Peer Effects and Ethnicity in Uganda: Impacts of Coethnic and High-Ability Peers on University Performance[*]

Click here for the latest version.

Isaac Ahimbisibwe[†]

January 23, 2024

Abstract

Empirical research has documented the negative impact ethnic diversity has on several political and economic outcomes in Sub-Saharan Africa, including economic growth, political engagement, conflict, and contributions to public goods. However, we know relatively little about educational peer effects in such settings, which are generally characterized by high ethnic diversity and cross-ethnic mixing. This paper studies the effect of coethnic and high-ability peers in student groups on academic outcomes at a large public university in Uganda, a country with pronounced ethnic heterogeneity and segregation. I link data on student-level university admissions with subsequent grades. Upon admission, dorm assignments are random conditional on gender, providing exogenous variation in peer group formation. On average, high-ability peers (irrespective of ethnicity) and coethnic peers (irrespective of ability) positively affect a student's performance. Whereas the coethnic peer effect disappears by the year of graduation, the high-ability peer effect persists and even increases in magnitude over time. The effect of high-ability coethnic peers on performance is statistically indistinguishable from that of high-ability noncoethnic peers. The results of the heterogeneous effects analysis suggest that the entire coethnic peer effect is driven by students with little exposure to other ethnicities prior to enrolling at the university. The pattern of results is consistent with both psychological and peer-to-peer learning explanations that reflect the specific context of this study.

[*]Thanks to the seminar audiences at the World Bank DIME-KDI School Development Impact Conference Nov 2023, UC Davis Development Tea May 2023, the CEGA's Working Group on African Political Economy initiative meeting in Kigali, Rwanda Jun 2022, as well as Bulat Gafarov, Paco Martartorell, and Aleksandr Michuda. I deeply appreciate the continuous guidance provided by Travis Lybbert, Scott Carrell, and Stephen Boucher. I am thankful for the permission granted by the former academic registrar, Mr. Alfred Masikye, to utilize de-identified secondary records from Makerere University Kampala (MUK) for this research. I received assistance from MUK's deputy academic registrar admissions, Mr. Charles Ssentongo, and his team, including Mr. Justus Karegyeya and MUK's data team. It is important to note that the views expressed in this research are my own and do not represent the stance of MUK as an institution. I acknowledge the support from the UC Davis Henry A. Jastro Graduate Research Award, and the ARE Summer Research funding from the Erika C. H. Meng Award and the Jean Kinsey Graduate Fellowship. This study was reviewed by the UC Davis Institution Review Board.

[†]University of California, Davis

1 Introduction

Understanding the determinants of academic and other outcomes for students in higher education continues to be a priority for university administrators and policy-makers. While significant progress has been made in understanding the role of peer effects on academic performance (Sacerdote, 2011; Foster, 2006; Zimmerman, 2003) and other outcomes, such as major choice De Giorgi et al. (2010), and cheating (Carrell et al., 2008). Most of this research has been conducted in the West. Whether or not these results translate to developing countries, such as those of Sub-Saharan Africa (SSA), is unclear. Indeed, since peer effects reflect social dynamics that can change dramatically across cultural contexts, it seems likely that these effects could operate quite differently in non-Western settings. One specific reason to doubt the external validity of the existing peer effects literature on SSA is the degree and nature of ethnic diversity that characterizes much of the region. Such heterogeneity combined with ethno-linguistic differences may, for example, complicate student collaboration, thereby muting the positive effects of high-ability peers on student performance.

Uganda, the setting for this study, consists of over 50 ethnicities (Uganda Bureau of Statistics, 2016). These ethnicities are geographically segregated, although there is considerable ethnic mixing in the capital of Kampala. Several studies link high ethnic heterogeneity in SSA to several poor economic outcomes, such as public goods provision, economic growth, and firm productivity, and to negative effects on social indicators, especially social trust.[1] A prevalent bias in favor of coethnic interaction partly explains these documented costs associated with high ethnic diversity in SSA. Coupling this diversity with strong ethnic segregation, as is the case in Uganda, further exacerbates mistrust (Alesina and Zhuravskaya, 2011). This added friction to social interaction, cooperation, and collaboration is costly in general but may be especially apparent in student performance at universities that draw from disparate ethnic regions and, hence, where many students first interact intensively with ethnicities other than their own.

This paper leverages the higher education context in ethnically diverse and segregated Uganda to explore the effects of coethnic and high-ability peers on academic outcomes. This unique empirical setting raises a number of questions that this paper studies. Does the share of coethnic peers within a student's peer group affect academic performance more or less than the share of high-ability peers? Do high-ability coethnic peers matter more than high-ability noncoethnic peers? Does the context of Ugandan higher education translate into coethnic peer effects stronger for some students than others? The contribution of this paper to the peer effects literature hinges on providing credible answers to these questions in this distinctive setting.

In the empirical stage for this analysis, I link administrative records of student applications, admissions, and post-admission academic performance from a large public university in Uganda. These records include students enrolled in most of the STEM, social sciences, and business degrees in the years 2009-2017 at this prestigious national university that is centrally located and, by admitting students from across the country, creates a microcosm of Uganda's rich ethnic

[1] For example, cross country quality of government (Alesina and Zhuravskaya, 2011); cross country public policies (Easterly and Levine, 1997); productivity of a firm in Kenya (Hjort, 2014); public goods provision in Uganda (Habyarimana et al., 2007) Additionally, regarding public goods, Gisselquist et al. (2016) show that high ethnic diversity may lead to welfare gains. For ethnicity and social trust, see Alesina and La Ferrara (2000)

heterogeneity. For the purposes of this research, this feature is particularly interesting given the strong geographic segregation of ethnicities in Uganda, which means that many students arrive at the university with little prior exposure to other ethnicities but are suddenly surrounded by the full diversity that constitutes the country as a whole. In the analysis that follows, I classify students who graduated high school from their districts of origin as those with less prior exposure to other ethnicities and for whom the ethnic diversity on campus is most salient.

Being surrounded by coethnic peers at this large university might provide a sense of belonging and stability, thereby enhancing academic performance. Interacting with high-ability students can similarly improve performance in this setting because contact with instructors is limited (e.g., office hours are not offered), so learning from peers is important. In addition to testing the direct effects of coethnic and high-ability peers, I also estimate the interaction effect of these two peer types since homophilous coethnic sorting could hamper or help learning from peers depending on the academic ability of these coethnic peers. In this analysis, I rely on exogenous variation in the share of coethnic and high-ability peers in a given student's peer group to test for these direct and interaction effects.

The administrative data I use in this paper provide students' demographic and academic characteristics, including whether they were admitted on merit scholarships, which I take as an indicator of high ability. These records do not, however, report student ethnicity. I overcome this limitation by exploiting linguistic and cultural characteristics common to Uganda and SSA, where surnames reflect one's native languages and, thus, ethnicity. To do so, I apply a machine learning algorithm common in computational linguistics introduced in Cavnar and Trenkle (1994) and recently adapted by Michuda (2021) to the Ugandan context to a national administrative dataset of 2016 voter registrations that includes over 14 million Ugandans. This external data set provides training data I use to build a classification model that predicts ethnicities using student surnames.

This paper's causal identification of peer effects hinges on the random assignment of incoming students into dorms, which provides exogenous variation in peer groups. Specifically, a peer group in this analysis consists of students admitted to majors in the same school and assigned to live in the same dorm. Upon admission and conditional on gender, dorm assignment is random. Since there is excess demand for dorm beds, actual residence in dorms is not guaranteed. Some end up living off-campus, but dorm assignments shape campus life for some of these students, as they may engage in extracurricular activities within their assigned dorm. In addition to exogenous assignment to peer groups, the econometric strategy exploits idiosyncratic year-to-year variation in coethnic composition. Moreover, each student's course list is predetermined at the time of admission, and students do not meet their classmates and dormmates until orientation week. Thus, the results in this paper are not driven by selection into peer groups. I control for dorm, classroom (course-by-year), and major fixed effects to account for correlated shocks and differences that might confound my estimates.

The results of this analysis indicate that the coethnic peers are as important as high-ability peers in this setting, especially in the first year. That is, I find that coethnic peers (irrespective of ability) and high-ability peers (irrespective of ethnicity) increase a student's performance in the first year. Specifically, adding five coethnic peers to a peer group of size 25, which would increase the number of coethnic peers in a group from the twenty-fifth to the seventy-fifth percentile, increases a student's performance by 0.19 percentage points. Additionally, the same change of adding high-ability peers to a group of 25 increases a student's performance by 0.15 percentage points. Both effects are significant at the 5% level and are about 0.02 standard deviation change in a student's performance in the first year.

Nevertheless, the effect of coethnic peers disappears by the time a student graduates but that of high-ability peers persists and even increases. Specifically, the effect of coethnic peers in the third year, which is the final year for almost all the majors in this setting, is half of that observed in the first year. Yet the effect of high-ability peers in the third year is 1.5 times that of the first year in magnitude. Lastly, although I find that suggestive evidence shows that coethnic peers matter more than high-ability noncoethnic peers as a student advances during university education, the effects of both types of peers are statistically indistinguishable.

Beyond the average effects, heterogeneous impacts indicate that the effect of coethnic share is mostly driven by students of assumed high ethnic salience. For example, adding five coethnic peers to a group of 25 increases the academic performance of a high ethnic salience student by 0.05 standard deviations, which is 2.5 times the average effect. Nevertheless, like the mean effect of coethnic peers, this effect on students with high ethnic salience fades as a student progresses. On the contrary, coethnic peers have a positive and significant effect on high-ability, not low-ability, students that persist into the third year, suggesting that the benefits of coethnic peers throughout a student's university career can be reaped by those posed to succeed when they enter university.

Qualitative insights from the specific university context of this study align with potential underlying explanations for these results, including peer-to-peer learning and cultural and psychological factors. In this setting, both coethnicity and academic ability are readily and generally observable. Incoming freshmen can easily identify coethnic peers through physical features and cultural characteristics, including names and language. As the academic year unfolds, they also learn who among their peers are high-ability because publicly posted scores and grades reveal academic merit scholarship status or through frequent interactions. Given the prevalence of ethnic student organizations and activities on campus, which suggests a degree of homophily that shapes student life, it is natural for incoming students to seek out coethnic connections and support. Such connections can be critical to a student's successful transition to a novel setting of high ethnic diversity and, possibly, latent inter-ethnic tensions that may prevail on campus.

This university setting is also characterized by classical lecture-style instruction with few opportunities to interact with faculty or consult with teaching assistants, which makes informal peer-to-peer learning especially important. For both incoming and continuing students having high-ability peers in one's peer group can therefore provide an advantage. If anything, the benefit of such informal peer tutoring increases as students progress to more advanced courses in their degree programs.

The finding that a higher coethnic share matters on average and especially so for students of higher ethnic salience is suggestive of other psychological mechanisms. Enrolling at a large, centrally located national university may increase ethnic identity salience and attachment, as social identity theory in Tajfel (1982) predicts. The presence of coethnic peers in one's university environment may thus be beneficial for such students. This is similar to the finding reported in Okunogbe (2018), showing that the ethnic pride of Nigerian youth increases when they do national service in a region where they are not part of the ethnic majority. Also, since several students are forced to navigate a space that is diverse compared to their pre-university schools, having coethnic peers precludes inter-ethnic barriers. Moreover, I find the differential effect of the share of coethnic peers on students of high ethnic salience observed in the first year disappears as a student progresses. This indicates that through frequent cross-ethnic interactions at the university, these types of students make cross-ethnic networks and factors other than the shared ethnic identity of peers begin to matter more for academic performance. This phenomenon can be interpreted by the contact hypothesis in William (1947). This might also explain why the effect of high-ability peers increases over time.

This paper contributes to several strands of literature on peer effects in college. Although mixed, prior evidence largely indicates that post-secondary peer effects meaningfully impact education outcomes, such as major choice and academic performance. For example, Zimmerman (2003) and Sacerdote (2001) exploit random roommate assignments at US colleges to study roommate peer effects. Zimmerman (2003) finds significant but small peer effects when using pre-treatment academic characteristics to measure peer quality and also detects nonlinear effects that are conditional on the student's SAT scores. Sacerdote (2001) finds null effects using the ability of a peer but significant nonlinear effects at Dartmouth on academic outcomes. Additionally, he finds strong effects on some social outcomes (e.g., fraternity membership). Carrell et al. (2009) argues that roommates are a small part of one's college life, which might explain why Zimmerman (2003) and Sacerdote (2001) find no strong dorm or roommate peer effects. Exploiting exogenous assignments at the United States Air Force Academy, where students are assigned to peers with whom they spend a majority of time together, Carrell et al. (2009) find stronger academic peer effects than roommate peer effects. More recently, Mehta et al. (2018) use a panel data set that tracks students' time allocation and friendships at Berea College and found that peers have an effect on study efforts.

These peer effects studies primarily focus on college peers in the West. Their main econometric specifications include the average quality of peers measured by pre-treatment academic characteristics on the right-hand side variables. Given the setting, these papers also control for race, usually a binary indicator for white or black. In the SSA region, however, high ethnic diversity introduces new complexity and nuance to peer effects. For high-ability noncoethnic might have a negative or null effect on academic performance if high ethnic diversity leads to inter-ethnic rivalries and discrimination that spill into classrooms. I find the opposite: the identity of a high-ability peer does not matter. High-ability peers (irrespective of ethnicity) affect a student's academic performance, suggesting that peer effects observed in studies in the West also exist in this setting. I find that coethnic peers are also important in the first and second years.

This paper also contributes to the literature exploring the role of ethnic diversity on economic and social outcomes in SSA more broadly (Easterly and Levine, 1997; Habyarimana et al., 2007; Alesina and Zhuravskaya, 2011; Miguel, 2004; Gisselquist et al., 2016; Alesina and

La Ferrara, 2000; Håkansson and Sjöholm, 2007; Hooghe, 2007). In contrast to these more general studies, this analysis focuses on a different question, albeit one with clear importance and policy relevance. Understanding peer effects from social networks play out in higher education institutions with high ethnic diversity may enable more informed admissions and other academic processes, which often feature explicitly or implicitly in facilitating (or potentially undermining) cross-ethnic cooperation among young adults. High-ability peers affecting academic outcomes more than coethnic peers as students progress may suggest that Ugandan youth are less ethnically biased or able to adapt to ethnic diversity. However, it is important to note that coethnic peers might have lasting impacts on social networks outside school or other outcomes that are unavailable in my data.

Peer effects in higher education in SSA have been understudied for several reasons, including data constraints. A few studies that have explored college peer effects in the region use data from a South African university (Garlick, 2018; Corno et al., 2019). Nevertheless, Garlick (2018) focuses on peer effects under two different assignment rules (random and residential tracking), while Corno et al. (2019) focuses on how exposure to roommates of another race changes one's stereotypes. Race (white vs black) is salient in South Africa for historical reasons and general population composition, unlike other African countries. Therefore, I add to the literature by studying higher education peer effects at a university in a context about which we know very little. I find that in this setting, coethnic and high-ability peer effects exist, especially in the first year.

