the periods preceding and following competitive elections. Using the multinomial logit, they further observe that increase in ethnic salience corresponds with decreasing salience in class/occupational identity. Ethnic salience during competitive elections is seen as a tool for mobilizing political power as well as allocation of resources. Ethnicity as a tool for winning competitive elections partly explains Kenya's 2007/08 post-election violence (Miguel, 2004).

2.2 Social Identities and Social Cohesion

Literature on the relationship between ethnic identification (or diversity) and social cohesion has been approached from two angles. On one hand, ethnic identity is viewed as an undermining factor in achieving social cohesion. This view is based on the observation that identities, and in particular ethnic identities, seem to be at the centre of most of the conflicts in both the developed and African nations, such as in Burundi, Guinea, Liberia, Rwanda, Sierra Leone, and Sudan (Hagg and Kagwanja, 2007).

A number of studies, including Miguel (2004), advocate a salience of national identity over ethnic identity. This is based on a study conducted on Kenya and Tanzania, which indicates that in the latter country—where national identity is salient over ethnic identity—there was higher inter-ethnic cooperation compared to Kenya—where ethnic divisions are more pronounced. National identity is thought of as a strategy (through re-categorization) that can improve cooperation and social cohesion in ethnically diverse societies (Robinson, 2009) as well as improve important economic outcomes such as provision of public goods. Thus, early research work assumed that ethnic and national identity were negatively correlated.

The second strand of literature takes a different direction and advocates that ethnic and national identities are not mutually exclusive. In more recent views, the two phenomena may be positively or negatively correlated or uncorrelated with social cohesion. Those who front these views argue that ethnic identity (or diversity) is not the problem. Rather, other factors including poverty, competition for natural resources, poor governance, and historic episodes are the main reasons for the observed lack of social cohesion (Elbadawi and Sambanis, 2000; Robinson, 2009; Tarimo, 2010). The modern approach to explaining the link between ethnic identity and social cohesion has therefore moved from focusing on re-categorization from ethnic to national identity, to the issue of managing ethnic diversity. The argument behind this view is that ethnic diversity is seen as a resource, since it enables people to exchange knowledge and skills as well as perspectives.

2.3 Components of Social Cohesion

This study adapts the description of social cohesion proposed by Langer and Stewart (forthcoming) and KIPPRA (2013), where social cohesion is conceptualized as a triangle with three vertices: horizontal inequality, generalized trust, and national identity. These components are broadly comparable to the components discussed in similar studies, including United Nations (2007) and Markus (2010).

'Horizontal inequalities' is a term used in reference to inequalities among groups (Stewart, 2009). One particular form of horizontal inequality important for social cohesion is "ethnic inequality", which has been used to refer to inequality in income (or other measure of well-being) along ethnic lines. It is hypothesized as likely to lead to political inequality and increase animosity as well as conflict, and thus curtail development (Alesina, Michalopoulos and Papaioannou, 2012; Cederman, Weidman and Gleditch, 2011). Some of the explanations for inequality across groups include differences in geographic endowments and regional differences in development. Some of the channels thought to explain these group inequalities are political inequality, heightened perceptions of discrimination and undersupply of public goods (Alesina, Michalopoulos and Papaioannou, 2012).

As pointed out earlier, trust is a belief that people will meet their obligations or do as expected without evidence. Generalized trust is thought to be influenced by both individual experiences and community characteristics. Alesina and Ferrara (2000), for instance, identify moral or cultural attitudes; past experience of individuals and groups; family, ethnic or racial ties; length of interactions; and legal institutions as some of the factors that determine trust. Using USA data to test their various hypotheses, they conclude that religious beliefs and ethnic origins do not significantly affect trust. The strongest factors that lessen trust are found to be the following: a recent history of traumatic experiences; belonging to a group that historically felt discriminated against (such as black and women); being economically unsuccessful in terms of income and education; and living in a community with a high degree of income disparity. Descriptions of national identity have been summarized in sub-section 2.1.

3. Methodology

3.1 Measurement Framework

This study relies on two broad approaches to data analysis. First, analysis of variance (ANOVA) is used to compare group means of the binary variables, and to examine the question relating to the ways in which an individual who preferentially categorizes themselves as a Kenyan differs from others who reveal a preference for an ethnic label. In the next measurement step, limited dependent variable models are used to provide reasons for social identity preferences. Ordinary least squares are used to explain the correlates of the social cohesion index.

Literature shows that probit and logit models are commonly used for explaining identity preferences (Robinson, 2009; Bossuroy, 2011). This study follows the tradition of Akerlof and Kranton (2000) who incorporate and treat identity as an argument in the standard utility function, implying that individuals have demands for social identities. The utility function in which social identity is incorporated takes the form:

$$U_j = U_j(a_j, a_{-j}, I_j) \dots \dots \dots (1)$$

Where: U_j is utility of individual j ; a_j are actions of j ; a_{-j} are actions of others, and I_j =identity or self-image of individual j . In a reduced-form specification, social identity or self-image in a social context is defined as:

$$I_j = I_j(a_j, a_{-j}; c_j, \epsilon_j, P) \dots \dots \dots (2)$$

Where identity depends on j 's actions a_j and the actions of others, a_{-j} the social categories assignable to j , c_j , j 's own characteristics ϵ_j , and the degree to which j 's own given characteristics go with the social standards or ideals of j 's categories, as indicated by recognized social prescriptions, P .

The social prescriptions, P , are the social constraints that affect an individual's self-categorization. Similarly, there are costs (c_j) and unobservables (ϵ_j) that influence demand for social identities. It is model (2) that forms the variety of binary regression models that are estimated in this paper. Data limitations restrict the right hand side variables to mainly socio-demographic factors.

3.2 Modelling Approach

Interest lies primarily in the response probability:

$$P(y = 1/x) = P(y = 1/x_1, x_2, \dots, x_k) \dots \dots \dots (3)$$

Where: y is a binary indicator while x denotes the full set of explanatory variables. Following Long (1997), it is supposed that there is an unobserved or latent variable y^* that generates the observed y 's. Those with smaller values of $y^*=1$ are observed as $y=0$, while those with larger values of y^* are observed as $y=1$. The latent variable y^* is assumed to be associated linearly with the observed explanatory variables (x 's) through the reduced form model:

$$y_i^* = x_i\beta + \varepsilon_i \dots\dots\dots(4)$$

The latent variable y_i^* is linked to the observed binary variable by the equation:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > \tau \\ 0 & \text{if } y_i^* \leq \tau \end{cases} \dots\dots\dots(5)$$

Where τ is the threshold or cut-off point.

The probit model assumes normal distribution in error terms while the logit model assumes a logistic distribution. The general model estimated takes the form:

$$y_i = X_i\beta + \varepsilon_i \dots\dots\dots(6)$$

Where y takes the value 1 for 'Kenyans' and 0 for 'Ethnics'; X_i is a vector of values for the i^{th} observation, which include age, sex, education, location and ethnic diversity; β_i is a vector of parameters to be estimated; and ε is the error term. The more specific model being estimated is:

$$y_i = \beta_0 + \beta_1 Education + \beta_2 Age + \beta_3 Agesqd + \beta_4 Female + \beta_5 Location + \beta_6 ELF + \varepsilon_i$$

To examine the relationship(s) between social cohesion and social identity, the components of social cohesion (Inequality, Identity and Trust) are introduced as dependent outcome variables in the regression models. The conventional hypotheses regarding the relationships between the independent variables (such as education, location and diversity - or ELF) and identity are discussed in the section on literature review.