2 Background

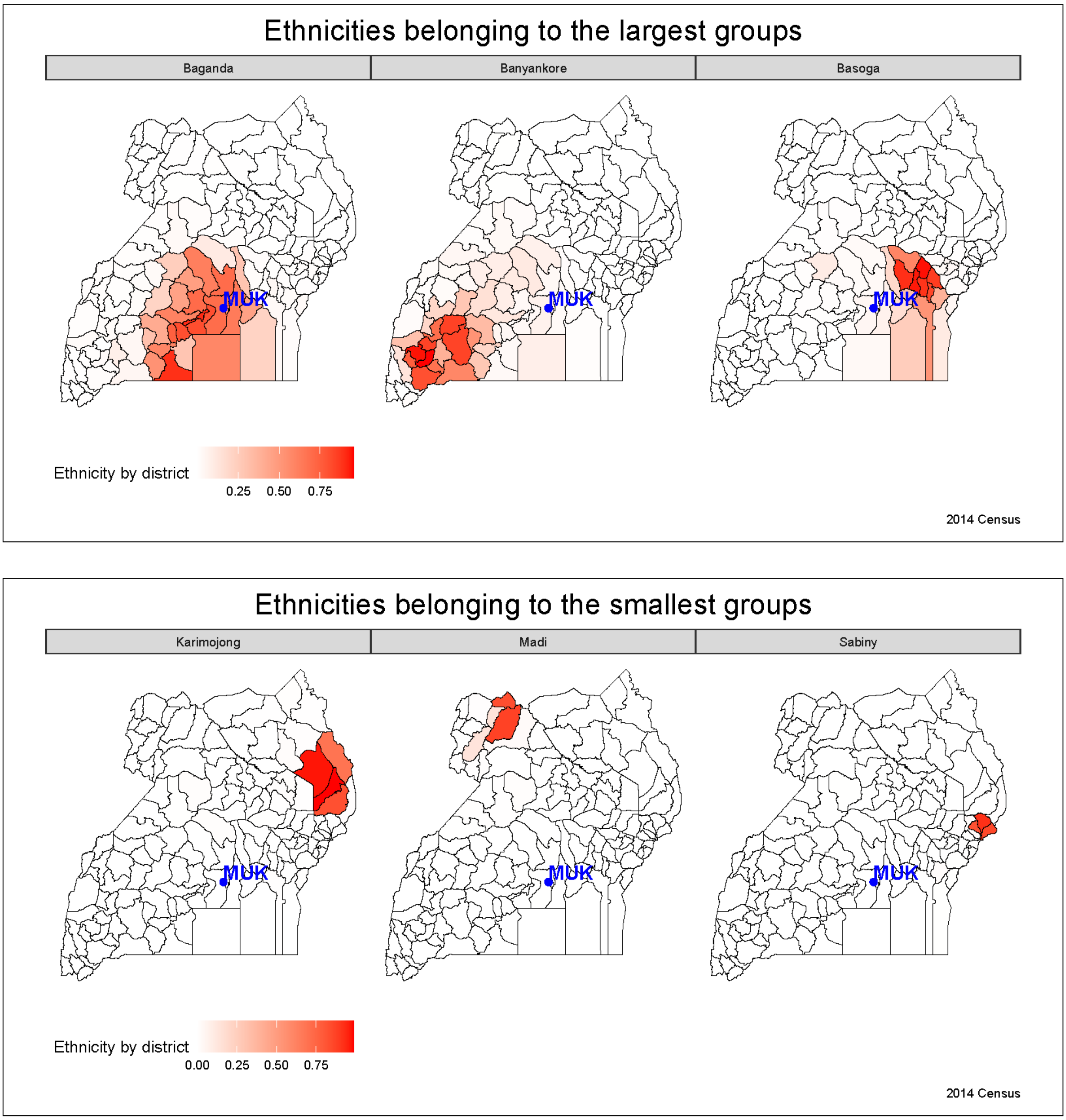
2.1 Ethnicity in Uganda

Uganda has over 50 ethnic groups that belong to three broader Bantu-speaking tribes (UBOS, 2006). The largest nine ethnic groups constituted 71% of the population according to the 2002 Uganda Population and Housing Census. Groups may differ by traditions (e.g., dressing), language, food, economic activities, and sometimes by physical characteristics (e.g., skin tone). This pronounced ethnic diversity is also characterized by distinct geographic segregation as shown in Figure 1. Indeed, Alesina and Zhuravskaya (2011) rank Uganda the 4^{th} most segregated countries in the world based on a spatial segregation index.

Historic migration and ethnic kingdoms drive these segregated settlement patterns. Bantu-speaking groups are clustered in the country's South, Central, and Western parts, while Nilotic and Nilo Hamites peoples are clustered in the Northern and Eastern parts. For purposes of the analysis that follows, I retrace current ethnic borders to historic kingdoms (see Appendix Section 8.2). Inter-region migration is limited except for rural-to-urban migration into the capital, Kampala, for economic opportunities. By contrast, rural-to-rural migration across ethnic clusters rarely opens economic opportunities and is limited due to cultural reasons.

Although ethnic divisions existed in pre-colonial Uganda, some were exacerbated during British colonialism (Tornberg, 2013). The first post-independence government made efforts to reduce the importance of ethnic identities by abolishing historic kingdoms and preaching national unity, an effort that met with resistance from some kingdoms, especially those with

Figure 1: Geographic Segregation.



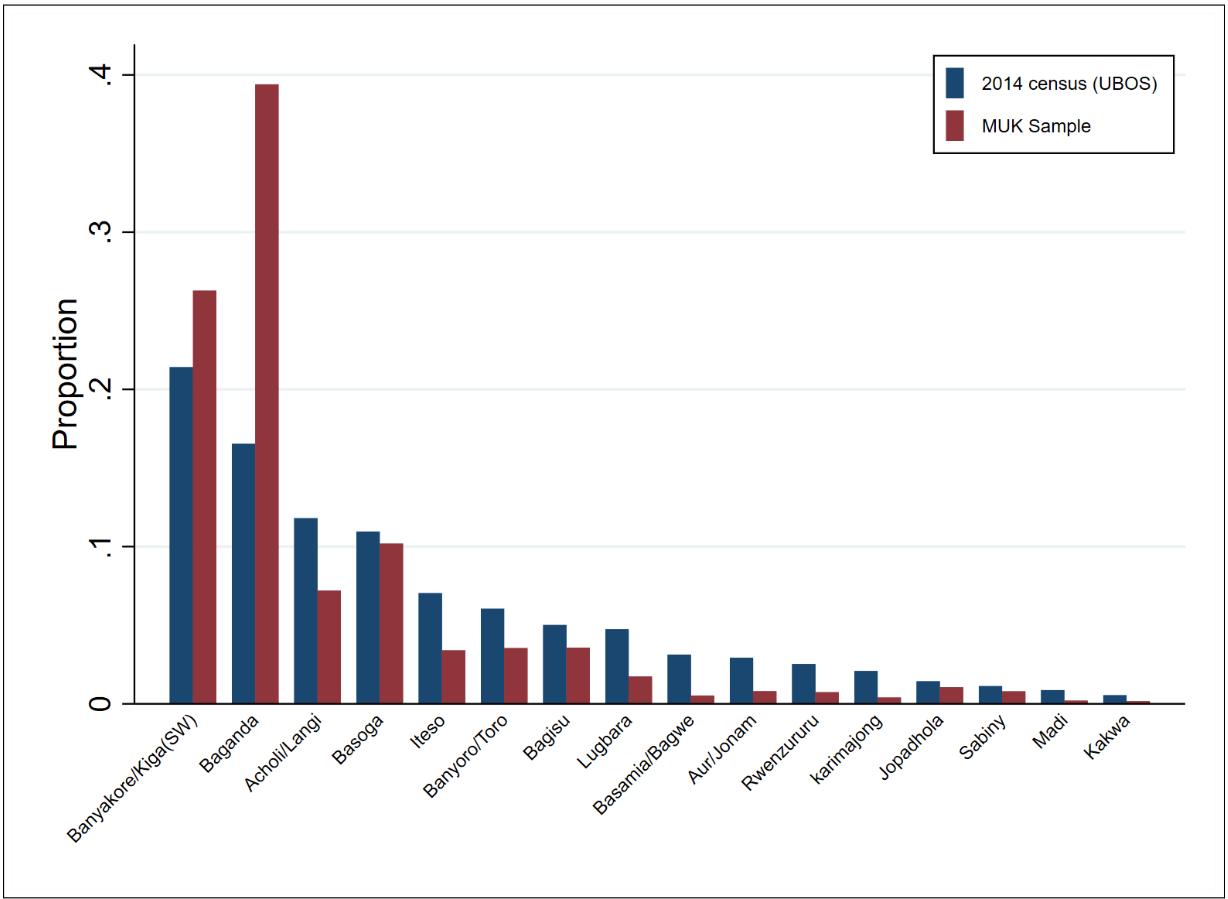
Notes: Ethnicity by district is the proportion of each ethnicity within a district. Data source: 2014 census. District shape files can be downloaded from https://data2.unhcr.org/en/documents/details/83043.

economic or political power. The current government allowed ethnic groups to reinstate their historical kingdom; some ethnic groups did. While current inter-ethnic competition and recent historical conflicts can be traced to political and sometimes historical factors (Mamdani, 2001), inter-ethnic competition or outright conflict is generally not as intense as in neighboring countries.

Although English, the official language of Uganda, is spoken in public offices and taught in schools, native linguistic diversity is high.[2] Differences between native languages are correlated

[2]WorldAtlas reports Uganda's language diversity index of 0.929, which indicates that most Ugandans speak at least one native language.

Figure 2: Distribution by ethnicity: student sample vs general population



Source: MUK admissions, 2009-2017 and 2014 Census (UBOS)

with physical distance, implying that one may partially comprehend the language of a neighboring tribe. Luganda is the most familiar native language because it is native to the Kampala region. I exploit this language diversity to predict ethnicity in Section 4.

2.2 Ugandan Higher Education and Makerere University Kampala

Although Uganda has one of the youngest populations in the world, post-secondary school education is low: the post-secondary enrollment rate for college-age Ugandans was only 6.85% during the 2017/18 academic year (NCHE, 2018). Nine public and 44 private universities offered degree programs during the 2018/19 academic year (NCHE, 2018), of which Makerere University Kampala (MUK) ranks first in quality and size.

MUK is well-known in the SSA as it is one of the oldest universities in the region. It was established in 1922 as a technical school to facilitate training workers for the British colonial government. It is centrally located in Kampala and admits students from across the country. For some students, it is at this university that they meet and interact with people of different ethnicities for the first time. With the exception of Baganda, the diversity of the MUK student population mirrors that of the country as a whole (see Figure 2).

3 Empirical Setting

3.1 Applications, Admissions, and Sample Definition

The Ugandan public university and pre-collegiate nationwide system offer a unique setting that I exploit to identify coethnic peer effects. First, national pre-collegiate exit exams and public university merit scholarships are centrally administered. Second, the Uganda Examination Board, an organization separate from MUK, runs an algorithm for all MUK admissions. Thus, there is no room to manipulate the composition of its student population.

Students are admitted under two schemes: (I) National merit scholarship and (II) self-funding scheme. A student lists up to six majors in order of preference during application. Admission to a major (cutoffs) is a function of the student's preference set, admission in national exams, and the university's capacity. A student's major (and course list) are predetermined during admission, 3-4 months before enrolling. Each major non-extension major is housed within a school, which is a smaller unit within a college. A school is locally termed as faculty or department, but I will adopt the 'school' term for simplicity.

Students in the same majors take almost all their first year classes together since courses have predetermined sequencing. Still, they interact with students from other majors within classrooms, usually within the same school, who share the same course requirements on a daily basis. In addition, students within a school usually share common spaces, such as computer labs, food canteens, study rooms, and libraries.

Students cannot select into different sections within the same major, as sections do not exist in this setting. Because of this, most student's course sequence is also pre-determined before a student reports to campus. The performance data show that over 98% of classes in each year are non-elective. Moreover, The university offers evening and day class options as 'different' majors when a degree, such as business administration, is in demand. Still, the day class is a 'different' major from the evening class, and students must apply and get admitted to either the day or evening class cohort separately. For example, students who intend to obtain a Bachelor of Business Administration degree can apply and be admitted to either the day cohort or the evening cohort. Students admitted to the day cohort cannot take classes and sit for their exams with students admitted to the evening cohort. I restrict the sample to day cohorts as evening majors do not qualify for the national merit scholarship. This is important because the merit scholarship is my measure of high-ability as I define in the coming sections.

Students stick with the majors offered during admissions but can apply to change within the first two weeks of their freshman year. Approvals depend on the capacity of the intended major and student performance and are thus rare. I find major change cases are less than 2.75% in the ten-year period of my sample.[3] Non-STEM majors, especially business and social sciences, tend to have relatively large class sizes.

[3] I compare the major student's enrollment and the major at the admissions and find a mismatch of 2.75%. This number includes students whose major switch applications were approved and possibly some data entry errors when entering admissions data.

3.2 Dorm Assignment and Defining Peer Groups

Conditional on gender and upon admission, dorm assignment is random. MUK has nine single-sex large dorms: three are female and six are male dorms. There are more incoming students assigned to dorms than there are beds to accommodate them. I observe dorm assignments but not the subsequent residence status and room assignments. Each student's admission letter indicates the assigned hall, which determined by the administration by simple random assignment. Students must formally apply to their assigned dorm for residence, at which point a dorm administrator and committee allocates beds according to a university-wide priority list that favors students on national merit scholarships in majors and schools perceived to be especially rigorous, such as medicine and engineering.

The remaining beds are then assigned to students according to the order of their dorm application. While students who are not allocated a bed in their assigned dorm must arrange for their own housing off-campus, their initial dorm assignment continues to shape campus life as assigned students have access to shared spaces with entertainment and dining facilities in these dorms. Extracurricular activities such as student government elections are also organized by dorm assignment irrespective of residence.

In general, a peer group consists of individuals with shared or similar characteristics who interact in social or other settings. Specific definitions of peer groups are context-specific. Carrell et al. (2009) define a peer group as a squadron at the US Air Force Academy, while Foster (2006) consider a peer group to be students living on the same dorm floor and Pre-collegiate studies, such as Carrell and Hoekstra (2010), use a cohort definition. In the MUK context, I define a peer group, as illustrated in Figure 3, as students within the same school that are assigned to the same dorm. Although most of the prior studies on college peer effects observe roommates, I do not observe room allocations and residence status, so I restrict my definition to dorm assignment.

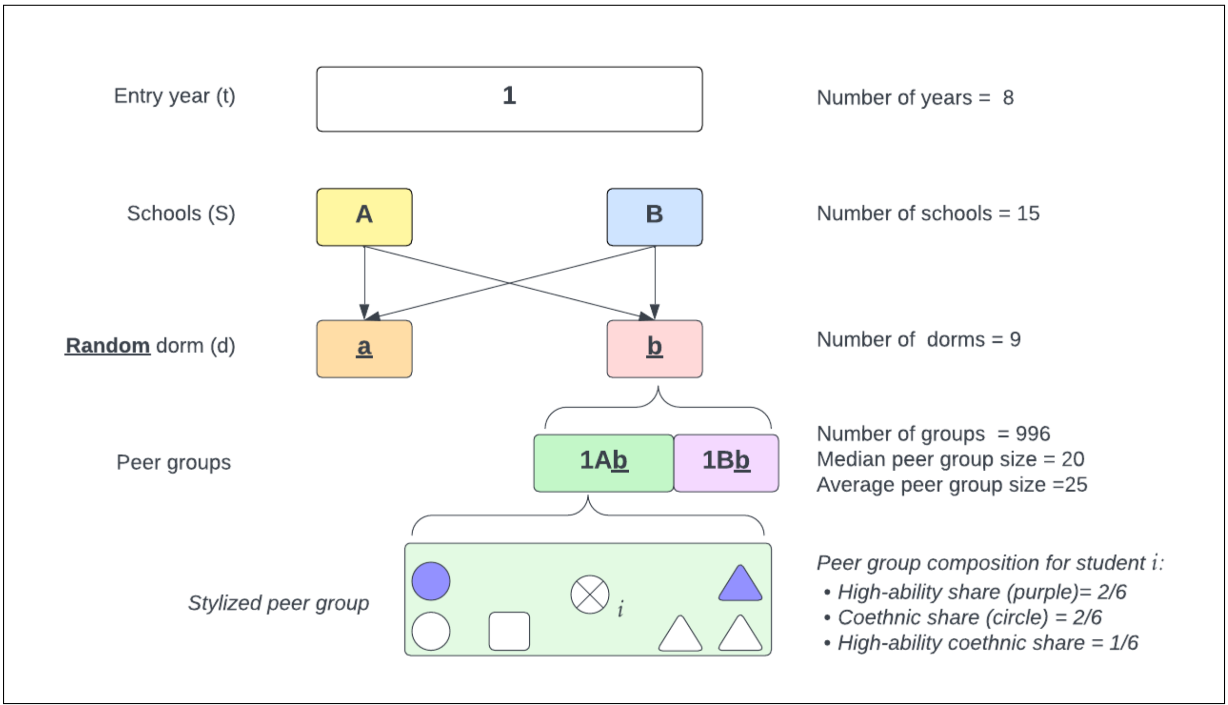
By focusing on the cohort-residential peer groups, I use a "strict" definition of a peer group, but it also allows me to study peers with whom a student spends most of their time. For instance, students within the same school and dorm may spend a lot of time together, such as walking to and from classes, attending classes together, and sharing common spaces at school and within the dorm. One may argue that a major hall year is a better peer group since students in the same major take 100% of their classes. However, the focus of this paper is coethnic and high-ability shares, and using a way smaller peer group definition reduces variation in the coethnic share as most of the coethnic share of the smallest ethnicities will be zero in a lot of peer groups.

3.3 Identifying Peer Effects at MUK

Estimating peer effects may be econometrically challenging for three reasons: self-selection (Hoxby, 2002), endogeneity (Manski, 1993), and correlated common shocks (Bramoullé et al., 2009). This section highlights the characteristics of this setting that provide solutions to these issues related to measuring peer effects.

Self-selection arises when people choose to join a group based on some pre-treatment characteristics. As stated in Hoxby (2002) *".. if everyone in a group is high achieving, many observers assume that achievement is an effect of belonging to the group instead of a reason*

Figure 3: Peer Group Construction



Notes: This diagram illustrates a peer group definition. Students in these defined peer groups are much more likely to interact regularly with each other, including those of the same or different ethnicity. High-ability students are defined as those on merit scholarships, a status that is widely known among all students.

for belonging to it." In the case of colleges, self-selection exists because students can select into classrooms, majors, and sometimes, dormitories. Peer effects literature typically employs two strategies to deal with selection in peer effects papers. First, conditional on some pre-treatment characteristics, such as gender and ability, peers result from random assignment (Sacerdote, 2001; Zimmerman, 2003; Carrell et al., 2009; Foster, 2006). However, random assignments into classrooms in US studies are difficult. So, with the exception of Carrell et al. (2009), these studies use a setting where roommates at some universities are randomly assigned.

The second approach involves exploiting natural variation in a cohort or group composition. The idea behind this approach is that year-to-year variation (e.g., gender, race, and class size) observed at the group level is a reflection of a natural variation in a general population (idiosyncratic). This approach has been used in pre-collegiate peer effects studies (Carrell et al., 2018; Hoxby, 2000)

My approach leverages characteristics of this setting described in Section 3.1. Peers are classmates who potentially live together. As aforementioned, conditional gender dorm assignment at MUK is random. Since I do not observe roommate assignments and residence status, this paper estimates the intent to treat (ITT) of the peer qualities defined later. Unlike most US universities, students do not select courses or majors post-admission, which has the convenient feature that students do not sort into classes or classrooms based on characteristics or exposure (or not) to different types of peers.

The reflection problem is the endogeneity problem challenge, which arises from a feedback loop of peers. This is a challenge because a student's and their peers' outcomes are simultaneously determined. One of the approaches in the literature is to use preexisting characteristics that are exogenous to the dependent variable, such as race and gender. For example, Carrell and Hoekstra (2010) uses the presence of family problems when studying peer effects of children linked to domestic violence on academic outcomes. I use pre-collegiate characteristics, as most of the literature, to exploit exogenous variation in treatment variables, which are coethnic share and high-ability share within a student's peer group.

In Uganda, students' ethnic identities are determined at birth. An argument may be made that ethnicity is part of multifaceted identities, a function of collective cultural traits, and that an individual's ethnicity may change through self-identification (Sen and Wasow, 2016). I am not concerned that this exists in Uganda to the extent that it would confound my estimates. First, I follow the official categorizations of ethnic groups in UBOS (2006), and admissions do not have ethnic quotas or any form of affirmative action based on ethnicity. Thus, there is no incentive to change one's ethnic identity during university applications. Second, I use linguistic characteristics to predict ethnicity instead of self-identification. I describe these variables in Section 4.2 below.

The last main challenge is contemporaneous common shocks, especially if they are correlated with academic performance. My setting uses random assignments at the same university, which reduces the possibility of such shocks. Nevertheless, there may be shocks that affect some peer groups differently. Thus, the main regressions include all group fixed effects, such as dorm and classroom, to account for observed characteristics that might confound the main effect.

4 The Data

4.1 Academic and Demographic Characteristics

4.1.1 Pre-university Characteristics

The analysis in this paper uses several data sources: MUK's administrative records on academic and demographic characteristics observed from applications, admissions, and post-admission academic performance for students entering the university during 2009-2017, and ethnicity is predicted by student surnames.

I observe students' application data from 2009 to 2017. The student applications include the student's name, type of application, admission scheme (merit scholarship or private scheme), and offered majors, as well as age and religious identity. All student records are de-identified pre-analysis, although most student admission data, such as major, are publicized on university notice boards and in newspapers.

4.1.2 Measure of High-ability

Every year, 4,000 students are admitted to public universities on a government merit sponsorship basis of performance in high school national exams, most of which enroll at MUK relative to other public universities (HESFB Uganda, 2012). These scholarships are awarded to the top students within a major, and the number of spots per major is relatively constant across years. Merit scholarship application forms are submitted at the time of national exam registration before students take their exams. Therefore, almost all A-level graduates are automatically considered for the government merit scholarship, as the sponsorship does not require a separate application. Students are ranked based on their high school GPA within their preference set, and the top students are offered a scholarship until each major's scholarship spots are filled up. That is, the scholarship is determined by high school GPA.[4]

High school GPA is a proxy for ability as it may pick up a student's innate ability, effort during high school, and success in an academic context. I therefore use this as an imperfect but informative proxy of "academic potential," which accounts for both a student's subject combination and performance in this selected subject combination. Each major has a high school subject combination required for a student's successful college career from a university's perspective. I define "high-ability" students within each major as those enrolled with the national merit scholarship. Lastly, it is usually public knowledge which of a student's peers are admitted through merit as university registration numbers differ by merit status. Moreover, admission lists are usually published in newspapers and university notice boards.

4.1.3 University Academic performance

I observe student transcripts from 14 departments belonging to six colleges. Each student's transcript lists all courses taken, credit units, and performance in percentages by semester year of study during which the course was taken. Therefore, I can observe these students' classmates and how they have progressed from matriculation to completion of coursework. Unlike schools in the West, letter-grade ranges assignment is the same across all majors, and professors do curve grades. Professors at MUK do not assign letter grades. They submit each student's performance on a 0-100% scale, and the central system assigns the letter grades. Also, most majors take three years to complete, and thus students take a lot of courses per semester (a min of six, and some majors require students to take up to ten courses in some semesters).

4.2 Ethnicity

University applications and admissions do not capture the ethnic identity of students, although ethnicity is one of the most salient identities among Ugandans. I overcome this by exploiting linguistic differences reflected in surnames. Ugandans' surnames are in their native languages.[5] This naming pattern is not random or unique to Uganda. Historically African parents chose names intentionally. However, with the arrival of colonists, first names are now in foreign languages, such as English (in Anglophone countries) or French (Francophone countries). The meanings of most Ugandan surnames can be traced to the father's tribal clan and religiosity or prevailing conditions at the time of birth, among others. These are linguistic characteristics I use to predict one's ethnicity. Data Appendix 8.2 describes how I trace ethnic boundaries from

[4]There are a few variations. For example, Ugandan public universities have a gender affirmative action policy that awards a 'free' 1.5 additional points to every girl during admission. This 1.5 free point is also awarded to girls when they are being considered for non-merit admission schemes. In addition, a small proportion of the merit scholarship is awarded through the district quota to the top four students graduating from their district of origin who did not obtain the merit scholarship through the direct route. Therefore, the number of district quota spots is proportional to ethnicity size. District quota applications are made at the same time as the national merit applications.

[5]Trevor Noah mentions the same pattern in South Africa in his book "Born a Crime" (PP.). Also, see this https://www.bbc.com/news/world-africa-37912748 for another example

current administrative units to historical kingdoms.

Using surnames to trace one's identity is not new in economics and other fields. For example, surnames have been used in mobility studies to trace wealth across generations within a family in the West (Barone and Mocetti, 2016; Clark and Cummins, 2015). Some studies have also used surnames to predict ethnic identity across several countries. For instance, Bhusal et al. (2020) use surname frequency in the Nepalese historical censuses to predict one's caste in their paper studying how revolutions may have altered political representations and inclusion in Nepal. Using fuzzy matching and naïve Bayes machine learning techniques on historical records, Monasterio (2017) studies surnames and ancestry in Brazil.

4.2.1 Predicting Ethnicity and Constructing Coethnic Share

More related, Michuda (2021) exploits rural-urban linkages in Uganda and applies machine learning on representative Uganda surnames to predict the rural origin of Uber drivers in Kampala. His study explores how Uber drivers adjust their online hours when their probable ancestral homes experience a negative weather shock. Therefore, agroecological zones form a basis for his predictions. His procedure, like the machine learning section in Monasterio (2017), follows a text categorization procedure developed by Cavnar and Trenkle (1994). This process has been widely used in computational linguistics and involves breaking down a name into N-grams.

Following the literature, there are two common approaches: use frequencies to predict probabilities as in Bhusal et al. (2020) and train a machine learning algorithm on some training data set by applying tools, such as gradient boosting. Method (I) computes simple probability using the frequency of each surname. Suppose $\{E_1, E_2, ..., E_n\}$ is a set of all ethnicities in a population. Also, suppose $N_{s \in E_i}$ is the number of times a surname, s, belongs to an ethnicity, E_i. Then the probability of belonging to a particular ethnicity is computed as:

$$\frac{N_{s \in E_i}}{\sum_{\forall n} N_{s \in E_i}} \tag{1}$$

To illustrate, consider the surname "AHIMBISIBWE": it appears 17,559 times in the name training data, of which 13,904 occurrences in the Ankole region/ethnicity. Therefore, there is a 79.2% probability that a student with the surname "AHIMBISIBWE" is of Ankole ethnicity.

Method (II), which is my preferred, follows Michuda (2021) and computational linguistics and begins by breaking a surname into N-grams. Taking "AHIMBISIBWE" as an example, Method (II) breaks this surname into 1-grams ("A", "H", " M", "B", "I", etc); 2-grams ("AH", "HM", "MB", "BI", etc); 3-grams ("AHI", "HIM", "IMB", etc..) and so forth. The algorithm can now count the number of frequencies each n-gram appears in a surname and in each region. The most common weighting approach used in linguistics is the term frequency-inverse document frequency (tf-idf) that combines approaches developed by Luhn (1957) and Jones (2004). I then apply gradient boosting on N-grams and tf-idf features on an external data set described in Section 4.2.2, producing a classification model that I apply to students' surnames.

The second method is preferred to the first in this paper for two reasons. First, by following the tf-idf weighting procedure, the algorithm picks each surname's most unique linguistic characteristics. Second, it does not require an exact match in the name database.

This algorithm predicts N probabilities if we have N ethnicities. Given that languages are not exclusively unique, some probabilities are non-zero or 1. We can, therefore, interpret the predicted probabilities as a measure of 'confidence' that a student belongs to a particular ethnicity. One approach is to use ethnicity corresponding to the top predicted ethnicity (the ethnicity a prediction is most confident about), as is common in the literature.

I can then compute the share of coethnic peers using two approaches: (A) and (B). Firstly, by single ethnic identity assignment (A), I assign an individual a single ethnicity category corresponding to the group the algorithm is most confident about. This is common in studies employing machine learning algorithms to predict ethnicity, religion, or area of origin in the literature. This method assumes that individuals' top predicted ethnicity corresponds to the 'true' ethnicity and treats ethnicity as a categorical variable without considering potential measurement errors. Using top predicted is common in the literature. The average probability corresponding to ethnicity in the algorithm is most confident equals .792 (median=.861), which is high.

Secondly, by joint probability estimation, I consider all the probabilities that a given surname belongs to different ethnicities. This method acknowledges potential measurement errors associated with using categorical variables for ethnicity. It estimates the probable fraction of coethnic peers in a peer group by considering the joint probabilities of two individuals belonging to the same group.

That is, student i's share of coethnic peers in a peer group G, S^E_{iG} is computed as:

$$\text{Using category assignment (A)}: \quad S^E_{iG} = \frac{\sum_{k\forall\neq i} \text{Number of coethnics}}{N_G - 1} \tag{2}$$

$$\text{Using joint probability estimation (B)}: \quad S^E_{iG} = \frac{\sum_{e=1}^{16} \sum_{\forall k\neq i}^{N_G-1} \Pi_{e_i}\Pi_{e'i}}{N_G - 1}, \tag{3}$$

where N_G is the peer group size, Π_{ei} is the predicted probability that an individual i belongs to ethnicity group e. Lastly, I collapse ethnicities to 16 ethnic/language groups as described in Appendix 8.2. The main analysis uses the probable coethnic share in a group in equation (3), but the results remain unchanged when I use the share of coethnic peers computed using equation (2). Throughout this paper, I use the share of coethnic peers and the probable share of coethnic peers synonymously for simplicity and ethnicity to mean the most probable ethnicity in the empirical and results sections.

4.2.2 Training Data

I use nationwide voter registration to train the machine learning model (gradient boosted). These data contain names, voter ID numbers, date of birth, sex, polling station, and area of residence. The area of residence is given for all units of administration parish, sub-county, county, and district. I link these voter data to spatial administrative and public data containing ethnic boundaries traced from historic kingdoms described in Appendix 8.2.

People register to vote from a polling station within their parish of residence. Moreover, in many cases, people who live in cities outside their areas of origin often register to vote in their ancestral homes. Because voter registration is manual, it only takes place once between elections. A few Ugandans own cars to travel, so most walk, as long-distance public transportation is costly. Thus, the cost of registering to vote in a village different from their residence is high.

4.3 Descriptive Statistics

Table 1 provides summary statistics for demographic and academic characteristics in Panel A and peer group averages in Panel B. About half of the student population is female, and about 31% are high-ability (enrolled through national merit scholarships). Most of the students have declared religion, and as expected in this context, most students are either Catholic or Anglican. The average age of incoming students is 20. Lastly, about 36% of the students in the sample graduated from a high school with their home district (these are the type of students I assume to have higher ethnic salience). Because of Uganda's high ethnic segregation (Figure 1), this group of students may not have interacted with peers of different ethnic groups.

Given my peer group definition, the average group size is 25, which is small, albeit the SD for peer group and cohort sizes are large. Close to 75% of the peer groups are of size 50 and below, as the Appendix Figure 6 portrays. STEM majors, which comprise most of my sample majors, usually admit a few students relative to other majors.

The average coethnic share is 25%, and the average co-ethnic share on merit is 7%. To explore how treatment intensity (shared ethnicity) may vary by ethnicity and if MUK is representative of Uganda's ethnicity distribution, I plot the distribution of ethnicities in Figure 4. The CDFs in this figure show that 80% of peer groups have a coethnic share of 0.2 or less.