3.3 Data

3.3.1 Sample selection and variables

The data used for this analysis is from the Knowledge Attitudes Practices and Perceptions (KAPP) survey. The purpose of the survey was to capture knowledge, attitudes, practices and perceptions of Kenyans regarding issues on ethnicity and social cohesion. The sampling frame was constructed by the Kenya National

Bureau of Statistics (KNBS). The survey design consisted of a total of 69 randomly selected districts across Kenya's eight (8) provinces. A total of 3,300 households were randomly sampled, and randomly selected household members interviewed in each of the selected households. In total, data was collected from 1,807 rural households and 974 urban households covering 57 districts out of the study target of 69. Overall, the household response rate was about 85 per cent.

In one of the questions, the KAPP Survey asked respondents to self-categorize or describe themselves. Self-identification or categorization is a common approach used to identify the preferred identity of an individual in a given context and time. Some examples of possible categories/groups individuals categorize themselves in include nationality, tribe, ethnicity, occupation and gender.¹ Respondents were offered various options of how individuals describe themselves, which included nationality, language, ethnicity, race, religion and gender as well as economic terms, such as working class, middle class or a farmer. Their responses allow for a broad study of individuals based on their self-categorization.

3.3.2 Measuring Social Identity, Ethno-linguistic Fractionalization and Social Cohesion

As a measure of identity, individuals were asked to provide a self-description. The question was posed as follows: "Some people describe themselves in terms of their nationality, language, ethnic group, race, religion or gender and others describe themselves in economic terms, such as working class, middle class or a farmer. How would you describe yourself?" From their responses they were categorized as "Kenyan" (i.e., those who preferred a national outlook) and "ethnics" (those who preferred an ethnic orientation). This outcome variable was coded as binary.

One of the predictor variables is the ethno-linguistic fractionalization index (ELF), a measure of ethnic diversity along ethnic and linguistic lines. It is constructed using concentration formulas developed in earlier studies by Hirschman (1945, 1964). The ELF was computed separately for each county and is computed as:

$$ELF = 1 - \sum p_i^2$$

Where p_i is the probability that any two individuals randomly picked from the population will belong to different ethnic groups. To compute the ELF, all respondents were grouped into six ethnic clusters. These were Kikuyu, Luhya,

¹Self categorization is used narrowly to entail the preferred identity chosen by an individual when presented with a number of social categories such as nationality, language, race, ethnic group, religion, gender and social class—within a defined context.

Luo, Kalenjin, Kamba and a sixth group encompassing individuals who do not belong to these five larger groups. It should be noted that the five largest groups constituted nearly 70 per cent of the entire sample/population.

In the KAPP survey, the interest is on individuals who either identified themselves first and foremost as Kenyans or those who identified themselves using an ethnic label (besides ethnicity, ethnic labels also included those who identified with race, tribe, clan, or region). Nearly 40 per cent of the respondents initially described themselves first and foremost as Kenyans, while 27 per cent initially described themselves first and foremost using an ethnic label (Table 3.1).

Table 3.1: Preferred Group identities (sample size = 2,443)

Preferred Social Identity	Sample Proportions (Standard Errors in Parentheses)
Kenyan	39.8 (0.014)
Ethnic	27.1 (0.013)
Other (religious, occupational, gender, class, and marriage identities)	33.2 (0.013)

Source: Authors' computations from the KAPP Survey 2010

4. Results and Discussion

4.1 Descriptive Statistics

Descriptive statistics from the analytic sample are presented in Table 4.1 and additional summary statistics in Appendix Table A1.3. The mean years of education are 7.8, which is consistent with other nationally representative data sets (such as the Kenya Integrated Household Budget Survey 2005/06, which has a mean of about 8 years). Women are over-represented in the sample, and only about 40 per cent were males (whereas the ratio of females to males is about 1:1 in the most recent census count of 2009). Those who indicated that “other ethnic groups can be trusted” were 81 per cent suggesting that, based on individuals attestations, generalized trust ‘for other ethnic groups’ is relatively high. About 31 per cent of the respondents said they were affected by the post-election violence. The stated level of pride at being “a Kenyan” is high given that about 94 per cent of the individuals said that they were either “very proud” or “proud” to be Kenyan. The sample shares of the larger ethnic groups in Kenya closely mirror their actual share as revealed by previous population census counts.

The null hypothesis for the normality of most of the variables, including inequality (public goods distribution), education, location and diversity (as measured by the ethno-linguistic fractionalization index) cannot be rejected. However, some variables such as gender, “preferred group identity”, “trust of other ethnic groups”, and “pride to be Kenyan” fail the normality tests. The variables that fail the normality test are analyzed using non-parametric statistics, since the t-tests are not reliable (see Appendix Tables A1.4).

4.2 Measuring and Explaining Social Cohesion

Following the conceptualization of social cohesion proposed by Langer and Stewart (forthcoming), social cohesion is characterized as having three components: horizontal inequality, generalized trust and national identity, all of which are used to generate a social cohesion index (SCI) using the principal components method. In this study, the components are measured on a scale of 0 to 1 (or 0 to 100%) where zero is assigned to the least desirable attribute. The steps used to compute the SCI are outlined in Appendix 1. An SCI score of zero would imply the “highest degree of perceived inequality”, “complete lack of generalized trust” and “no pride at all at being Kenyan”, while a score of one (or 100%) would suggest that there is “a perception of equity,” “complete generalized trust in individuals of other ethnic groups,” and that all individuals are “very proud or proud to be Kenyan”.

Table 4.1: Descriptive statistics from the analytic sample

Variable	Sample size	Proportion (%) or mean	Std. deviation	Minimum	Maximum
Age (in years)	2,481	38.628	15.2386	19	65
Age squared (in years)	2,481	11.52	8.0452	1	25
Discrimination (1 = common)	2,077	47.43	0.4995	0	100
Education (schooling in years)	2,480	7.81	5.0172	0	17
Ethno-linguistic fractionalization (ELF)	2,406	33.78	0.2713	0	78.92
Ethnicity (1 = ethnicity is important)	2,433	80.54	0.3960	0	100
Female (1 = female)	2,432	58.06	0.4936	0	100
Location (1 = rural)	2,471	65.55	0.4753	0	100
Parents' ethnicity (1 = same ethnicity)	2,487	94.15	0.2346	0	100
Post-election violence (PEV) (1= affected)	2,435	30.99	0.4625	0	100
Preferred group identity (PGI)	1,569	59.95	0.4901	0	100
Pride (1 = proud to be Kenyan)	2,480	94.18	0.2341	0	100
Inequality (1 = public goods are fairly distributed)	1,889	38.56	0.4869	0	100
Satisfaction (1 = satisfied with life)	2,408	43.79	3.0168	0	100
Social cohesion index (SCI)	1,842	70.32	0.2341	0	100
Trust of other ethnic group (1 = Trust)	2,434	81.67	0.3870	0	100
Trust of own ethnic group (1 = Trust)	2,425	93.68	0.2434	0	100
Kikuyu (1 = Kikuyu)	2,406	22.53	0.3250	0	100
Luhya (1 = Luhya)	2,406	15.15	0.2651	0	93.94
Luo (1 = Luo)	2,406	9.41	0.2135	0	96.15
Kalenjin (1 = Kalenjin)	2,406	10.78	0.2246	0	100
Kamba (1 = Kamba)	2,406	10.36	0.2418	0	100
Other ethnic (1 = not large ethnic group)	2,406	31.76	0.3552	0	100

Source: Authors' computations from the KAPP Survey, 2010

From the computations, the aggregate or national SCI has a value of about 68 per cent. The respective rates for rural and urban areas are 70 per cent and 63 per cent, respectively. The SCI does not seem to vary much by sex. Males and females had SCIs of about 67 and 68 per cent, respectively (Table 4.2).