5 Empirical Strategy

5.1 Mean Effects

5.1.1 Direct Effects of Coethnic and High-ability Peers

As aforementioned, a peer group refers to students admitted to majors in the same school f and assigned to dorm d in year t. For simplicity, I will index the peer group fdt with G in the estimation equations in this section. To estimate the direct effects of coethnic or high-ability peers on academic outcomes, I use a model that exploits variation in coethnic composition across peer groups within a year and year-to-year variation:

$$y_{ijcG} = \beta_0 + \phi_1 S^E_{iG} + \phi_2 S^H_{iG} + \beta_2 \mathbf{X}_{iG} + \beta_3 \bar{\mathbf{X}}_G + \delta_j + \alpha_c + \lambda_d + \theta_m + \gamma_s + \varepsilon_{ijcG}, \qquad (4)$$

where y_{ijcG} is the first year percent grade that student i of ethnicity j and belonging to group G obtained in course c. S^E_{iG} is the probable coethnic share of in i's peer group defined in Section 4.2 in equation(3), and S^H_{iG} is the share of high-ability peers (coethnic and noncoethnics). The main estimation controls for δ_j, which is i's most probable ethnic group, $\mathbf{X}_{iG}$ is a vector of i's background characteristics and includes i's own ability, and $\bar{\mathbf{X}}_G = \frac{\sum_{\forall k \neq i} X_{iG}}{N_G - 1}$ is the vector of

Table 1: Descriptive Statistics

	N	Mean	SD
Panel A: Student characteristics			
High-ability	25,487	0.31	0.47
Age	25,487	20.16	1.43
Female	25,487	0.48	0.50
High ethnic salience	25,487	0.36	0.48
Anglican	25,487	0.37	0.47
Catholic	25,487	0.31	0.47
Muslim	25,487	0.09	0.29
Pentecostal	25,487	0.06	0.24
Seventh Adventist	25,487	0.02	0.13
Unspecified Religiosity	25,487	0.05	0.22
Other Religions	25,487	0.01	0.09
Panel B: Peer group variables			
Peer group Size	996	25.64	21.64
High-ability share	996	0.35	0.24
Coethnic share	996	0.24	0.19
High-ability coethnic share	996	0.08	0.12
Low-ability coethnic share	996	0.15	0.15
Panel C: Course level			
All year grades (%)	1,061,905	67.83	9.61
Year One grades (%)	321,538	66.90	9.65
Year Two grades (%)	343,950	67.50	9.80
Year Three grades (%)	330,626	68.52	9.26

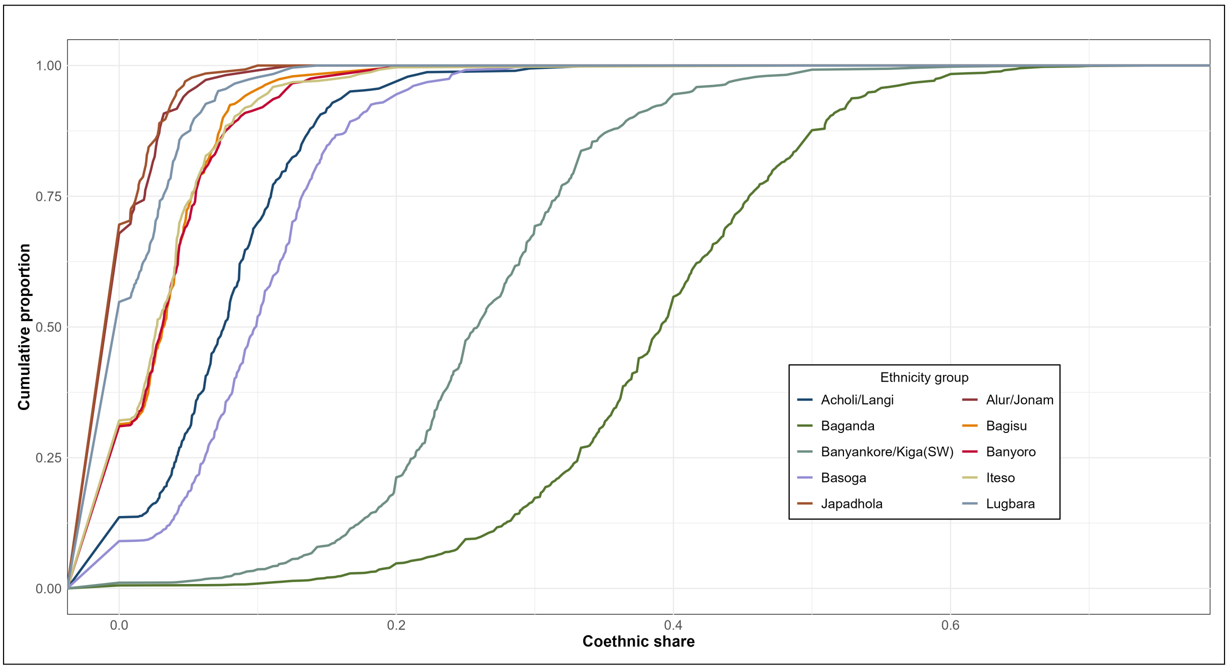
Notes: Data are from MUK and are restricted to students admitted to non-extension day majors at six colleges for 2009-2017, excluding 2010. A peer group comprises of students admitted to majors within a school major in the same year and assigned to the same dorm. Unspecified religion indicates whenever religious identities are not provided or entered as "Christian". Christianity is usually a correction of several or nondenominational religions in this context. Religion "Other" includes the smallest religions (where the count is less than 100 in the sample), such as Bahai, Jehovah's Witness, traditional religions, and Intambiro. Apart from age, Panel A variables are constructed to be binary.

exogenous variables (the average background characteristics of i's peers, except high-ability). Additionally, α_c, λ_d, θ_m, and γ_s represent classroom, dorm, major, and high school subject combination fixed effects (FE). Lastly, ε_{ijcG} is the error term. I cluster standard errors at the peer group to account for the potential error correlation across individuals in a group.

The coefficients of interest are ϕ_1, which captures the effect of attending lectures and potentially living with coethnic peers in this setting, and ϕ_2, which captures the effect of attending class and potentially living with high-ability peers irrespective of their ethnicity. I take several steps to ensure that ϕ_1 and ϕ_2 are unbiased. I control for several FE to deal with bias arising from correlated shocks.

First, correlated shock in this setting may arise from differences across classrooms. Therefore, I include classroom FE to control for unobserved differences in courses, such as performance, instructor effects, and classroom diversity. In addition, classroom FE should control for differences in major by year since the student's major and course list are determined at

Figure 4: Distribution of Coethnic Share across Peer Groups.



Notes: Data used to produce these distributions are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. Coethnic share is computed as the leave-me-out proportion of coethnic peers in a group. This figure plots the 10 largest ethnic groups by the total number of MUK students (out of the 16 total ethnic groups).

the time of admission. Nevertheless, students may take courses with peers admitted to majors outside their schools if cohort sizes are small and major course requirements are related. This implies nonrandom exposure to coethnic peers because of the systematic differences in the share of some ethnicities in the MUK sample and general population. Thus, α_{ct} also controls for this systematic difference in ethnic exposure across students in addition to controlling ethnicity FE. Relatedly, I include major FE, θ_m to control for differences between majors. Each regression will control for ethnicity, major, and classroom FE at the minimum in the results section.

Second, I control for dorm FE to control for factors, such as renovation and dorm conditions, that might affect academic performance. In addition, cultures are different across dormitories. For example, Ricart-Huguet and Paluck (2023) show that cultures, such as outgoing and academic mindedness, are different across MUK dorms to the extent that they affect interpersonal outcomes.

Third, although evaluated at the same cutoffs, students entering the same major may occasionally take different subject combinations during upper high school. Therefore, I include high school subject group FE, γ_s, to capture the differences in types of incoming students. When computing high school weighted GPA, each major has different requirements to capture incoming students' academic preparedness. Take Bachelor of Commerce, for example, the required HS subjects are math and economics, but students who take one of the two and those who take both can qualify if they perform above the cutoffs. Students graduating with math and economics have a higher perceived potential for success in Bachelor of Commerce classes than those graduating with one of the two subjects. Therefore, controlling γ_s captures the unobserved differences in academic preparedness across students in the same major.

Concerning self-selection, dorm assignment is random, and each student's course list and classmates are predetermined before entry at the time of admission, as mentioned in Section 3.1. Two lines of concern can be made for potential sources of self-selection. First, although dorm assignment is random, the on-campus residence may be biased to STEM students admitted through the merit scholarship. This is only statistically meaningful if merit scholarships are correlated with ethnicity and if dorm assignment was not random.

Correlation between ethnicity and obtaining a scholarship in a STEM major is possible if the top secondary schools are concentrated in one ethnic region, where students from that region graduate with the highest A-level scores to qualify for the merit scholarship. As Panel B of Figure 2 shows, the student population is biased towards the two largest ethnic groups. Coincidentally, the most elite secondary schools in the country belong to these regions because of historical reasons. Nevertheless, this is not an issue, as dorms and majors do not have ethnic quotas and equation (4) controls for i's probable ethnic group, which controls for differences in the levels of stratification. Also, when I regress merit ethnicity fixed effects, I find the explained variation is less than 1%.

Additionally, students may select into majors by manipulating the rank of their choices. This might cause selection into majors even though an organization separate from the university handles admissions and even though obtaining admission is quasi-experimental. This is possible since the ranking of program cutoffs does not change from one year to another, although actual cutoffs may change. This is not a concern as a peer group of classmates who potentially live together, and dorm assignment is random.

As a test, I provide balance tests in Table 2, which presents evidence against selection. Each column is an independent estimation similar to specification (4). I run these regressions at the aggregated to the student level (not course level). Each pre-university characteristic is regressed against the coethnic share in Panel A, while in Panel B, each pre-university characteristic is regressed against the high-ability coethnic share. Panel B also controls for a student's ability. Additionally, I regress the share of coethnic or high-ability peers on all the pre-university characteristics and report the estimates and the F stat in Appendix Table A4. The correlation between pre-university characteristics and the primary variable of interest would be high and significant if nonrandom sorting into peer groups existed.

From Table 2, the correlation between each student's characteristics and the share of co-ethnic peers (Panel A) or the share of high-ability peers (Panel B) is practically zero and not significant in all regressions, providing evidence that students are not selecting into peer groups. In addition, the F stats from Appendix Table A4 are small, 0.84 and 2.15 when the share of coethnic (column one) or share of high-ability (column two) peers are regressed against all the pre-university characteristics, respectively. This indicates that the results presented in this paper are unlikely to be biased because of nonrandom sorting.

5.1.2 Interpreting the Magnitude of ϕ_1 and ϕ_2

As Section 3.2 describes, not all students assigned to a dorm end up residing in their assigned dorm due to capacity constraints. Living off campus does not, however, exclude a student from dorm-based peer groups; it just alters the nature and frequency of interactions. Given my peer

Table 2: Evidence against Selection

	(1) Age	(2) Anglican	(3) Catholic	(4) Muslim	(5) Pentecostal	(6) SDA	(7) High Ethnic Salience	(8) Other Religion	(9) High-ability
Panel A: Coethnic share as the independent variable									
Coethnic share	-0.00	-0.01	0.02	0.01	-0.02	-0.01	0.04	0.00	0.02
	(0.11)	(0.04)	(0.03)	(0.02)	(0.02)	(0.01)	(0.04)	(0.01)	(0.03)
R-squared	0.13	0.07	0.05	0.27	0.03	0.03	0.11	0.06	0.17
N	25,487	25,487	25,487	25,487	25,487	25,487	25,487	25,487	25,487
Panel B: High-ability share as the independent variable									
High-ability share	0.06	-0.00	-0.02	-0.01	0.02	-0.01	0.03	0.01	
	(0.08)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	
R-squared	0.13	0.07	0.05	0.27	0.03	0.03	0.13	0.06	
N	25,487	25,487	25,487	25,487	25,487	25,487	25,487	25,487	25,487

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. Each column is an independent regression that regresses a pre-university characteristic against the share of coethnics. All regressions include school-by-year (not classroom), ethnicity, and hall FEs. Also, all regressions in panel B control for own ability. Standard errors are in parentheses and clustered at the peer group level.
*p<0.1, **p<0.5, ***p<0.01

group definition, consider two types of students based on the extent of interactions with others in a given peer group: fully compliant and partially compliant. Fully compliant students live in their assigned dorms and can thus interact as dorm residents with other fully compliant peers and, in other ways, with their partially compliant peers. Partially compliant students live off-campus and, therefore, do not interact as dorm residents with others in the peer group I construct for them. Both types of students are likely to interact daily (within and across each type) in classes and study groups.

If I observed residents, I could estimate the local average treatment effect of coethnic and high-ability peers using dorm assignment as an instrument for dorm residence to account for endogenous dorm residency. Since I do not observe residence, ϕ_1 and ϕ_2 in equation (4) are effectively reduced-form peer effects estimates based on dorm assignment. These reduced-form effects are a data-weighted average of the peer effects for fully compliant peers and partially compliant peers.[6]

I expect the reduced-form peer effects in Equation (4) to be less than the local average treatment effect of coethnic and high-ability peers. Also, the existence of partially compliant peers will likely attenuate peer effects. To illustrate the logic behind this claim, consider a parallel with Carrell et al. (2009), who use dorm floors to reconstruct peer groups at the Air Force. Their empirical setting allows them to construct pseudo-peer groups that span the relevant peer group (squadrons). Interactions are expected to still exist within these pseudo-peer groups but at a reduced rate than the true squadron-based peer groups. Although these pseudo-peer groups comprised 66.6% of peers from squadrons, the presence of peers with whom students interact less frequently attenuates estimated peer effects. Analogously, in the MUK context, I expect that reduced-form peer effects based on dorm assignment and, therefore,

[6]Even knowing the overall proportion of each cohort residing in dorms (i.e., residence compliance) does not necessarily solve this issue as I cannot use this compliance rate as the first stage to inversely weight reduced-form in the equation proposed in Bloom (1984) and as equation (A4b) in Appendix Section 8.3 without additional (strong) assumptions.

including both resident and non-resident students in peer groups will underestimate the true peer effects operating in this setting.

5.1.3 Coethnic and High-ability Interaction Effects

Although evidence in the literature is mixed, the effect of the high-ability peers (ϕ_2) is expected to be positive. Classrooms in the setting are relatively large, and services, such as office hours, do not exist, making peer-to-peer learning important. The sign of the coethnic share coefficient (ϕ_1) is largely ambiguous. The coefficient ϕ_1 could be zero on average, although it could be positive or negative for several reasons.

Bayer et al. (2020) show that minority students in introductory economics classes report a lower sense of belonging than non-minority students, and some studies (e.g., Walton and Cohen, 2011) show that interventions to increase a sense of belonging improve academic outcomes for students. Additionally, some students might suffer from imposter syndrome exacerbating their sense of belonging. Thus, having coethnic peers might be significant for some student as it may increase a sense of belonging during student interactions.

On the other hand, students of shared ethnicity may gravitate toward one another for cultural reasons, such as language, traditions, and beliefs. These homophilous tendencies and coethnic bias during interactions in diverse societies might lead to ethnic-based sorting into study and friendship groups. In this case, the effect of coethnic peers on academic performance will be indirectly through high-ability coethnic peers. For example, it might be detrimental for coethnic peers to isolate if they are, on average, low-ability compared to noncoethnic peers. A low-ability student might benefit from a higher share of high-ability than a higher coethnic share in a peer group.

To capture the effect of the pre-university academic quality of coethnics, I use an equation similar to equation (4).

$$y_{ijcG} = \beta_0 + \phi_1 S_{iG}^{EL} + \phi_2 S_{iG}^{EH} + \phi_3 S_{iG}^{E'H} + \beta_2 \mathbf{X}_{iG} + \beta_3 \bar{\mathbf{X}}_G + \delta_j + \alpha_{ct} + \lambda_d + \theta_m + \gamma_s + \varepsilon_{ijcG}, \quad (5)$$

where S_{iG}^{EH}, S_{iG}^{EL} and $S_{iG}^{E'H}$ are the probable shares of high-ability coethnic, low-ability coethnic, and high-ability noncoethnic peers, respectively. All other terms are the same as those in equation (4).