The overall index is suggesting that based on the aggregated values of the three attributes, inequality, trust and identity, rural areas are more socially cohesive than urban areas. The computed SCI and its components scores (identity, trust

Table 4.2: Summary for SCI by location and sex

SCI by location	Mean (percentage)
Urban	63.21
Rural	69.75
Total (urban and rural)	67.71
SCI by Sex	Mean (percentage)
Male	67.29
Female	67.80
Total (male and female)	67.60

Source: Authors' computations from the KAPP Survey 2010

and inequality) can serve as benchmarks against which computations using a similar methodology can be compared in future.

Langer and Stewart (forthcoming) who use the Afrobarometer survey and a different methodology report a SCI and a variance adjusted SCI ranging from about 50 per cent to 63 per cent in 2005 through 2011. The values are not strictly comparable since the overall method (and definitions) used are not similar.

4.3 Identity and Components of Social Cohesion

The main objective of ANOVA is to use t- or p-values to test whether sample proportions for different social identities are the same; for example, it can be

Table 4.3: Proportion differences by social identities (two sample tests by preferred group identity)

Variable	Difference in Mean*	Standard Error	z score	p-value
Ethno-linguistic fractionalization index	-0.0207	0.0243	-0.8516	0.3944
Ethnic identity (1 = important)	0.0479	0.0205	2.3332	0.0196
Female (1 = female)	0.0550	0.0257	2.1359	0.0327
Gini index	0.0040	0.0252	0.1602	0.8727
Inequality (1 = public goods distributed fairly or no inequality)	-0.0681	0.0285	-2.3937	0.0167
Location (1=rural)	0.0797	0.0243	3.2807	0.0010
Post election violence (1 = affected)	-0.1318	0.0229	-5.75	0.0000
Proud to be Kenyan (1 = proud)	-0.0119	0.0105	-1.13	0.2571
Social Cohesion Index	-0.0299	0.0271	-1.1024	0.2703
Trust of other ethnic groups (1 = trust)	-0.0041	0.0201	-0.21	0.8370
*Difference = prop(Ethnics) – prop(Kenyan) and *Ho: Difference = 0 and Ha: Difference ≠ 0				

tested whether or not the proportion of persons stating that they are “Kenyan” is different from the proportion preferring “an ethnic-based social identity” with respect to a particular attribute.

The analysis begins with the three identified components of social cohesion—trust, identity and inequality. The level of trust by ‘Kenyans’ and ‘ethnics’ is not statistically different; that is, ‘Kenyans’ and ‘ethnics’ are similar in this respect. The null hypothesis of equal proportions in this case cannot be rejected at the 5 per cent level (Table 4.3).

In a similar vein, “Kenyans” and “ethnics” do not differ in their “pride at being Kenyan”, in the way they are “generally satisfied with life”, and in their “experiences with ethnic-based discrimination”. The results on trust and pride (which failed the normality test) do not change even when the Wilcoxon-Mann-Whitney test, a non-parametric test, is used (see Appendix Table A1.5).

The SCI does not differ statistically between “ethnics” and “Kenyans”. Although Kenyans have a larger mean for the SCI, the difference in the SCI between the two groups is statistically insignificant. But there is evidence to suggest that there is a statistically significant difference between the underlying distributions of the social identities “Kenyans” and “ethnics”) by the SCI ($p=0.0166$) (Appendix Table A1.5). The independent sample t-tests and the rank sum tests thus suggests that preference for an “ethnic identity” need not be detrimental to social cohesion or to its components (in this case identity and trust) (Table 4.3 and Appendix Table A1.5).

Individuals who preferentially self-categorize themselves with an ethnic group (rather than the nation) can be “trusting of other ethnic groups” and “proud of country” as much as individuals who prefer a national identity. However, “ethnics” and “Kenyans” seem to differ on their perceptions of inequality. When asked whether public goods are distributed fairly, a large proportion of “Kenyans” (41%) said the distribution was fair relative to nearly 35 per cent for “ethnics”. The difference between the proportions is statistically significant (at 5%).

The above finding suggests that “ethnics” perceive greater inequality in the distribution of public goods compared to “Kenyans”. A reasonable speculation is that inequality may shift preferences from the ‘national identity’ and towards an ‘ethnic identity’ (since it is less pragmatic to argue that the preference over one or the other identity could be driving the perception of inequality). The proportion of “ethnics” who hold that ethnic identity is “important in defining their identity” is significantly larger than that of “Kenyans”. With respect to sex, a larger proportion of females prefer an ethnic identity to a national one (63% and 54%, respectively) and the differences in the proportions is significant.

4.4 Socio-Economic Determinants of Self-Categorization

The regression results from the probit model are presented in Table 4.4. Logit models were also estimated for comparison. The logit results are not reported, but they are similar to the probit results. The estimated coefficients should be interpreted as magnitudes of the correlations of the associated variables with social identities rather than as causal effects. Focus is on the role of education, age, location, gender and ethnic diversity on preferences for social identities.

4.4.1 Preferred group identity

The results examine socio-economic determinants of self categorization into the two social identities—national identity and ethnic identity. Different model specifications are tried, focusing on a few explanatory variables and their interactions. Table 4.4 summarizes the preferred specifications.

Table 4.4: Correlates of social identities: Dependent variable is preferred group identity (takes the value 1 if ‘Kenyan’ and 0 if ‘Ethnic’)

Variables	Probit Estimates (Marginal Effects)	Probit Estimates (Marginal Effects)
	Number of obs 1,459 Wald chi ² (6) 34.49 Prob > chi ² 0.0000 Pseudo R ² 0.0287 Log pseudolikelihood -954.64695	Number of obs 1,459 Wald chi ² (6) 35.45 Prob > chi ² 0.0000 Pseudo R ² 0.0306 Log pseudolikelihood -952.7393
Constant	–	–
Education (years of schooling)	0.0127*** (0.0039)	0.0275*** (0.0095)
Location+ (1 = rural)	-0.0255 (0.0518)	-0.1562 (0.1108)
Education*Location	-	-0.0194* (0.0102)
Age (in years)	0.0013 (0.0088)	0.0015 (0.0088)
Age squared (in years)	0.0000 (0.0001)	0.0000 (0.0001)
Female+ (1= female)	-0.0948*** (0.0361)	-0.0938*** (0.0362)
Ethno-linguistic fractionalization index	0.0242 (0.0756)	0.0020 (0.0777)

Note: Marginal effects and robust standard errors are reported. +Are discrete changes of dummy variable from 0 to 1. The numbers in parentheses are standard errors and *, ** and *** signifies significant at 10%, 5% and 1%, respectively.

From the probit estimates, education and sex are statistically significant (at the conventional levels) in their association to preferred social identities. The coefficients of the ELF and location variables are not statistically significant at the conventional levels. Education is positively correlated with a preference for a national identity, and an additional year of schooling increases the probability of self identification as a Kenyan by about 1.3 per cent (in regression without interaction term). In the specification that includes an interaction of education and location, rural areas have lower marginal effects of education, on average. Specifically, the average change in the predicted conditional probability of one being Kenyan for a one year increase in education differs between rural and urban areas by about 2 percentage points. Education could be a proxy of a wide variety of factors/effects, including the fact that the less educated individuals are more likely to be less prosperous and inclined to feel more alien to the state. Females are less likely than males to prefer a national identity. The difference is about 9.5 per cent less for females.

The positive sign of the ethno-linguistic fractionalization index (ELF) suggests that on average, ethnic diversity is associated with national rather than ethnic self-identification. In other words, it would be expected that individuals from counties (or regions) that are more ethnically diverse would be more likely to prefer a national identity. Rural residents are also less likely to prefer a national identity (Table 4.4).

4.5 Correlates of the Components of Social Cohesion

An attempt is made to explain the correlates of social cohesion by first examining the correlates of each of its three components—‘identity’, ‘inequality’ and ‘trust.’ The difference between the dependent variable ‘preferred group identity’ and ‘identity’ as one of the components of social cohesion should be made clear at this point. The ‘preferred group identity’ emanates from self categorization of respondents. This is interpreted as an ‘indicator of preference’ over the competing or complementary social identities. This study measures the ‘identity’ component of social cohesion using ‘pride to be Kenyan’, which is interpreted as a measure of ‘degree of attachment to the nation’ (rather than a preference).

‘Pride’ seen in this way, is perhaps a better measure of ‘national identity’ as a component of social cohesion since it allows the capture of ‘attachment to the nation’ without penalizing those who would prefer an ‘ethnic identity’ and still feel attached to the nation. But why would it be a better measure? The argument is that if ‘identification with the nation’ (usually called “national identity”) is chosen as one of the components of social cohesion (and use a scale of 0 to 1 where 1 means

‘best outcome’ or ‘preference for national identity’) there will be an inherent assumption that ‘preference for ethnic identity’ is bad for cohesion. Clearly, this measure will be inconsistent with the earlier observation that preference for an ‘ethnic identity’ need not be detrimental to social cohesion.

4.5.1 Correlates of pride

The correlates of “pride”, which is used as a proxy for the identity component of social cohesion, are summarized in Table 4.5. Identity is measured as a binary variable that takes a value of 1 when an individual reports being ‘very proud’ or ‘proud to be Kenyan’ and zero (0) when they are ‘not proud’ (or ‘not proud at all’).

Education and location are statistically significant in their association with identity (pride). An additional year of education is associated with about a 0.3 per cent increase in “pride to be Kenyan” or national identity. Location seems to matter for pride and rural residents have a nearly 4 per cent larger likelihood of being proud to be Kenyan relative to their urban counterparts (Table 4.5). The interaction of education and location (not reported) does not improve the overall results.

Table 6: Correlates of social cohesion: Dependent variable is “proud to be Kenyan” (“proud to be Kenyan” = 1 if individual is proud or very proud)

Variables	Probit Estimates (Marginal Effects)
	Number of obs 2325 Wald chi ² (6) 13.67 Prob > chi ² 0.0336 Pseudo R ² 0.0168 Log pseudolikelihood -500.3631
Constant	–
Education (years of schooling)	0.0027** (0.0013)
Age (in years)	0.0003 (0.0029)
Age squared (in years)	0.0000 (0.0000)
Female+ (1= female)	0.0057 (0.0130)
Location+ (1 = rural)	0.0388*** (0.0164)
Ethno-linguistic fractionalization index	0.0109 (0.0219)

Note: Marginal effects and robust standard errors are reported for probit. +Are discrete changes of dummy variable from 0 to 1. The numbers in parentheses are standard errors and *, ** and *** signifies significant at 10%, 5% and 1%, respectively.

4.5.2 Correlates of Inequality

Table 4.6 summarizes the correlates of inequality, measured using a dummy variable with value 1 when an individual perceives that public goods are fairly distributed across Kenya's regions, or zero otherwise.

About 43 per cent of the sampled individuals agreed that public goods were fairly distributed. An attempt is made to interact education and location, and location and ELF. None of these interactions seems to improve the results. The interaction between location and ELF makes the ELF variable significant at 10 per cent, but the interaction term itself is not significant at the conventional levels.

In the preferred specification, besides education, the other variables are not statistically significant at the conventional levels. Education and ethno-linguistic fractionalization index are negatively correlated with perceived inequality. This suggests that increases in education (and more ethnic diversity) are associated with "rising perceptions of inequality" in the distribution of public goods (note that inequality is declining as the measure moves from 0 to 1). Relative to males, females tend to have larger 'perceptions of inequality in the distribution of public goods'. The negative sign of the ELF may capture the rural urban divide, since urban areas tend to be more ethnically diverse. The result that rural residents are more likely to perceive less inequality is thus consistent with the negative sign of the ELF.

Table 4.6: Correlates of social cohesion: Dependent variable is Inequality (inequality = 1 if individual perceives no inequality in the distribution of public goods, and 0 for "inequality")

Variables	Probit Estimates (Marginal Effects)
	Number of obs 1776 Wald chi ² (6) 10.13 Prob > chi ² 0.1194 Pseudo R ² 0.0071 Log pseudolikelihood -1180.6316
Constant	–
Education (years of schooling)	-0.0065* (0.0035)
Age (in years)	-0.0051 (0.0073)
Age squared (in years)	0.0000 (0.0001)
Female+ (1= female)	-0.0135 (0.0328)
Location+ (1 = rural)	0.0040 (0.0453)
Ethno-linguistic fractionalization index	-0.1048 (0.0687)

Note: Marginal effects and robust standard errors are reported for probit. +Are discrete changes of dummy variable from 0 to 1. The numbers in parentheses are standard errors and *, ** and *** signifies significant at 10%, 5% and 1%, respectively.

The findings that (i) rural residents are more proud to be Kenyan (Table 4.4) and (ii) rural residents are more likely to perceive less inequality (Table 4.6) may agree with intuition. Although rural areas have inferior developmental scores than urban areas of Kenya (such as poverty scores), rural areas have less disparity across individuals. It is likely then that perception of inequality (because individuals are likely to compare themselves with their immediate neighbours) should be higher in urban areas. It is this ‘higher’ perception of inequality that could be lowering the level of pride or national attachment much more in urban than in rural areas (Tables 4.5 and 4.6).

4.5.3 Correlates of trust

Correlates of the third component of social cohesion, trust, are summarized in Table 4.7. From the probit results, the relatively large value of the p-value from the Likelihood Ratio test, of about 16.5 per cent, suggests that the model as a whole is not significant at the conventional levels. This does not change even with the introduction of interaction terms. Trust does not seem to be convincingly explained by the proposed model. Location and ELF are positively associated with trust of other ethnic groups (Table 4.7).

Table 4.7: Correlates of Trust: Dependent variable is “Trust of other ethnic group” (Trust = 1 if individual says yes)

Variables	Probit Estimates (Marginal Effects)
	Number of obs 2,284 Wald chi ² (6) 2,284 Prob > chi ² 0.1649 Pseudo R ² 0.0085 Log pseudolikelihood -1087.5035
Constant	–
Education (years of schooling)	-0.0010 (0.0023)
Age (in years)	-0.0030 (0.0049)
Age squared (in years)	0.0000 (0.0001)
Female+ (1= female)	-0.0180 (0.0216)
Location+ (1 = rural)	0.0677** (0.0317)
Ethno-linguistic fractionalization index	0.0959** (0.0456)

Note: Marginal effects and robust standard errors are reported for probit. +Are discrete changes of dummy variable from 0 to 1. The numbers in parentheses are standard errors and *, ** and *** signifies significant at 10%, 5% and 1%, respectively.

4.5.4 Correlates of social cohesion index (SCI)

Unlike the components of SCI discussed above (i.e. identity, inequality and generalized trust), the SCI is not a binary variable, but a variable that varies from 0 to 1 (or 0% to 100%) across the counties. Therefore, a binary probit is not estimated as in the preceding estimations.

It is important to note that SCI is computed per county. In the subsequent estimations, where it appears as a dependent variable, the independent variables (education, age, age squared, sex (female) and location) are computed mean of each county. Mean of location can be interpreted as the proportion of individuals residing in rural areas within a given county, while the mean of sex (female) can be interpreted as the proportion of females (in the sample) within a given county. The ELF is computed by county. The data used for the correlates of SCI is attached in Appendix Table A1.3. The variables incorporated in the SCI model are not significant at the conventional levels besides the coefficient for education (Table 4.8).