Therefore, the setting provides four sources of variation of interest in the share of peers that is: (A) high-ability and coethnic; (B) low-ability and coethnic; (C) high-ability and noncoethnic; and (D) low-ability and noncoethnic. Coefficient ϕ_1 in equation (4) captures the average effect of (A) and (B), while ϕ_2 captures the average effect of (A) and (C). If students are sorting on ethnicity when forming study groups, (A) should matter than (C). In such cases, we can think of the coethnic peers operating indirectly through high-ability ethnic peers.

5.2 Heterogeneous Peer Effects

To estimate heterogeneous effects, I estimate equation (4), including the interaction of the two treatments with the dummy variable that captures the heterogeneous dimension listed below.

$$y_{ijcG} = \beta_0 + \phi_1 S^E_{iG} + \phi_2 S^H_{iG} + \varphi_1 S^E_{iG} \times d_i + \varphi_2 S^H_{iG} \times d_i + \beta_2 \mathbf{X}_{iG} + \beta_3 \bar{\mathbf{X}}_G + \delta_j + \alpha_c + \lambda_d + \theta_m + \gamma_s + \varepsilon_{ijcG}, \tag{6}$$

where d_i can be gender, ability, or assumed level of ethnicity salience, while φ_1 and φ_2 are the differential impacts on d_i of coethnic and high-ability share, respectively. All other terms are the same as those in equation (4).

If coethnic and high-ability peers matter for academic performance, the average effects may be conceptually different depending on dimensions, such as the size of each ethnicity at the university and level of prior exposure to noncoethnic Ugandans. If increasing a sense of belonging is a channel through which coethnic peers might work, then the effect of coethnic peers could be zero for large ethnic groups who are less likely to suffer a low sense of belonging. However, coethnic peers may matter positively for small ethnic groups with limited exposure to different ethnicities prior to University.

Also, coethnic peer effects in the presence of high ethnic heterogeneity may also matter due to inter-ethnic impacts. For instance, there are ethnic groups that share values with other groups or portray less in-group bias. In such cases, fewer co-ethnic peers may not matter as those students would easily integrate with other ethnicities. More broadly, if inter-ethnic uncongenial relationships exist in Ugandan societies, they could spill over into schools, creating 'bad' peers. Nevertheless, this is unlikely in Uganda, as inter-ethnic tensions are not that common. The analysis will, therefore, explore heterogeneity in other dimensions.

5.2.1 Differential Effects by Gender

Several studies report differential peer effects on academic and non-academic outcomes by gender in several settings. For example, Carrell and Hoekstra (2010) study peer effects of kids exposed to domestic violence on test scores and disciplinary incidents in a classroom and find that peer effects are significant and stronger for boys, not girls. Additionally, Stinebrickner and Stinebrickner (2006) use HS GPA to study peer effects on sudy habbits and academic perforamcne at Berea College and find that HS GPA captures the effect on study habits and significant peer effects on girls. More recently, using a field experiment at a public school in Peru, Zárate (2023) finds low peer effects on academic outcomes but stronger on social skills, such as network connectivity, and psychological measures of social skills, such as altruism, which vary by gender.

Given the coethnic bias reported in the literature, it is likely that friendships are formed along ethnic lines. For example, using a setting in SSA similar to Uganda, Salmon-Letelier (2022) report that friendship networks form along ethnicity or religion lines in Nigeria's state and federal unity schools, respectively. If such homophily exists, it may create differential impacts by gender since friendship groups overlap with study groups.

There is also long-standing anthropological literature, such as de la Cadena (1995) exploring ethnicity and gender that finds women are more ethnic than men in the community of Cusco. Studying how information affects homophily, Gallen and Wasserman (2023), finds that women portray homophile tendencies more than men in an online college mentoring platform. Also, Jackson et al. (2022) track university students' friendships and study partnerships in their

Caltech Cohort study and find assortative homophily by gender and ethnicity exists and persists substantially over time among friendship and study groups.

5.2.2 Differential Effects by Ethnic Salience

Having a coethnic in a peer group might be useful for students with high ethnic salience due to migrating from their home regions to attend university and experiencing a "diversity shock" when they arrive at the campus. Migrating from one's ethnic region to attend a university located in a different ethnic region with different cultures could cause immigrant students to be aware of their own ethnic identity, leading to greater attachment to their own ethnicities. This is the phenomenon in Okunogbe (2018), who finds greater ethnic pride among Nigerian youth randomly assigned to serve in a region where the ethnic majority is different from their own ethnicity. These hypotheses also align with the psychology literature on social identity, which suggests that the salience of one's ethnic identity increases when one migrates away from one's native region.

In addition, such students could face social isolation as they encounter cultural barriers, which may increase their stress levels and contribute to a lack of sense of belonging. Moreover, students from certain ethnicities may experience discrimination from other groups, leading them to isolate themselves.[7] These students are forced to navigate a new learning environment where classrooms are more diverse than their high schools. Yet, several studies report generally low trust levels in addition to high in-group bias in highly ethnically diverse societies. Having a high proportion of coethnic peers in their peer group can be beneficial for students with high ethnic salience, especially if they belong to small ethnic groups.

5.2.3 Differential Effects by Degree Type

Studies on post-secondary education have reported differences by subject type. For example, Carrell et al. (2009) find peer effects are stronger in math and science courses, smaller in social sciences, and absent in foreign languages and physical education at the Air Force Academy. Studying peer effects from the field of study at an Italian university, Brunello et al. (2010) find that peer effects are stronger in the 'hard' sciences (engineering, math, and natural sciences) but absent in social sciences and humanities. I do not observe course names, but I observe the course code (e.g., STA101) and the degree type. MUK offers degrees in either arts or sciences. Arts degree comprises a wide range of degrees, such as business-related, social sciences, and humanities, and so do science degrees.

There are other reasons in this context to anticipate why peer effects might differ by degree type. For example, classrooms in arts degrees may differ from those in science classes, as they are, on average, larger. Additionally, the proportion of high-ability peers in arts degrees is smaller due to the design of the national merit scholarship scheme. Generally, larger class sizes would reduce interaction with the professor by increasing the student-to-teacher ratio. Since student-teacher interactions outside the classroom are limited in this setting, class size effects are more likely to manifest through peer effects. Also, large classrooms might increase the need

[7] Another potential reason for the isolation of certain ethnicities is inter-ethnic conflicts and competition spilling into classrooms, although ethnic conflicts are not common in this setting.

Table 3: Mean Effects in Year One: Coethnic vs High-ability Share

	(1)	(2)	(3)	(4)
Coethnic share	1.054**	1.064**	1.046**	0.936**
	(0.47)	(0.47)	(0.47)	(0.47)
High-ability share	0.799***	0.833***	0.848***	0.735**
	(0.29)	(0.29)	(0.29)	(0.28)
High-ability	3.790***	3.792***	3.791***	3.787***
	(0.09)	(0.09)	(0.09)	(0.09)
R-squared	0.326	0.326	0.328	0.328
N	321,452	321,452	321,452	321,452
Dorm FE	No	Yes	Yes	Yes
Individual Controls	No	No	Yes	Yes
Group Controls	No	No	No	Yes

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school assigned to the same dorm. Each column is an independent regression, but the outcome is the course grades in all the specifications. The differences between each specification are indicated at the bottom and come from the controls. All regressions control for own ethnicity, gender, own ability, major FE, and HS subject combination FE, but gender drops out (2)-(4) since dorms are single-sex. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level.
*p<0.1, **p<0.5, ***p<0.01

for a sense of belonging and may reduce intimate cross-cultural interactions among students. It is easier to sort based on ethnicity as the probability of the coethnic presence of ethnicity is high.

5.2.4 Differential Effects by Ability

Heterogeneous peer effects by a student's own ability and peers' average ability have been shown to exist in the literature. For example, Zimmerman (2003) finds students in the middle of the verbal SAT distribution have negative peer effects from low-ability roommates. Also, Carrell et al. (2009) find suggestive evidence of non-linearity peer effects. Verbal SAT peer effects are strong for students in the bottom third of the distribution. Given reduced student-teacher interaction in this setting, peer effects may exist through study partnership channels, especially for low-ability students.

6 Results

6.1 Mean Effects

Table 3 estimates various specifications of equation (4). All specifications control for ethnicity as described in section 5, ability, gender, student's major, and classroom and HS subject combination fixed effects. The difference between specifications is shown at the bottom of each column. It comes from controls in each regression, as I begin with a simple regression and progressively add more controls. Since I do not know the residence status of the students, the coefficients reported in this should be interpreted as intent to treat effects.

Given no evidence of selection, as reported in Section 5, we do not expect the coefficients to change significantly as we move from column (1) to (4). The table shows that the share of coethnic and high-ability matters significantly for academic performance. The effect of coethnic share is stable at around one percentage point (pp) , while that of high-ability peers is around 0.8pp. Adding dorm fixed effects and individual and group controls does not alter the effects.

The results in specification From specification (4) imply that adding five coethnic peers to a typical peer group of size 25 (corresponding to the average group size) increases a student's performance by 0.19pp ($5/25 \times 0.936$). This effect is equivalent to 0.02 standard deviations in a student's performance in the first year. The effect of high-ability share is 0.735, which implies that adding two more high-ability peers to a typical group of size 25 increases a student's performance by 0.15pp ($5/25 \times .735$), which is also about 0.02 standard deviations change in a student's performance.

For context, adding five coethnic peers to a typical peer group of size 25 (corresponding average size as Table 1 shows) is equivalent to moving from the group where the number of coethnic peers corresponds to the twenty-fifth percentile to a group where the number of coethnic peers corresponds to the seventy-fifth percentile.[8] For simplicity, I will interpret the results as the effect of adding either five coethnic or high-ability peers to a group of 25.

Table 3 also shows the effect of own ability is much larger than the effect of coethnic and high-ability share. The table shows that high-ability students perform about four percentage points higher than low-ability peers. This difference is large as it corresponds to a change in grade that would move a student whose first-year grade is equal to the average from the second-class lower (Fairly Good) performance range to a second-upper (Good) range. The average of grades reported in Table 1 is equivalent to second class lower degree type in this setting.

Results show positive and direct peer effects from a higher share of high-ability (irrespective of ethnicity) and coethnic peers. However, it is likely that the high ability of coethnic peers might matter, while high-ability noncoethnic peers do not. To test this, I break down the treatment variables into four: (A) high-ability coethnic peers, (B) low-ability coethnic peers, (C) high-ability noncoethnic peers, and (D) low-ability noncoethnic peers and compute the share of each as described in Section 5. If high-ability coethnic peers matter while high-ability noncoethnic peers do not, (A) should be positive and significant while (C) should not, or at least (A) should be larger than (C), indicating that coethnic peers matter indirectly through high-ability peers.

These results in Table 4 largely follow the pattern observed in Table 3. When running these regressions, I exclude category (D). Therefore, the reported coefficients should be interpreted relative to that reference group. From the preferred specification (4), adding five high-ability coethnic or noncoethnic peers to a group of size 25 increases a student's course grade by 0.18pp relative to low-ability noncoethnic peers, which is equivalent to 0.02 standard deviations. The same table also shows that low-ability coethnic peers have a large and positive effect on academic

[8] It is important to note that distribution of high-ability peers may be different from that of coethnic peers in a group. I use a marginal change of 5 peers in a typical group size for simplicity of interpretation.

Table 4: The Effect of High-ability Coethnic and High-ability Noncoethnic peers on Academic Performance in Year One.

	(1)	(2)	(3)	(4)
(A) High-ability coethnic share	0.928*	0.980*	1.053**	0.865*
	(0.52)	(0.52)	(0.51)	(0.50)
(B) Low-ability coethnic share	0.686*	0.692*	0.696*	0.648
	(0.41)	(0.41)	(0.41)	(0.41)
(C) High-ability noncoethnic share	0.963***	0.992***	0.991***	0.875***
	(0.32)	(0.32)	(0.32)	(0.32)
R-squared	0.326	0.326	0.328	0.328
N	321,375	321,375	321,375	321,375
Dorm FE	No	Yes	Yes	Yes
Individual controls	No	No	Yes	Yes
Group controls	No	No	No	Yes

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school, and assigned to the same dorm. Each column is an independent regression, but the outcome is the course grades in all the specifications. The differences between each specification are indicated at the bottom and come from the controls. All regressions control for own ethnicity, gender, own ability, major FE, and HS subject combination FE, but gender drops out (2)-(4) since dorms are single-sex. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level.
* p<0.10, ** p<0.05, *** p<0.01

performance. However, it is imprecise.

Lastly, estimates in both Table 4 and Table 3 don't change or change very little when I add dorm FE, individual controls, and group controls. This is consistent with exogenous assignment into peer groups and that correlated shocks are less likely to drive results reported in this paper. The results do not suggest that high-ability coethnic peers matter more than high-ability noncoethnic peers in the first year, which is contrary to my prior. Taken together, these results show that, on average, high-ability peers directly and significantly impact every student's grades regardless of their ethnicity. Also, these results show that coethnic peers have a positive effect on grades, although they suggest high-ability coethnic matter more than low-ability coethnic peers.

6.2 Persistence of Mean Effects

All the results presented thus far focus on student performance during their first year at MUK. By extending the analysis to the subsequent years of their education, I test for the persistence of these peer effects. If peer effects from social networks persist as a student advances throughout their college career, then the effects of high-ability and coethnic peers observed in Section 6.1 should also be evident in the follow-on years. Since selection into courses is limited, this setting allows me to explore the persistence of peer effects. I estimate specification (5) separately for each of the three academic years of undergraduate education at MUK and present the results in Table 5. Panel A compares the effect of coethnic share to that of high-ability share, while Panel B compares high-ability coethnic to noncoethnic peers.

Column (1) restates the results in Section 6.1 to facilitate comparison. The effect of coethnic share persists into the second year, but it is almost half of the magnitude of the first year's effect in the third year. That is, the effect of adding five coethnic peers into a peer group of size 25 is 0.10pp in the third year, which is not statistically different from zero and is almost half of the effect observed in the first year, as reported in Table 3. In comparison, the effect of high-ability share persists and even increases in the third year. From Panel A, adding two high-ability peers to a group of 25 increases a student's performance by 0.22pp in the third year Yet, the same change increases student performance by 0.15pp in the first year. Thus, the effect of high-ability peers in the third year is 1.5 times the effect observed in the first year, and that of coethnic peers is one-half of what is observed in the first year.

Panel B shows results similar to those in Panel A. Relative to low-ability noncoethnic peers, the effect of high-ability coethnic and noncoethnic in the third year is positive and significant. In contrast, that of low-ability coethnic peers in the third year is not significant and is about 25% lower than the effect observed in the first year. Additionally, the effects of high-ability peers (coethnic and noncoethnic) relative to low-ability peers increase from the first to the third year.

The results in Table 5 show that the role of shared identity, if not coupled with ability, falls as time goes on. However, the effect of high-ability peers rises as time goes on, although the effect of high-ability coethnic peers increases more than that of high-ability noncoethnic peers. The results are suggestive of evolving study groups or social networks.[9] For example, students might form stronger study bonds with high-ability peers as time goes on.

6.3 Heterogeneous Peer Effects

Results in Section 6.1 show that, on average, going to school and potentially living with coethnic and high ability affects academic performance, but the effect of the coethnic share falls as time goes on. I now turn to see whether there are differential impacts on different dimensions mentioned in Section 5.2, as such differential effects might shed light on some mechanisms.

6.3.1 Differential Impacts by Gender

Table 6 presents the differential effects by gender. As dorms are single-sex, I control for gender instead of dorm FE. Controlling for dorm FE does not change the results, but gender drops out. The table also represents results across the years and p-values corresponding to testing the significance of the treatment variables for girls: coethnic share ($\phi_1 + \varphi_1$) and high-ability share ($\phi_2 + \varphi_2$) in equation (6). These results reveal several patterns.