Counties or regions with higher ‘mean years of education’ have a lower SCI score on average. The positive coefficient for the location variable suggests that regions with lower levels of urbanization are expected to have larger SCI scores. The expectation on education would be that regions with “larger mean for education” would have larger social cohesion indices. However, from the results and data (see Appendix A6) it is known that urban areas tend to have larger “mean values for years of schooling” and less social cohesion (lower indices). The

Table 4.8: Correlates of SCI: Dependent variable is “social cohesion index” (whose components are trust, inequality and pride)

Variables	Ordinary least squares
	Number of obs 43 F(6, 36) 2.58 Prob > F 0.0350 R-squared 0.3007 Adjusted R-squared 0.1842 Root MSE 0.06187
Constant	0.4489 (0.4805)
Age (in years)	0.0177 (0.0242)
Age squared (in years)	-0.0004 (0.0003)
Education (years of schooling)	-0.0130** (0.0051)
Ethno-linguistic fractionalization index	0.0182 (0.0452)
Female+ (1= female)	-0.1041 (0.0838)
Location+ (1 = rural)	0.0028 (0.0403)

Note: The numbers in parentheses are standard errors and *, ** and *** signifies significant at 10%, 5% and 1%, respectively.

negative coefficient for education may therefore be a result of its interaction with the location variable.

It is noteworthy that the ethno-linguistic fractionalization index is positively correlated with the SCI. This result suggests that counties that are more ethnically diverse would be expected to be more cohesive (have larger SCI scores). Diversity appears not to be bad for cohesion as is commonly found in the literature.

The results insinuate that regions with a larger proportion of females (or female respondents) would have a lower SCI score. Practically, counties or regions across Kenya are not expected to have large variations of the proportion of females at a point in time or even across time. To this extent, not much weight would be attached to this result.

An attempt is made to link the SCI to national development, as proxied by major determinants of economic growth such as the density of roads networks across counties. This will enable an analysis of the role of social equity in fostering social cohesion. The ordinary least squares results are summarized in Table 4.9.

Only education is significant at the conventional levels. Road network density and SCI seem to be negatively correlated. The poverty rate (across the counties) seems to enter the model with a negative sign, suggesting that regions with higher poverty rates have lower SCI scores.

Table 4.9: Correlates of SCI: Dependent variable is “social cohesion index” (whose components are trust, inequality and pride)

Variables	Ordinary least squares
	Number of obs 43 F(8, 34) 1.89 Prob > F 0.0939 R-squared 0.3080 Adjusted R-squared 0.1452 Root MSE 0.06525
Constant	1.4836 (1.3789)
Age (in years)	0.0099 (0.0131)
Log of age squared (in years)	0.0099 (0.0131)
Education (years of schooling)	-0.0131* (0.0069)
Ethno-linguistic fractionalization index	0.0265 (0.0490)
Female+ (1= female)	-0.1126 (0.0865)
Location+ (1 = rural)	0.0028 (0.0432)
Log of road network density	-0.0077 (0.0118)
Poverty rate (in 2005)	-0.0002 (0.0008)

Note: The numbers in parentheses are standard errors and *, ** and *** signifies significant at 10%, 5% and 1%, respectively.

5. Summary, Conclusions and Policy Implications

5.1 Summary and Conclusions

Several disciplines and authors believe that identity, usually defined as “a person’s sense of self, as well as his/her assigned categories”, can resolve confounding issues and that it allows a new view of possibilities for solving many economic problems. This study focused on two broad forms of self categorization, namely “ethnic identity” and “national identity.” It examined not only possible socio-demographic correlates of self identification, but also their nexus with social cohesion. Two major areas of policy concern were examined. The first is the presumption that widespread preference for national identity over ethnic identity could be beneficial for social cohesion. Second was an examination of the general factors that can be associated with attachment to the nation or ethnic group.

Using three components of social cohesion—horizontal inequality, generalized trust and national identity—social cohesion index was generated using the principal components method. From the computations, the aggregate or national social cohesion index (SCI) has a value of about 68 per cent. The respective rates for rural and urban areas are 70 per cent and 63 per cent, respectively. The SCI does not seem to vary much by sex. Males and females had SCIs of about 67 and 68 per cent, respectively. The computed SCI and its components scores (identity, trust and inequality) can serve as benchmarks against which computations using a similar methodology can be compared in future.

5.1.1 Identity and components of social cohesion

This study adapted horizontal inequality, generalized trust and national identity as the key components of social cohesion. In some studies, ‘preferred group identity’ is used as one of the components of social cohesion and is measured in such a manner that identifying with an ethnic group is negatively associated with social cohesion. The inherent assumption is that ‘preference for ethnic identity’ is bad for cohesion. However, this study has shown that preference for an ‘ethnic identity’ need not be detrimental to social cohesion. A key conclusion would be that a measure that avoids penalizing those who preferentially identify with an ethnic group (such as pride) could be a better measure of “identifying with the state”.

The study examined differences between ‘Kenyans’ and ‘ethnics’ based on social cohesion scores as well as the three identified components of social cohesion—trust, identity and inequality. Primarily, the levels of trust, ‘pride’ (a proxy for identity),

and the social cohesion index does not differ statistically between ‘Kenyans and ‘ethnics.’ The findings suggest that preference for an “ethnic identity” need not be detrimental to social cohesion or to its components (in this case identity and trust). Nevertheless, ‘Kenyans’ and ‘ethnics’ seem to differ on their perceptions of inequality. Specifically, ‘ethnics’ perceive greater inequality in the distribution of public goods compared to ‘Kenyans’. A reasonable speculation is that inequality may shift preferences from the ‘national identity’ and towards an ‘ethnic identity.’

These findings may suggest that addressing inequality and perceptions of inequality may form a robust approach to promoting widespread preference for a national identity. Socio-economic inequalities may not be conducive for realizing social cohesion.

5.1.2 Socio-economic determinants of self categorization

Focus was on the role of education, age, location, gender and ethnic diversity (measured by the ELF) as determinants of preferred group identity. The most robust factor seems to be education. Education is positively correlated with a preference for a national identity, and an additional year of schooling increases the probability of self-identification as a Kenyan by about 1.3 per cent. In the specification that includes an interaction of education and location, rural areas have lower marginal effects of education, on average.

Education can capture a wide variety of effects including the fact that the less educated are less prosperous and feel much more alien to the state. Sex is also significant and females are less likely than males to preferentially pick national identity. The other variables of age, location and ethnic diversity are not statistically significant. Even so, the positive sign of the ethno-linguistic fractionalization index (ELF) is noteworthy, suggesting that on average, ethnic diversity is associated with national rather than ethnic self identification.

5.1.3 Socio-economic factors influencing social cohesion and its components

The variables incorporated in the social cohesion index (SCI) model are not significant at the conventional levels, except the coefficient for education. Counties or regions with higher ‘mean years of education’ have a lower SCI score on average. The role of education in enhancing or undermining social cohesion seems confounding, since it would be expected that regions with higher average levels of education should be more socially cohesive. It appears education is the channel through which individuals are able to discern inequalities, such as through

increased mobility. Perhaps improvements in education must be supported by an equally strong drive to address inequality (or at least perceptions of inequality) in order for the country to remain socially cohesive.

Ethnic diversity appears not to be bad for cohesion as is commonly found in the literature. The ELF index is positively correlated with the SCI. An attempt is made to link the SCI to national development, as proxied by major determinants of economic growth such as the density of roads networks across counties. Road network density and SCI seem to be negatively correlated. Regions with higher poverty rates have lower SCI scores.

Correlates of the components of social cohesion were examined. The ‘identity’ component of social cohesion is measured using ‘pride to be Kenyan’. This is interpreted as a measure of ‘degree of attachment to the nation’ rather than a preference. Education and location are statistically significant in their association with identity (pride). An additional year of education is associated with about a 0.3 per cent increase in “pride to be Kenyan” or national identity. Rural residents have a nearly 4 per cent larger likelihood of being proud to be Kenyan relative to their urban counterparts.

In the preferred specification for inequality, only education was statistically significant at the conventional levels. Education and ELF index are negatively correlated with perceived inequality. As for trust, the relatively large value of the p-value from the Likelihood Ratio test, of about 16.5 per cent suggests that the model as a whole is not significant at the conventional levels. Trust does not seem to be convincingly explained by the variables focused on.

The variables incorporated in the SCI model are not significant at the conventional levels, besides the coefficient for education. Counties or regions with higher ‘mean years of education’ have a lower SCI score on average. The positive coefficient for the location variable suggests that regions with lower levels of urbanization are expected to have larger SCI scores.

5.2 Policy Implications and Unresolved Questions

- Addressing inequality, or perceptions of inequality, is one of the pathways to promoting a socially cohesive society. It could be inferred that ‘patriotism campaigns’ and other similar interventions by the government may not have the desired impacts (on social cohesion) if inequality is unchanged.
- Due to data limitations, this study has not been able to provide complete explanations of the components of social cohesion (especially trust). This can be an area for further research. There is need to focus on how specific

government policies impact on ethnic and national consciousness as well as social cohesion in sectors such as education, land and settlement, health, the environment and industry. This kind of research can be supported by good quality data on the attributes of social cohesion such as trust, pride and peace across time and space.

- In many studies, identity has been shown to determine behaviour and perceptions (Salant and Rubinstein, 2008). Individuals assigned a particular identity flaunt the characteristic(s) associated with that identity. This is not examined in this study, but has important implications on policy. In particular, deliberate interventions that can be made to set expected virtues from the various identities individuals are free to choose.
- To analyze issues discussed in this study more effectively, existing tools for capturing data on social cohesion, such as those used by KIPPRA and NCIC, can be improved. As an example, in capturing data on identity, responses can be based on preferences as is done in some versions of the World Values Surveys (where respondents choose from categories such as “above all, I am Kenyan” or “above all, I am a member of tribe x”).

References

- Akerlof, G. and R. Kranton (2000), "Economics and Identity", *Quarterly Journal of Economics*, 115(3), pp. 715-753.
- Alesina A. and E. Ferrara (2000), The Determinants of Trust. NBER Working Paper Series No. 7621 Cambridge, MA: National Bureau of Economic Research.
- Alesina A. F., S. Michalopoulos and E. Papaioannou (2012), Ethnic Inequality. NBER Working Paper Series No. 18512, Cambridge, MA: National Bureau of Economic Research.
- Bossuroy, T. (2011), Ethnic Solidarity and the Individual Determinants of Ethnic Identification. A SALDRU Working Paper Number 69. Cape Town: Southern Africa Labour and Development Research Unit.
- Bratton , M. and M.S. Kimenyi (2008), Voting in Kenya: Putting Ethnicity on Perspective. Department of Economics Working Paper 2008-09, University of Connecticut.
- Cerdaman, L. E., N. B. Weidman and K.S. Gleditch (2011), "Horizontal Inequalities and Ethnoactionalist Civil War: A Global Comparison", *American Political Science Review*, 105(1), pp. 478-495.
- Council of Europe (2005), Concerted Development of Social Cohesion Indicators: Methodological Guide. Strasbourg: Council of Europe.
- Eifert, B., E. Miguel and D.N. Posner (2010), "Political Competition and Ethnic Identification in Africa." *American Journal of Political Science*, 54, pp. 494-510, available at <http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2010.00443.x/pdf>. Accessed on 22nd July 2013.
- Elbadawi, I. and N. Sambanis (2000), "Why are there so many Civil Wars in Africa? Understanding and Preventing Violent Conflict." *Journal of African Economies*, Vol. 9(3), pp.244-269.
- Fearon, J.D. (1999), "What is Identity (as we now use the word)? Mimeo, Stanford University.
- Gellner, E. (1983), *Nations and Nationalism*. Ithaca, NY: Cornell University Press.
- Government of Kenya (2011), Sessional Paper No. 8 of 2012 on National Cohesion and Integration. Nairobi: Government Printer.
- Hagg, G. and P. Kagwanja (2007), "Identity and Peace: Reconfiguring Conflict Resolution in Africa", *African Journal on Conflict Resolution*, Vol. (7) No. 2, pp .
- Hirschman, A.O. (1945), *National Power and the Structure of Foreign Trade*. Berkeley: University of California Press.

- Hirschman, A.O. (1964), "The Paternity of an Index", *American Economic Review*, (September), pp. 761-762.
- Horowitz, D.L. (1985), *Ethnic Groups in Conflict*. Berkeley, CA: University of California Press.
- Hyden, G. (1983), "Problems and Prospects of State Coherence." In D. Rothchild and V. A. Olorunsola (eds). *State versus Ethnic Claims: African Policy Dilemma*, pp. 67-84. London: Westview Press.
- KIPPRA (2013), *Inequalities and Social Cohesion in Kenya: Evidence and Policy Implications*. Nairobi: Japan International Cooperation Agency.
- Langer, A. and F. Stewart (forthcoming), "Measuring Social Cohesion: The Case of Kenya." Unpublished manuscript .
- Markus, A. (2010), "Mapping Social Cohesion: The Scanlon Foundation Surveys", available at www.globalmovements.monash.edu.au. Accessed on 15th July 2013.
- Masella, P. (2011), "National Identity and Ethnic Diversity". *Journal of Population Economics*. DOI 10.1007/s00148-011-0398-0.
- McLeod, S. A. (2008), "Social Identity Theory", available at <http://www.simplypsychology.org/social-identity-theory.html> .
- Miguel, E. (2004), "Tribe or Nation? Nation Building and Public Goods in Kenya versus Tanzania", *World Politics*, 56(3), pp. 327-362.
- Posner, D. (2004), "The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi", *American Political Science Review*, 98, 4, pp. 529-545.
- Posner, D. (2005), *Institutions and Ethnic Politics in Africa*. New York: Cambridge University Press.
- Robinson, A.L. (2009) "National versus Ethnic Identity in Africa: State, Group and Individual Level Correlates of National Identification. Working Paper No. 112. Afrobarometer .
- Tajfel, H. (1972), "Social Categorization". In S. Moscovici (ed.), *Introduction a la Psychologie Sociale*, Vol. pp. 272-302. Paris: Larousse.
- United Nations (2007), *A System of Indicators for Monitoring Social Cohesion in Latin America*. Santiago: United Nations.
- Weber, E. (1976), *Peasants in Frenchmen: The Modernization of Rural France, 1880-1914*. Stanford: Stanford University Press.

Appendix

Appendix 1: Computing the Social Cohesion Index for Kenya using the KAPP 2010 Data

Following earlier work (e.g. Stewart and Arnim, forthcoming) the adapted components of social cohesion are: trust, inequality and identity. The computation of the SCI uses the 2010 KAPP Survey data.

Inequality is measured by a response to the question “please specify ... how strongly you agree or disagree with” the statement: “Public goods are distributed fairly across Kenya’s regions.” Public goods were defined to include public schools, public hospitals, law enforcement, and roads. Respondents who “strongly agreed” or “agreed” were recoded as 1 (implying no inequality) while those who “disagreed” or “strongly disagreed” were recoded as zero (implying existence of inequality). The few “neutral” cases were dropped.

Identity (a concept that means attachment to the nation rather than any other social identity) is measured using a question on pride that asked individuals “How proud are you to be Kenyan.” The responses were dichotomized into a binary outcome where 1 was “very proud” or “proud” and zero (0) was “not proud” or “not proud at all.” Trust is measured using the question: “How much do you trust ... people of another ethnic group?” The responses were dichotomized into those who trust completely or trust somewhat (recoded as 1) and those who do not trust (recoded as 0).

A social cohesion index is developed using these measures of trust, inequality, and identity (pride). Principal Component Analysis is used. The following were the main steps followed:

1. The three components were checked for inclusion or exclusion in the overall Index. The recommended criteria is to check if the eigenvalues are near 1 (preferably 1 and above). For the three components, all were retained since their eigenvalues were close to 1 as indicated in Table A1.1 below.