First, girls perform lower than boys by 0.743pp, significant in the first year, but they perform

[9] I will explore this mechanism when I run surveys later. It is possible that students hang out with coethnic peers at the start of their university career because it is more organic. However, as time goes on, networks may evolve as they learn which of their peers are high-ability (coethnic or noncoethnic). They might form stronger networks with high-ability coethnic peers, weaker networks with low-ability coethnic peers, and somewhat strong networks with high-ability noncoethnic peers, as high-ability peers may be perceived as more beneficial for academic performance. The survey will ask students about their study and friendship groups throughout their undergrad career to see if they are constant or changing overtime.

Table 5: Persistence of Mean Effects in Follow-up Years

	Year One	Year Two	Year Three
Panel A: Coethnic vs High-ability			
Coethnic share	0.936**	1.136**	0.491
	(0.47)	(0.48)	(0.45)
High-ability share	0.735**	0.839***	1.101***
	(0.28)	(0.28)	(0.28)
R-squared	0.328	0.384	0.380
N	321,375	343,761	330,158
Panel B: High-ability (Coethnic vs Noncoethnic)			
High-ability coethnic share	0.867*	1.201**	1.347***
	(0.50)	(0.52)	(0.50)
Low-ability coethnic share	0.639	0.775*	0.476
	(0.41)	(0.41)	(0.39)
High-ability noncoethnic share	0.879***	0.951***	1.166***
	(0.32)	(0.32)	(0.32)
R-squared	0.326	0.383	0.378
N	310,867	333,187	320,752
Dorm FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Group controls	Yes	Yes	Yes

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school and assigned to the same dorm. Each column is an independent regression, but the outcome is course grades in all regressions. All regressions control for own ethnicity, own ability and major, HS subject combination, and classroom FE. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level.
* p<0.10, ** p<0.05, *** p<0.01

better than boys by 1.061pp in the third year. The difference in performance between boys and girls is not significant in the second year.

Second, the coethnic share is positive but not significant for both boys in the first and third years, but it is marginally significant in the second year. Although economically meaningful, the differential impact by gender is not significant across years. These results suggest that, unlike boys, girls might be benefiting from a higher coethnic share. The p-value testing the sum of the significance of coethnic share and the interaction of the coethnic share and female dummy is largely significant in the first and second years and marginally significant in the third year. These results imply that adding five coethnic peers in a group of 25 increases academic performance for boys by 0.13pp, which is equivalent to a 0.01 standard deviation change in academic performance in the first year. On the contrary, the same in coethnic peers would increase girl's performance by 0.25PP, equivalent to 0.03 standard deviation in the first year. Thus, the effect of coethnic share on boys is about 30% lower than the average effect in the first year, yet the effect on girls is about 30% larger than the average effect observed in Table 3.

Table 6: Differential Effect by Gender: Coethnic vs High-ability Share

	Year One	Year Two	Year Three
Coethnic share	0.630	0.928*	0.136
	(0.55)	(0.56)	(0.53)
High-ability share	0.661**	0.821***	1.101***
	(0.31)	(0.31)	(0.31)
Female	-0.743***	0.243	1.061***
	(0.21)	(0.21)	(0.21)
Coethnic share × Female	0.605	0.398	0.668
	(0.48)	(0.49)	(0.46)
High-ability share × Female	0.136	-0.021	-0.074
	(0.42)	(0.42)	(0.41)
p-val Coethnic share: Female	0.015	0.001	0.097
p-val High-ability share: Female	0.059	0.053	0.013
R-squared	0.328	0.384	0.380
N	321,375	343,761	330,158
Dorm FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Group controls	Yes	Yes	Yes

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school and assigned to the same dorm. Each column is an independent regression, but the outcome is course grades in all regressions. All regressions control for own ethnicity, major, HS subject combination, and classroom FE. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level. The table also reports p-values for coethnic share and high-ability share of female students. These tests correspond to $\phi_1 + \varphi_1$ and $\phi_2 + \varphi_2$ in equation (6).
* p<0.10, ** p<0.05, *** p<0.01

Third, the differential effect of high ability by gender is small, insignificant, and sometimes negative. Still, the share of high-ability peers has a positive and significant effect on boys and girls. As the average effect reported in Table 3, the share of high-ability persists and increases into the third year for both boys and girls.

6.3.2 Differential Impacts by Ability

Table 7 shows the differential effects by own ability across the years. The table also reports results across the years and p-values corresponding to testing the significance of the treatment variables for high-ability: coethnic share ($\phi_1 + \varphi_1$) and high-ability share ($\phi_2 + \varphi_2$) in equation (6). High-ability students perform higher than low-ability peers, as Section 6.1 already reported. The results reveal several other patterns.

First, the effect of coethnic share on the academic performance of low-ability students is not significant across all the years and is negative in the third year. On the other hand, the effect of

Table 7: Differential Effect by Ability Type: Coethnic vs High-ability Share

	Year One	Year Two	Year Three
Coethnic share	0.064	0.274	-0.126
	(0.48)	(0.48)	(0.46)
High-ability share	0.716**	0.911***	1.234***
	(0.31)	(0.32)	(0.30)
High-ability	3.259***	3.013***	2.816***
	(0.20)	(0.20)	(0.19)
Coethnic share × High-ability	2.467***	2.513***	1.846***
	(0.51)	(0.51)	(0.48)
High-ability share × High-ability	-0.015	-0.269	-0.429
	(0.42)	(0.44)	(0.41)
p-val Coethnic share: High-ability	0.000	0.000	0.003
p-val High-ability share: High-ability	0.082	0.115	0.046
R-squared	0.328	0.384	0.380
N	321,452	343,840	330,236
Dorm FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Group controls	Yes	Yes	Yes

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school and assigned to the same dorm. Each column is an independent regression, but the outcome is course grades in all regressions. All regressions control for own ethnicity, own ability, and major, HS subject combination, and classroom FE in addition to individual and group controls. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level. The table also reports p-values for coethnic share and high-ability share of high-ability. These tests correspond to $\phi_1 + \varphi_1$ and $\phi_2 + \varphi_2$ in equation (6).
* p<0.10, ** p<0.05, *** p<0.01

high-ability share on low-ability students is positive and significant across all the years and even higher in the third year than in the first year. For example, adding five high-ability peers to a group of 25 increases a slow student's performance by 0.14pp and 0.25pp in the first and third year, respectively. Although the differential impact by ability is not significant across all the years, it is negative and economically meaningful in the third year, implying that high-ability peers have a larger effect on low-ability students than high-ability students.

Second, the differential impact of coethnic share is large and significant in Year One and largely persists into Year Three. Table 7 shows adding five coethnic peers to a group of 25 increases the performance of high-ability students by 0.51pp, which is about 3.5 times the average effect reported in Table 3. This effect is about 0.05 standard deviation of the first year's performance. This differential impact of coethnic share on high-ability students persists significantly into the third year, albeit at a reduced magnitude.

6.3.3 Differential Impacts by Degree Type

I estimate the differential effect by degree type and report it in Table 8. Since I control for the Major FE, the arts major dummy drops out of the regressions. Although not shown, the results change significantly when I control for the degree type dummy instead of major FE changes. The table also reports results across the years and p-values corresponding to testing the significance of the treatment variables for arts majors: coethnic share ($\phi_1 + \varphi_1$) and high-ability share ($\phi_2 + \varphi_2$) in equation (6).

This table shows the differential impact of high-ability share by degree type is large and significant across all the years and more than doubles from the first to third year. The results show that adding five high-ability peers to a group of 25 increases a student in the art's major performance by 0.33pp in the first year, which is almost 1.5 times the average effect reported in Table 3. Moreover, this effect increases to 0.362pp in the third year, which is 3.3 times the effect reported 5. The effect of high-ability share on a student who is a degree major in the third year is very large, as it corresponds to 0.08 standard of academic year in the third year.

Lastly, the table also shows the differential impact of coethnic share by degree type is significant in the first year but not in the second and third years. Adding five coethnic peers to a group of 25 increases the academic performance of a student in the arts degree by 0.22pp more than for a student in the science majors.

6.3.4 Differential Impacts by Ethnic Salience

I proxy high 'ethnic salience' using a dummy variable equal to one if a student graduated high school from a district of origin and zero otherwise.[10] Since most Ugandan districts are ethnically segregated, these students have generally had much less exposure to other ethnicities prior to enrolling at MUK than their peers of the same ethnicity who graduated high school outside their district of origin (e.g., as a boarding student in or near Kampala). I estimate the differential impacts by ethnic salience and present these results in Table 9. The table also presents these results across the years, which show interesting patterns and p-values corresponding to testing the significance of the treatment variables for arts majors: coethnic share ($\phi_1 + \varphi_1$) and high-ability share ($\phi_2 + \varphi_2$) in equation (6).

First, students with assumed high ethnic salience perform significantly lower than those with low assumed ethnic salience. However, this negative difference reduces over time and is no longer significant in the third year. That is, students of assumed high ethnic salience perform 0.63pp, significant at 1% lower than those of assumed low-ethnic salience in the first year, but the coefficient of this dummy increases to -0.17 and is no longer significant in the third year.

Second, the effect of coethnic share on students with low assumed ethnic salience is not significant across the years. However, Table 9 shows that students of high ethnic salience type benefit from a high share of coethnic peers in the first and second year. The differential effects of coethnic in the first, second, and third years are 2.088pp (significant), 1.323pp (significant), and 0.515 (insignificant), respectively. The table also reports the p-values of treatment variables at the bottom, which show that coethnic peers are important for students with high ethnic salience

[10] I treat Kampala Metropolitan area, which includes Kampala and Wakiso as the one district as these two share the cities and there are clear boarders between these.

Table 8: Differential Effect by Degree Type: Coethnic vs High-ability Share

	Year One	Year Two	Year Three
Coethnic share	0.586	1.101**	0.650
	(0.51)	(0.52)	(0.49)
High-ability share	0.455	0.384	0.034
	(0.36)	(0.37)	(0.36)
Coethnic share × Arts degree	1.080**	0.188	-0.074
	(0.49)	(0.49)	(0.46)
High-ability share × Arts degree	0.975*	1.321**	2.979***
	(0.59)	(0.60)	(0.57)
p-val Coethnic share: Arts degree	0.006	0.017	0.283
p-val High-ability share: Arts degree	0.000	0.000	0.000
R-squared	0.328	0.384	0.380
N	321,375	343,761	330,158
Dorm FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Group controls	Yes	Yes	Yes

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school and assigned to the same dorm. Each column is an independent regression, but the outcome is course grades in all regressions. All regressions control for own ethnicity, ability, major and HS subject combination, and classroom FE. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level. The table also reports p-values for coethnic share and high-ability share of arts degree. These tests correspond to $\phi_1 + \varphi_1$ and $\phi_2 + \varphi_2$ in equation (6).
* p<0.10, ** p<0.05, *** p<0.01

in the first year and second year. These results imply that adding five coethnic peers to a group of 25 increases academic performance by 0.42pp more for students assumed to be of high-ethnic salience type than those assumed to be of low ethnic salience in the first year. That is, adding five coethnic peers to a group of 25 leaders leads to a 0.44pp increase in academic performance, which is equivalent to a 0.05 standard deviation in the first year and is 2.5 times the average effect reported in Table 3.

Third, although positive, the differential effect of high-ability share by ethnic salience is small and insignificant. The share of high-ability peers has a positive and significant effect on students assumed to be of low ethnic and high ethnic salience in the first year, which persists and increases for both types of students in the third year.

These results indicate that students who might suffer from cultural and diversity shock when they arrive at MUK to study benefit more from coethnic peers than high-ability peers. Nevertheless, the effect of coethnic share decreases from the first to the second and disappears by the time the student graduates.

Table 9: Differential Effect by Ethnic Salience: Coethnic vs High-ability Share

	Year One	Year Two	Year Three
Coethnic share	0.132	0.613	0.285
	(0.51)	(0.52)	(0.49)
High-ability share	0.716**	0.737**	1.044***
	(0.30)	(0.31)	(0.31)
High ethnic salience	-0.626***	-0.463***	-0.173
	(0.17)	(0.17)	(0.17)
Coethnics Share × High ethnic salience	2.088***	1.323***	0.515
	(0.48)	(0.48)	(0.46)
High-abillity Share × High ethnic salience	0.070	0.297	0.162
	(0.30)	(0.33)	(0.31)
p-val Coethnic share: High ethnic salience	0.000	0.000	0.130
p-val High-ability share: High ethnic salience	0.024	0.002	0.000
R-squared	0.328	0.384	0.380
N	321,375	343,761	330,158
Dorm FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Group controls	Yes	Yes	Yes

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school and assigned to the same dorm. Each column is an independent regression, but the outcome is course grades in all regressions. All regressions control for own ethnicity, ability, and major, HS subject combination, and classroom FE. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level. The table also reports p-values for coethnic share and high-ability share of students assumed to be of high-ethnic salience. These tests correspond to $\phi_1 + \varphi_1$ and $\phi_2 + \varphi_2$ in equation (6).
* p<0.10, ** p<0.05, *** p<0.01

6.4 Robustness checks

One area of concern revolves around the potential impact of measurement error on estimates of the coethnic share in a peer group. As aforementioned, I use the probable coethnic share in a peer group to account for this. Nevertheless, I get practically similar results when I re-estimate the results using a single ethnicity corresponding to the category the algorithm is most confident about. I discuss robustness in relation to using student-level aggregated data and, thus, a different set of fixed effects in this section.

6.4.1 GPA as the Dependent Variable

I re-estimate the average effects at the student level, using GPA as the outcome (not course-level grades), and report these results in Table 10. Panel A compares the ethnic and high-ability shares within a student's peer group. In contrast, Panel B compares the effect of higher-ability coethnic peers to high-ability noncoethnic peers relative to low-ability noncoethnics. In addition, (1) is the same as the main effects regressions reported in Section 6.1 and is included for comparison purposes. The results obtained using GPA as the outcome are similar to those reported in Section 6.1. Naturally, the magnitudes of coefficients are different since the outcome variables are different. The results are consistent when I use school-by-year in place of classroom FE.

From Panel A, the effect of high-ability and coethnic share is positive and significant when the outcome is GPA Similarly, from Panel B, high-ability coethnic and noncoethnic peers positively and significantly affect academic performance. However, panel B shows the effect of low-ability coethnic is precisely estimated when I use GPA. As in Table 4 of the main effects, using GPA as an outcome also suggests that high-ability coethnic matter as much as high-ability noncoethnic peers although. Although the effect of high-ability coethnic peers is larger than that of high-ability noncoethnic peers in panel B column (2), the difference of 0.02 is not significantly different from zero.

6.5 Discussion and Contextualizing Results

I find that the share of high-ability and coethnic peers positively and directly affects academic performance, although the effect of the latter does not persist. The results reported in Table 3 suggest a mean peer effect size of 0.02 SD for both peer types. These are reduced-form effects based on dorm assignment, which is likely to be an underestimate of the true peer effect (treatment effect on the treated). This effect is comparable to what zimmer2003 finds at Williams College as Figure 5 shows. The effect Garlick (2018) finds at the University of Cape Town (UCT) using randomly assigned residential peers assignment is larger than what find, although his reported confidence intervals are large. The estimate is Garlick (2018) also reduced form effect because the author observes dorm assignments but not roommates. However, compliance is high in Garlick (2018) and is probably characterized by students who enroll at UCT from out of the city.[11]

Interestingly, I find strong coethnic reduced-form peer effects—equal to 0.05 standard deviations, especially for students of assumed high ethnic salience, which is comparable to the average effect in Carrell et al. (2009) at the Air Force Academy. In short, in this setting with high ethnic diversity, I still find both ability and coethnic peer effects where high ethnic diversity is expected to dampen peer effects of higher ability. Ethnic diversity effects are unlikely to play a role at MUK.