Table A1.1: Principal components/correlation

Number of observations = 2,272

Number of components = 3

Trace = 3

Rho = 1.0000

Component	Eigenvalue	Difference	Proportion	Cumulative
Inequality	1.16633	0.184733	0.3888	0.3888
Identity (pride)	0.981598	0.129527	0.3272	0.7160
Trust	0.852071		0.2840	1.0000

2. Weights were computed for the three components. The weights for each of the variables were the eigenvectors of the first principal component (comp1) (Table A1.2). The first principal component yields values that assign larger weight to components that vary the most across counties (a component that does not vary across counties or the unit of analysis would have a weight of zero).

Appendix Table A1.2: Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Unexplained
Inequality	0.6860	-0.0142	-0.7274	0
Identity (pride)	0.5381	-0.6630	0.5204	0
Trust	0.4897	0.7484	0.4472	0

3. The generated weights were used to generate the SCI scores. The SCI scores were computed by county. The SCI for county I is a linear combination,

$$\text{Where: } SCI = p_1(x_1) + p_2(x_2) + p_3(x_3) / \sum_{k=1}^3 p_i$$

SCI is the social cohesion index

P_k represents the weight of each components (or variable) x_k for the first principal component

x_k represents the k^{th} component of SCI (the 3 components are inequality, identity and trust, respectively).

The SCI values and other mean values across the counties are summarized in Table A1.3.

Appendix Tables A1.3: Mean values of education, sex, ethno-linguistic fractionalization index, trust, location, pride, identity, SCI inequality, age, road density, and poverty rate across counties in Kenya

County	Mean Years of Education	Female	Ethno-linguistic Fractionalization Index	Trust of other Ethnic Group	Location (1 = rural)	Pride to be Kenyan	Preferred Social Identity	Social Cohesion Index	Perception of Inequality	Age in Years	Age Squared	Road Network Density per 100sqkm	Poverty Rate 2005
Nairobi	11.59	0.55	0.79	0.73	0.00	0.91	0.75	0.630	0.29	32.16	1,164.74	430.87	22.54
Kiambu	9.32	0.70	0.21	0.74	0.70	0.97	0.66	0.713	0.43	36.78	1,559.40	234.55	27.18
Kirinyaga	8.30	0.56	0.06	0.69	0.88	0.97	0.37	0.737	0.49	45.73	2,340.85	155.58	25.22
Murang'a	7.55	0.57	0.04	0.56	0.84	0.96	0.48	0.717	0.55	46.10	2,421.01	128.94	29.93
Nyandarua	10.00	0.80	0.00	1.00	0.00	0.40	1.00	0.519	0.25	28.50	836.25	112.66	46.32
Nyeri	8.88	0.61	0.03	0.70	0.56	0.92	0.53	0.743	0.57	39.47	1,782.46	120.36	32.75
Kilifi	7.02	0.39	0.32	0.98	0.86	0.92	0.32	0.805	0.59	34.41	1,343.38	30.26	71.38
Kwale	5.95	0.37	0.27	0.84	0.98	0.97	0.61	0.760	0.52	37.41	1,548.11	42.92	74.87
Lamu	7.14	0.64	0.46	0.85	0.84	0.97	0.60	0.665	0.22	37.46	1,558.17	13.36	32.71
Mombasa	11.25	0.28	0.61	0.89	0.05	0.96	0.62	0.694	0.32	35.07	1,322.11	435.27	37.63
Tana rive	7.38	0.38	0.20	0.75	1.00	0.88	0.50	0.722	0.60	36.64	1,562.46	7.31	76.94
Embu	5.99	0.71	0.39	0.83	0.72	0.80	0.52	0.706	0.48	39.16	1,736.02	114.15	42.02
Isiolo	10.04	0.47	0.35	0.75	1.00	1.00	0.58	0.575	0.15	37.86	1,621.49	11.18	72.62
Kitui	6.63	0.64	0.26	0.74	0.84	0.86	0.48	0.674	0.37	40.23	1,872.76	28.33	63.51
Machakos	6.61	0.65	0.00	0.78	0.97	0.81	0.54	0.658	0.39	45.43	2,312.89	96.00	59.58
Makueni	6.82	0.72	0.00	0.72	0.76	0.67	0.36	0.717	0.50	46.23	2,398.69	95.97	64.10
Marsabit	1.60	1.00	0.00	0.70	0.00	0.90	0.75	0.830	0.80	45.70	2,365.15	6.67	83.22
Meru	7.06	0.61	0.06	0.83	0.76	0.93	0.50	0.612	0.19	44.05	2,284.87	85.44	28.27
Tharaka-nithi	8.85	0.56	0.10	0.89	1.00	0.98	0.79	0.814	0.60	49.94	2,730.61	25.72	48.68
Garissa	4.44	0.47	0.24	0.67	0.67	0.95	0.65	0.687	0.42	40.89	1,940.40	6.77	49.20
Mandera	2.83	0.48	0.00	0.61	0.78	0.99	0.76	0.760	0.62	37.62	1,728.58	12.64	87.84

Wajir	1.84	0.49	0.00	0.81	0.95	0.99	0.74	0.833	0.72	41.58	1,966.43	8.77	84.02
Homa bay	6.50	0.60	0.07	0.92	0.74	0.94	0.77	0.682	0.30	38.44	1,730.40	100.29	44.07
Kisii	6.88	0.53	0.03	0.90	0.97	0.97	0.80	0.711	0.32	34.84	1,342.82	88.69	60.75
Kisumu	6.35	0.59	0.56	0.82	0.46	0.96	0.87	0.680	0.31	41.49	2,024.71	130.92	47.83
Migori	6.93	0.56	0.00	0.59	0.59	0.92	0.79	0.660	0.33	38.02	1,655.72	111.21	46.69
Nyamira	8.92	0.46	0.00	0.80	0.76	0.98	0.83	0.684	0.32	33.98	1,343.57	122.43	48.11
Siaya	6.35	0.49	0.11	0.87	0.84	0.96	0.67	0.672	0.26	36.20	1,528.44	85.00	35.28
Baringo	8.75	0.44		0.85	0.58	1.00	0.68	0.715	0.29	40.28	1,917.73		
Elgeyo marakwet	5.38	0.48	0.12	1.00	1.00	1.00	0.34	0.847	0.60	36.45	1,490.10	62.53	55.51
Kajiado	10.66	0.56	0.56	0.94	0.54	0.97	0.60	0.770	0.50	31.84	1,084.69	26.99	11.57
Kericho	8.39	0.49	0.07	0.86	1.00	1.00	0.35	0.650	0.16	36.28	1,483.53	60.20	44.22
Laikipia	8.23	0.54	0.20	1.00	0.89	1.00	0.56	0.806	0.50	41.78	2,042.05	32.71	50.53
Nakuru	7.46	0.70	0.66	0.81	0.76	0.97	0.64	0.775	0.52	34.77	1,380.89	127.30	40.05
Nandi	12.00	0.50	0.00	1.00	1.00	1.00	0.50	0.612	0.00	29.50	895.25	86.11	47.41
Narok	6.29	0.30	0.43	0.91	1.00	0.98	0.33	0.680	0.26	38.97	1,720.61	30.32	34.58
Samburu	7.51	0.60	0.57	0.86	1.00	1.00	0.47	0.725	0.41	40.63	1,952.64	11.73	73.01
Trans nzoia	6.72	0.71	0.37	0.90	1.00	0.94	0.72	0.641	0.22	39.22	1,693.47	73.03	50.18
Uasin gishu	8.83	0.82	0.42	0.89	0.32	0.91	0.56	0.599	0.20	34.43	1,347.89	123.72	51.30
West pokot	6.26	0.60	0.64	0.92	1.00	0.97	0.34	0.717	0.35	39.06	1,803.62	21.92	69.81
Bungoma	7.34	0.53	0.35	0.90	0.73	1.00	0.62	0.710	0.30	44.19	2,202.76	74.72	52.91
Busia	4.76	0.64	0.56	0.83	0.87	0.97	0.45	0.698	0.38	43.31	2,171.91	140.97	66.68
Kakamega	3.74	0.58	0.12	0.91	0.75	0.99	0.55	0.734	0.38	47.27	2,502.87	123.93	52.95
Vihiga	5.20	0.64	0.12	0.85	0.91	0.96	0.51	0.748	0.47	45.98	2,367.76	140.35	41.78
National	7.81	0.58	0.34	0.82	0.66	0.94	0.60	0.70	0.39	38.62	1,723.30	151.20	44.02

Appendix Table A1.4: Descriptive statistics and normality tests

Variable	Sample Frequencies	Proportion (or mean)	Std. Deviation	Min	Max	Normality*
Age (in years)	2,481	38.6158	15.2386	19	65	0.0000
Age squared (in years)	2,481	11.5165	8.0452	1	25	0.0000
Discrimination (1 = common)	2,077	0.4743	0.4995	0	1	1
Education (schooling in years)	2,480	7.8093	5.0172	0	17	0.0000
Ethno-linguistic fractionalization (ELF)	2,406	0.3378	0.2713	0	0.7892	0.0000
Ethnicity (1 = ethnicity is important)	2,433	0.8054	0.3960	0	1	0.0000
Female (1 = female)	2,432	0.5806	0.4936	0	1	1
Location (1 = rural)	2,471	0.6555	0.4753	0	1	0.0000
Parents' ethnicity (1 = same ethnicity)	2,487	0.9415	0.2346	0	1	1
Post-election violence (PEV) (1= affected)	2,435	0.3099	0.4625	0	1	0.0000
Preferred group identity (PGI)	1,569	0.5995	0.4901	0	1	1
Pride (1 = proud to be Kenyan)	2,480	0.9418	0.2341	0	1	1
Inequality (1 = agree public goods are fairly distributed)	1,889	0.3856	0.4869	0	1	0.0000
Satisfaction (1 = satisfied with life)	2,408	0.4379	3.0168	0	1	0.0000
Social cohesion index (SCI)	1,842	0.7032	0.2341	0	1	0.0000
Trust of other ethnic group (1 = Trust)	2,434	0.8167	0.3870	0	1	1
Trust of own ethnic group (1 = Trust)	2,425	0.9368	0.2434	0	1	0.0000
Kikuyu (1 = Kikuyu)	2,406	0.2253	0.3250	0	1	0.0000
Luhya (1 = Luhya)	2,406	0.1515	0.2651	0	0.9394	0.0000
Luo (1 = Luo)	2,406	0.0941	0.2135	0	0.9615	0.0000
Kalenjin (1 = Kalenjin)	2,406	0.1078	0.2246	0	1	0.0000
Kamba (1 = Kamba)	2,406	0.1036	0.2418	0	1	0.0000
Otherethnic (1 = not large ethnic group)	2,406	0.3176	0.3552	0	1	0.0000

*These are p-values of the Shapiro Francia W Test for normal data

Wilcoxon-Mann-Whitney test

Ranksom tests the hypothesis that two independent samples (i.e., unmatched data) are from populations with the same distribution using the Wilcoxon rank-sum test, which is also known as the Mann-Whitney two-sample statistic. The test is a non-parametric analog to the independent samples t-test and can be used when it is not assumed that the dependent variable is a normally distributed interval variable (you only assume that the variable is at least ordinal).

Appendix Table A1.5: Ranksum test or Wilcoxon-Mann-Whitney test by social identities (two sample test by preferred group identity)

Variable	z score	p-value	Probability – pgio>pgi1
Ethno-linguistic fractionalization index	-0.392	0.6949	0.494
Female (1 = female)	2.135	0.0327	0.527
Inequality (1 = public goods distributed fairly or no inequality)	-2.393	0.0167	0.466
Location (1=rural)	3.280	0.0010	0.540
Post-election violence (1 = affected)	-5.747	0.0000	0.434
Proud to be Kenyan (1 = proud)	-1.133	0.2573	0.494
Social Cohesion Index	-2.395	0.0166	
Trust of other ethnic groups (1 = trust)	-0.206	0.8371	0.498
Education	-7.029	0.0000	0.401

*Difference = prop(Ethnics) – prop(Kenyan) and *Ho: Difference = 0 and Ha: Difference \neq 0

Based on the ranksum test, the results suggest that there is no statistically significant difference between the underlying distributions of the preferred group identities by pride (p=0.2573) and trust (p=0.8371). There is a statistically significant difference between the underlying distributions of the preferred group identities by inequality (p=0.0167). These results are identical to earlier conclusions based on ANOVA. The difference between the underlying distributions of preferred group identities by sex (p=0.0327), location (p=0.0010), education (p=0.0000), and by SCI (p=0.0166) is statistically significant.

Chi-square (X²) test

Chi-square (X²) tests were conducted to examine potential median differences

among independent variables by social identity. The null hypothesis is that the K samples were drawn from populations with the same median. There was a significant difference in median values of 'Kenyan' and 'ethnics' on inequality ($p=0.020$) and the social cohesion index ($p=0.022$). The results indicate that there is no statistically significant difference in medians between the two social identities (Kenyan and ethnic) in relation to: ethno-linguistic fractionalization index.

Appendix Table A1.6: Chi-square (χ^2) test of equality of medians (by preferred group identity)

Variable	Mean	Median	Pearson chi2(1)	p-value
Ethno-linguistic fractionalization index	0.34	0.27	-0.333	0.855
Inequality (1 = public goods distributed fairly or no inequality)	0.39	0	-5.443	0.020
Post-election violence (1 = affected)	0.31	0	-32.373	0.000
Age (in years)	38.62	34.5	-11.935	0.001
Social Cohesion Index	0.70	0.61	-5.223	0.022
Education	7.81	8	47.677	0.000
Road density network	151.20	122.4315	2.770	0.096

Appendix Table A1.8: T-tests and one-way ANOVA: Two sample test of proportions by preferred group identity

Variable	Difference in Mean+	Standard Error	z score	p-value
EFL	-0.0207	0.0243	-0.8516	0.3944
Ethnic identity (1 = important)	0.0479	0.0205	2.3332	0.0196
Female	0.0550	0.0257	2.1359	0.0327
Inequality (1 = public goods distributed fairly or no inequality)	-0.0681	0.0285	-2.3937	0.0167
Location (1=rural)	0.0797	0.0243	3.2807	0.0010
Post-election violence (PEV)	-0.1318	0.0229	-5.75	0.0000
Proportion of Kalenjin	0.0303	0.0132	2.2919	0.0219
Proportion of Kamba	0.0191	0.0136	1.4096	0.1587
Proportion of Kikuyu	0.0353	0.0216	1.6344	0.1022
Proportion of Luhya	0.0001	0.0176	0.0044	0.9965
Proportion of Luo	-0.0499	0.0171	-2.9256	0.0034
Proportion of other ethnic group	-0.0349	0.0256	-1.3652	0.1722
Proud to be Kenyan (1 = proud)	-0.0119	0.0105	-1.13	0.2571

Satisfaction	0.0092	0.0253	0.3641	0.7158
SCI	-0.0299	0.0271	-1.1024	0.2703
Trust of other ethnic groups (1 = trust)	-0.0041	0.0201	-0.2100	0.8370
Trust of own (1 = trust)	0.0063	0.0117	0.5335	0.5937

+Difference = prop(Ethnics)–prop(Kenyan) and +Ho: Difference = 0 and Ha: Difference ≠ 0