While channels behind peer effects literature, in general, are unclear, I hypothesize on the mechanisms behind these results by discussing explanations for these results in this section based on the mean and heterogeneous effects reported in the results section and the characteristics of this context.[12] The suggestive channels at play in this context that I discuss in this section include peer-to-peer learning, friendships, and psychological and cultural reasons.

[11]The author mentions that people who do not live on campus in private residences, most likely with families (page 348).

[12]I do not test mechanisms directly. I am yet to start collecting primary data through surveys, for which I have already obtained IRB approval, including local IRB.

Table 10: Coethnic vs High-ability Share: Outcome as GPA

	Outcome variable: % course grades	Outcome variable: GPA
Panel A: Coethnic vs High-ability		
Coethnic share	0.936**	0.112**
	(0.47)	(0.05)
High-ability share	0.735***	0.066**
	(0.28)	(0.03)
R-squared	0.328	0.277
N	321,452	25,298
Panel B: High-ability (Coethnic vs Noncoethnic)		
High-ability coethnic share	0.867***	0.091*
	(0.51)	(0.05)
Low-ability coethnic share	0.639	0.076*
	(0.39)	(0.04)
High-ability Noncoethnic Share	0.879***	0.081**
	(0.32)	(0.03)
R-squared	0.328	0.277
N	321,375	25,298
Classroom FE	Yes	No
School-by-year FE	No	Yes
Dorm FE	Yes	Yes
Individual controls	Yes	Yes
Group controls	Yes	Yes

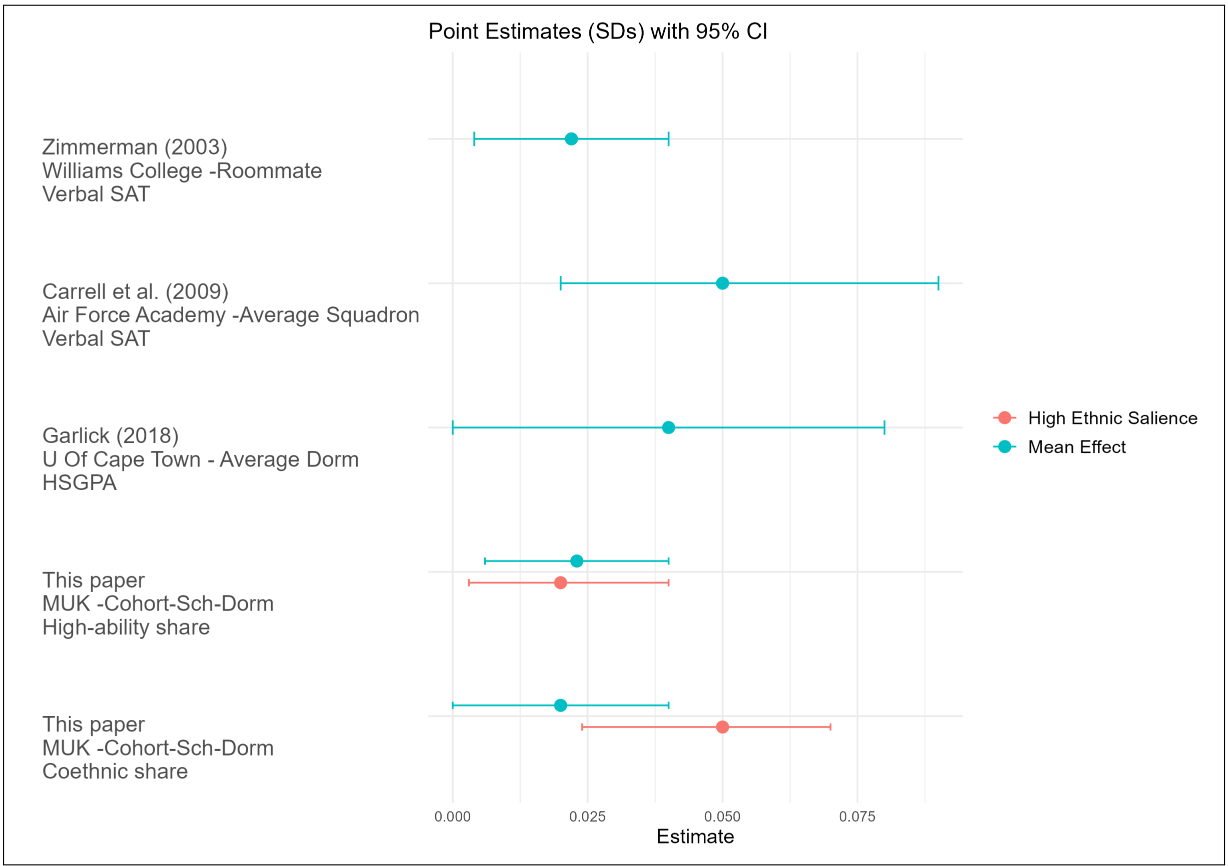
Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school and assigned to the same dorm. Each column is an independent regression, but the outcome is course grades in all regressions. All regressions control for own ethnicity and dorm, major, and HS subject combination FE in addition to individual and group controls. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

6.5.1 Peer-to-Peer Learning and Study Behavior

Table 3 shows that one's own ability positively and significantly affects academic performance, an effect that persists into the final year of most majors. High-ability peers may influence the academic performance of their peers by facilitating peer-to-peer learning, such as leading discussion groups. This is especially important as office hours (professors or TAs) do not exist and because of the classical style of lectures in this setting. Students in need of extra help might rely on high-ability peers for additional assistance.

Students can identify their high-ability classmates through several methods, especially during the academic year advances. Firstly, student registration numbers differ by the enrollment scheme, such as merit scholarship status (high-ability). Secondly, it is common for newspapers to publish the names of the top students in the country (those with a high chance of

Figure 5: Comparison to past Papers



This figure compares my average estimate and estimate of students of assumed high ethnic salience to past papers with randomly assigned peers and significant average effects. Carrell et al. (2009) Table 3 column (6) page 452 reports a coefficient of 0.382 on peer SAT verbal, equivalent to a 0.05 increase in GPA. Additionally, Carrell et al. (2009) estimate is about 2.5 times that reported in Zimmerman (2003) Table 3 column ("First semester") page 17. Garlick (2018) Table 4 column (1) reports a coefficient of 0.216 on the dormitory mean high school GPA, equivalent to a 0.04 SD. Lastly, the effects of coethnic and high-ability share in Table 3 are about 0.02 SD increase in academic performance. Additionally, I report the peer effects on students of high ethnic salience type, showing that coethnic peer effects are about 2.5 times the average effect for these types of students.

obtaining a merit scholarship) once the national exams are out. However, this usually occurs several months before students enroll at university, and newspapers are not delivered outside the largest cities. Lastly, it is typical for students' course grades, especially in midterm marks, to be publicly posted on department noticeboards. Consequently, it becomes easy to identify and seek assistance from high-ability peers in ways that can impact academic outcomes.

Another potential explanation through which high-ability peers can influence others is by affecting study efforts. Several studies utilizing time-use data have examined how a student's study behavior is influenced by the study behaviors of their peers (e.g., Mehta et al., 2019; Frijters et al., 2019). For instance, Mehta et al. (2019) show that students exhibit studious behavior if their peers, assigned randomly or connected through organic friendships, invest a lot of time studying at college or did during high school. It is conceivable that high-ability peers who have earned merit scholarships might have achieved it due to investing a significant amount of time into studying during high school or are doing so while at MUK. This intensified study behavior among high-ability peers could have a positive impact on the study behaviors of their peers.

6.5.2 Coethnic Friendships

Incoming freshmen can easily identify coethnic peers through physical features and cultural characteristics, including names and language. Shared ethnicity friendships are likely more organic due to shared identity since literature shows that coethnic bias exists in ethnically diverse societies. For example, Salmon-Letelier (2022) finds that ethnicity is important during friendship formation in Nigeria's state schools. Even studies outside the SSA report homophilous assortativity in student interactions in study groups and friendships based on gender and ethnicity (Jackson et al., 2022). Therefore, students might form ethnic-based friendships within randomly assigned groups explaining the suggestive evidence on why high-ability coethnic students might matter more for academic success as Table 4. This aligns with the Berea college freshman time-use (Mehta et al., 2019), which finds that using friends as peers is a stronger predictor of a student's propensity to study.

Nevertheless, the same table reports that high-ability noncoethnic peers also positively and significantly affect academic performance. Students may likely seek high-ability peers for academic help, irrespective of ethnicity. Thus, having high-ability coethnic peers is an added advantage because students may sort into friendships or study groups based on ethnicity when unaware of which of their peers are high-ability.

6.5.3 Cultural and Psychological Reasons

Lastly, these results also suggest cultural and psychological explanations at play. Many students migrating from rural districts might feel isolated as they navigate a diverse environment as they no longer belong to an ethnic majority. This could hamstring a sense of belonging for such students, which could have a negative effect on academic performance. As Table 9 shows, students of high ethnic salience perform lower than those of low ethnic salience in the first year.

In this case, a higher share of coethnic peers might be perceived equally or even more important compared to the share of high-ability peers by students experiencing a diversity shock. This mechanism might explain why I find coethnic peer effects in Table 3 are positive and significant, and even stronger in column one of Table 9 where the interaction of coethnic share and graduating HS from the district of origin is positive and significant.

If this mechanism is at play, this interaction should be even stronger for small groups (excluding Banyankore/Kiga and Baganda groups, which are the largest two groups that makeup 65% of the student population) as the smallest groups tend even to be more segregated as Figure 1 shows. Appendix Table A3 shows that the interaction is larger for smaller ethnic groups.

Nevertheless, this interaction could capture the influence of cultural shocks stemming from differences in diversity in the learning environment and between life in the city and rural areas. Beyond navigating diverse classrooms, migrated students encounter an urban lifestyle distinct from their rural upbringing. Additionally, it's conceivable that students who graduate from a high school within their district of origin might have predominantly resided at home, even though most Ugandan high schools offer boarding facilities. Such students might struggle to build a support network with peers, especially noncoethnic ones.

Table 9 also shows that the interaction's magnitude reduces from Year One to Year Three, and so does the mean effect of the coethnic share in Table 3. This pattern in the coefficients indicates that the importance of coethnic peers to students of assumed high ethnic salience goes away by the time they graduate. It is possible that cross-ethnic friendships emerge as these types of students acquaint themselves with peers through frequent interactions, making the coethnic share less important. The contact hypothesis, first introduced by William (1947), can explain this phenomenon.

Also, as students learn more about peers, ethnicity-based networks become less important compared to assortative matching based on attributes such as ability and study behaviors that matter more for academic success at college. This evolution of networks and information gain might explain why the effect of ability share increases with time.

7 Conclusion

Ethnic diversity has widespread and measurable impacts on a host of social, political and economic outcomes. In Sub-Saharan Africa, latent ethnic tension can deteriorate social trust and reinforce high coethnic favoritism. In the context of higher education, which brings students into close contact with ethnic diversity – often for the first time – ethnic heterogeneity may hamstring student collaboration and undermine academic performance with long-run implications. This paper provides causal estimates of peer effects on performance in the unique setting of higher education in Uganda, one of the region's most ethnically diverse and segregated countries.

I define a student's peer group as students admitted to majors within the same school in the same year who are assigned to the same dorm. This allows me to study the effects of peers with whom a student is likely to interact during school and non-school activities. Dorm assignments are random conditional on gender after a student is admitted, and courses are pre-determined at the time of admission before a student enrolls, providing an exogenous variation across peer groups. I find that coethnic peers (irrespective of ability) and high-ability peers (irrespective of ethnicity) have a positive and significant effect on grades in the first year. However, the mean effect of coethnic peers does not persist until a student graduates.

These mean results mask significant heterogeneity in coethnic peer effects. First, I find strong and positive coethnic peer effects for students of high ethnic salience that do not persist until a student graduates. These are students who graduated from secondary schools in their districts of birth and have relatively limited exposure to ethnicities different from their own prior to arrival at campus. I also find a strong positive coethnic peer effect for high-ability students, not low-ability students, that persists. This suggests that the benefits of coethnic peers can be reaped by those who have the capacity to succeed academically. The results also suggest coethnic peers have a larger positive impact on girls than boys.

These results have a number of implications for higher education policy and administration in Uganda and, perhaps, in comparable settings with high ethnic diversity. First, the positive impact of high-ability peers on academic performance underscores the importance of fostering an environment that encourages peer-to-peer learning. For example, universities could implement optimal peer group assignments where low-ability students are mentored by high-ability students. Second, the positive effect of coethnic peers in the initial years on students assumed to be of high ethnic salience suggests that there could be benefit of implementing programs that facilitate cross-cultural awareness, shared cultural events, and increase a sense of belonging.

Given the existence of ethnic student organizations in this setting, which suggests a degree of homophily that shapes student life, it is natural for incoming students of high ethnic salience to benefit from coethnic connections and support.

These results also suggest there might be a short-term cost to ethnic integration policies. For example, if a university peer group assignment algorithm breaks any homophily on ethnicity and enforces cross-ethnic mixing, it might have a negative effect on students who benefit from a higher share of coethnic peers, especially those assumed to be of high ethnic salience.

This paper points to several promising questions for future research. I find that a higher share of high-ability coethnic and noncoethnic peers increases a student's academic performance. At first glance, these findings suggest that college students at MUK may portray less coethnic bias during classroom interactions, such as study group formations that have an effect on economic outcomes. In such cases, the peer effects in this setting work through channels, such as study effort, as some studies using colleges in the West (e.g., Stinebrickner and Stinebrickner, 2006) report. However, these findings do not preclude other channels, such as coethnic cooperation and inter-ethnic competition. For example, high-ability coethnic peers might affect academic performance through cooperation with peers of shared ethnicity, while noncoethnic high-ability might increase competition where students of different ethnicities compete to the extent that increases academic performance.

Additionally, I study the first order of ethnic diversity on academic performance by focusing on coethnicity within a peer group. This paper does not study higher-order effects, such as the ethnic composition of noncoethnic peers, which is open for future research. For example, there might be an optimal pairing, tripling, quadrupling, etc., of ethnicities that could be beneficial or detrimental to academic performance. This kind of question requires going beyond studying the effect of ethnic diversity that would regress a Herfindahl index computed from ethnic shares within a student peer group on academic performance.

Lastly, this paper investigates short-term high-ability and coethnic peer effects by focusing on academic outcomes and finds that high-ability peers (irrespective of ethnicity) affect academic performance. However, it is unclear if a similar pattern of findings exists in the long term. Students may strategically engage during classroom interactions in a way that does not extend beyond the classroom. For instance, students might strategically select into study groups with higher-ability peers irrespective of ethnicity when doing homework but select into coethnic friend groups when forming non-education social networks. Cross-ethnic mixing at university may not change interethnic attitudes or social networks post-graduation if this happens. I focus these questions on the additional work I have initiated using the same setting of this paper.

References

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic growth*, 8:155–194.

Alesina, A. and La Ferrara, E. (2000). Participation in Heterogeneous Communities*. *The Quarterly Journal of Economics*, 115(3):847–904.

Alesina, A. and Zhuravskaya, E. (2011). Segregation and the quality of government in a cross section of countries. *American Economic Review*, 101(5):1872–1911.

Barone, G. and Mocetti, S. (2016). Intergenerational mobility in the very long run: Florence 1427-2011. Temi di discussione (Economic working papers) 1060, Bank of Italy, Economic Research and International Relations Area.

Bayer, A., Bhanot, S. P., Bronchetti, E. T., and O'Connell, S. A. (2020). Diagnosing the learning environment for diverse students in introductory economics: An analysis of relevance, belonging, and growth mindsets. *AEA Papers and Proceedings*, 110:294–98.

Bhusal, B., Callen, M., Gulzar, Saad;Pande, R., Prillaman, S. A., and Singhania, D. (2020). Does revolution work? evidence from nepal's people's war. *EGA Working Paper Series No. WPS-116. Center for Effective Global Action. Unversity Of California, Berkeley.*

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8(2):225–246.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.

Brunello, G., De Paola, M., and Scoppa, V. (2010). Peer effects in higher education: Does the field of study matter? *Economic Inquiry*, 48(3):621–634.

Carrell, S., Fullerton, R., and West, J. (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.

Carrell, S. E., Hoekstra, M., and Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review*, 108(11):3377–3415.

Carrell, S. E. and Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, 2(1):211–28.

Carrell, S. E., Malmstrom, F. V., and West, J. E. (2008). Peer effects in academic cheating. *Journal of Human Resources*, 43(1):173–207.

Cavnar, W. B. and Trenkle, J. M. (1994). N-gram-based text categorization.

Clark, G. and Cummins, N. (2015). Intergenerational wealth mobility in england, 1858–2012: Surnames and social mobility. *The Economic Journal*, 125(582):61–85.

Corno, L., Ferrara, E. L., and Burns, J. (2019). Interaction, stereotypes and performance. Evidence from South Africa. IFS Working Papers W19/03, Institute for Fiscal Studies.

De Giorgi, G., Pellizzari, M., and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2(2):241–75.

de la Cadena, M. (1995). *Women Are More Indian: Ethnicity and Gender in a Community near Cuzco*. Duke University Press.

Depetris-Chauvin, E. and Durante, R. (2017). One team, one nation: Football, ethnic identity and conflict in africa.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics*, 112(4):1203–1250.

Eifert, B., Miguel, E., and Posner, D. N. (2010). Political competition and ethnic identification in africa. *American Journal of Political Science*, 54(2):494–510.

Foster, G. (2006). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics*, 90(8):1455–1475.

Frijters, P., Islam, A., and Pakrashi, D. (2019). Heterogeneity in peer effects in random dormitory assignment in a developing country. *Journal of Economic Behavior & Organization*, 163:117–134.

Gallen, Y. and Wasserman, M. (2023). Does information affect homophily? *Journal of Public Economics*, 222:104876.

Garlick, R. (2018). Academic peer effects with different group assignment policies: Residential tracking versus random assignment. *American Economic Journal: Applied Economics*, 10(3):345–69.

Gisselquist, R. M., Leiderer, S., and Niño-Zarazúa, M. (2016). Ethnic heterogeneity and public goods provision in zambia: Evidence of a subnational "diversity dividend". *World Development*, 78:308–323.

Green, E. (2008). District creation and decentralisation in uganda.

Habyarimana, J., Humphreys, M., Posner, D. N., and Weinstein, J. M. (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review*, 101(4):709–725.

Hjort, J. (2014). Ethnic Divisions and Production in Firms. *The Quarterly Journal of Economics*, 129(4):1899–1946.

Hooghe, M. (2007). Social capital and diversity generalized trust, social cohesion and regimes of diversity. *Canadian Journal of Political Science / Revue canadienne de science politique*, 40(3):709–732.

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation*. *The Quarterly Journal of Economics*, 115(4):1239–1285.

Hoxby, C. M. (2002). The power of peers: How does the makeup of a classroom influence achievement? (research). *Education Next*, 2:57.

Hudson, M. C. and Taylor, C. L. (1972). *World handbook of political and social indicators*. Yale University Press New Haven, Conn.

Håkansson, P. and Sjöholm, F. (2007). Who do you trust? ethnicity and trust in bosnia and herzegovina. *Europe-Asia Studies*, 59(6):961–976.

Jackson, M. O., Nei, S. M., Snowberg, E., and Yariv, L. (2022). The dynamics of networks and homophily. Working Paper 30815, National Bureau of Economic Research.

Jones, K. S. (2004). A statistical interpretation of term specificity and its application in retrieval. *J. Documentation*, 60:493–502.

Luhn, H. P. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, 1(4):309–317.

Mamdani, M. (2001). *When Victims Become Killers: Colonialism, Nativism, and the Genocide in Rwanda*. Princeton University Press.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Mehta, N., Stinebrickner, R., and Stinebrickner, T. (2018). Time-use and academic peer effects in college. Working Paper 25168, National Bureau of Economic Research.

Mehta, N., Stinebrickner, R., and Stinebrickner, T. (2019). Time-use and academic peer effects in college. *Economic Inquiry*, 57(1):162–171.

Michuda, A. (2021). Rural-urban risk sharing networks: Uber driver labor supply responses to shocks in uganda. *SSRN*.

Miguel, E. (2004). Tribe or nation? nation building and public goods in kenya versus tanzania. *World Politics*, 56(3):327–362.

Monasterio, L. (2017). Surnames and ancestry in brazil. *PLoS ONE*, 12(5).

Morris, H. F. (1966). The uganda constitution, april 1966. *Journal of African Law*, 10(2):112–117.

NCHE (2018). The state of higher education and training in uganda 2018/19. A report on higher education delivery and institutions, The National Council for Higher Education.

Okunogbe, O. (2018). *Does Exposure to Other Ethnic Regions Promote National Integration?: Evidence from Nigeria*. The World Bank.

Ricart-Huguet, J. and Paluck, E. L. (2023). When the Sorting Hat Sorts Randomly: A Natural Experiment on Culture. *Quarterly Journal of Political Science*, 18(1):39–73.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates*. *The Quarterly Journal of Economics*, 116(2):681–704.

Sacerdote, B. (2011). Chapter 4 - peer effects in education: How might they work, how big are they and how much do we know thus far? volume 3 of *Handbook of the Economics of Education*, pages 249–277. Elsevier.

Salmon-Letelier, M. (2022). Friendship patterns in diverse nigerian unity schools. *Comparative Education Review*, 66(4):709–732.

Sen, M. and Wasow, O. (2016). Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics. *Annual Review of Political Science*, 19(1):499–522.

Ssentongo, J. S. (2016). 'the district belongs to the sons of the soil': Decentralisation and the entrenchment of ethnic exclusion in uganda. *Identity, Culture and Politics: An Afro-Asian Dialogue*, 17:60–96.

Stinebrickner, R. and Stinebrickner, T. R. (2006). What can be learned about peer effects using college roommates? evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, 90(8):1435–1454.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1):1–39.

Tornberg, H. (2013). Ethnic fragmentation and political instability in post-colonial uganda : understanding the contribution of colonial rule to the plights of the acholi people in northern uganda.

Turyahikayo, B. (1976). *Transafrican Journal of History*, 5(2):194–200.

UBOS (2006). 2002 uganda population and housing census. Analytical report, Uganda Bureau of Statistics.

Uganda Bureau of Statistics (2016). The national population and housing census 2014 – main report. Technical report, Kampala, Uganda.

Walton, G. M. and Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331(6023):1447–1451.

William, Robin M., J. (1947). *The Reduction of Intergroup Tensions: A Survey of Research on Problems of Ethnic, Racial, and Religious Group Relations.* Social Science Research Council, New York.

Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *The Review of Economics and Statistics*, 85(1):9–23.

Zárate, R. A. (2023). Uncovering peer effects in social and academic skills. *American Economic Journal: Applied Economics*, 15(3):35–79.

8 Appendix

8.1 Data Appendix

This section provides details that are not highlighted in the main data Section 4 of this paper.

8.1.1 Linking Student Data

MUK stores data on students' applications, admissions, and results in separate databases and offices. There is no unique identifier that can link databases in some cases.

STEP I: Computing GPA. My data cleaning process starts from the results database. These data list courses and course units (for some), and exam scores in percentages by program, department, semester, and year of study. They also list the calendar year when the exam was taken. These data cover 2008-2017. However, to match the admissions sample, I restrict the results sample to 2009-2017 years. I convert the exam scores from the percentage scale to letter grades using the information on the back of the transcripts and available in the code book. I then compute GPAs by semester and year.

STEP II: Merging with admissions. Each admitted student has two unique identifiers: student number and registration number. I use the latter to merge results and admissions data. There is a 93.9% merge rate at this stage.

Step III: Determining cohorts. The admissions and graduation programs are coded differently in many cases. The undergrad (graduation) program may admit students through different cohorts (e.g., evening and day classes). Take the graduation program "Bachelor of Science in Computer Science", for example; it is coded as "BCSCS". However, BCSCS students may be admitted to through two cohorts: day classes("CSC") or evening classes ("CSE"). This distinction was necessary because the cohort forms one's peer. I use the university codebook to ensure the admission, enrollment, and graduation programs are consistent. Since I restrict the sample to day majors, CSC appears in my final sample, while CSE does not.

Step IV: Merging with the name data. After correcting obvious misspellings in the names, I merged these data with data that predicted ethnicities. Merging on names in the training data gives a merge rate of 98.6%. Merging features produced by ML classification is irrelevant since ethnic predictions can be made for every surname.

Lastly, I deleted all the 2010 observations because the hall assignment is unavailable for many students admitted through a private scheme. The university officials in the admissions office mentioned that there was a problem/data bleach with the information system in 2010, where the university lost a lot of records.

8.2 Ethnic and Geographic Boundaries

The Ugandan parliament's gate has engravings of symbols and names of 15 administrative units at the time of independence from Britain. The administrative units were federal states, districts, or Territories (The Constitution of Uganda, 1964). The federal states were historical kingdoms, which included Ankole, Buganda, Bunyoro, and Toro, and the territory of Busoga. The districts included Acholi, Bugisu, Bukedi, Karamoja, Kigezi, Lango, Madi, Sebei, Teso, and West Nile.

Coincidentally, these kingdoms and districts' boundaries followed ethnic/tribe boundaries that existed before the British colonial government but were exacerbated by British colonists.

However, the colonial government introduced a notion of a district as an administrative unit, which initially was a way to group similar ethnicities in geographical proximity. Kingdoms were historically centralized and ethnically segregated, with a traditional king as a ruler. However, this was different for districts. Some districts, such as Sebei and Bugisu, were ethnically segregated but followed a different system of local political leadership, such as clans or chiefdoms. There were also districts (e.g., West Nile and Bukedi) that were a cluster of several, and sometimes unrelated, relatively small ethnic groups. For example, the West Nile comprised mostly Lugbara people but included smaller ethnic groups, such as the Alur and Kakwa.

President Obote abolished kingdoms in 1966 for political reasons and changed the status of federal states to districts, and split the formerly powerful federal state (kingdom) of Buganda into four districts (Morris, 1966).[13] Since then, the number of districts has increased to 135 over the years, with the highest increase happening under the current government for reasons such as service delivery and ethnolinguistic conflict management, among others. Some studies report political reasons as the most prominent explanations for new district creation (Green, 2008).

Most importantly, new districts are curved out of existing districts at the time of creation. It has been rare to create a district by carving out counties that initially belonged to two separate districts over the years. Interestingly, albeit unsurprising, new districts tend to be more segregated by ethnicity (Ssentongo, 2016). For example, the population of Nebbi is 96.2% of Alur ethnicity, although it was carved out of the West Nile district in 1974, which mainly comprised the Lugbara people. The creation of new districts sometimes begins with smaller ethnicities wanting to break away from the majority ethnicity in the original bigger district for reasons such as autonomy and bringing resources closer to them. But also, the government will offer a county a district status for political support.

I can trace current administrative units to historical kingdoms using publicly available data on administrative units from the Ugandan Ministry of Local Government. I complement the public data with data from the 2014 census from UBOS. The Census data contain the population breakdown by ethnicity for each district, confirming ethnicity within each district. That is, the census reports the number of each 66 ethnicities that reside in each district (i.e., 136 X 66 observations). I compute the proportion of each ethnicity in a district and rank these proportions from the highest to the lowest.

The top-ranked ethnicity informs the ethnic region that the district belongs to. The average proportion of the top ethnicity by population is 0.737 (the median is 0.813), indicating high ethnic segregation within each district. These UBOS data help me confirm the historic ethnic regions and give the final ethnic and geographic boundaries. I then create ethnic clusters by combining both the current and historical administrative units to give final ethnic geographical borders. Using just the 1962 districts and kingdoms that were created by the British colonial government would give wrong borders as the colonial sometimes bundled together ethnic groups

[13] District is the second-largest unit of administration after the federal government. The districts divide into counties. Counties divide into sub-counties. Sub-counties divide into parishes/villages, which divide further into cells/villages.

that did not have centralized governments, such as those found in the eastern parts of the country

Specifically, when retracing the ethnic borders, the ethnicity with the highest proportion in a district based on UBOS data combined with historical settlement patterns supersedes these geographic boundaries established by the British colonial government. Additionally, This study ignores the smallest ethnicities within each district. Take Abim district, for example, the population of Abim is 87% of Karimojong ethnicity and geographically belongs to the Karimojong subregion. Using both UBOS data and historic settlement, this study identifies Abim within Karimojong borders when running the ML algorithm. However, Abim comprises other small minority groups, such as Gimara (0.033%). By ignoring ethnic groups that make up 13% of Abim's population, I am implicitly assuming that the smallest ethnicities are forced to assimilate with the largest ethnic groups within that district, or they are immigrant groups.

I use two formulas when allocating each district to the ethnic border (I) proportion of the highest ethnicity in the district and (II) ethnic fractionalization index. The two methods should give very similar borders. I use both for consistency. The ethnic fractionalization index introduced in Hudson and Taylor (1972) gives the probability that two randomly from a region (a district in this setting) belong to two different ethnic groups. I.e.,

$$FRAC_j = \sum_{e=1}^{E} \pi_{je} \left(1 - \pi_{je}\right), \tag{A1}$$

where j indexes a district, π_{je} is the proportion of ethnic group e in district j. Using UBOS ethnicity breakdown data by district, county, and sub-county, (I) and (II) are highly correlated (-0.981).

Table A1: Ethnic fractionalization in a district

	N	mean	sd
Ethnic fractionalization index	135	0.388	0.26
Max proportion in a district	135	0.727	0.22

From Table A1, the average proportion of the largest ethnicity in a district is 0.727, and the median is even higher (median=0.802). This implies that it is rare to find districts with equal shares of ethnicities. The average probability that two individuals are randomly selected from a district is low, and the median is also lower (0.345).[14] However, I compute this probability for the whole country, and I get 0.933. This is the same value reported in Alesina et al. (2003). Therefore, although Uganda is ethnically diverse as a whole, its subnational units are not. When constructing the training sample, I restrict districts where the ethnic fractionalization index is low (< 0.5), and the max proportion in a district is 0.7 and above.

Even though UBOS reports that Uganda has over 50 ethnic groups, 45 (68.2%) of the 66 ethnic groups reported in 2014 census data contribute to less than 1% of the population each, and 22 ethnicities (33%) contribute a combined total of less than 1% of Uganda's population. The smallest ethnic groups are either non-Ugandan immigrant groups or indigenous groups. The immigrant ethnicities may be scattered across the country or segregated in the refugee

[14]This is based on 66 ethnic groups in the census data. When I use ethnic clusters/language groups from Table A2, this index falls to 0.235

resettlement areas.[15] The indigenous groups are tiny in that even though they are segregated, they only make up a small part of the district population. This leaves 32 unique ethnicities (out of 66) based on district and ethnicity clusters.

Students do not report ethnicity or places of origin during the application stage but their home districts. Although I observe home districts for most students, using reported districts would ignore cases of internal migration, especially rural-urban migration. Instead, I use students' surnames to predict their ethnicity as Ugandans' last names are almost usually in their native language, as Section 4.2 highlights. I combine ethnicities whose languages have high lexical similarity and mutual intelligibility to create a language group to proxy ethnicity.

Using language groups to proxy ethnicity has been used in several African studies to proxy ethnicity (e.g., Eifert et al., 2010; Depetris-Chauvin and Durante, 2017) as language and ethnicity usually overlap. The similarity in languages implies similarities in cultures, facilitating the ease of interaction in ethnically heterogeneous societies. Moreover, although not always, local languages in different follow a dialect continuum, which further informs my language groups/ethnicity. For example, historical and current Ankole and Kigezi people living in the SW part speak the same language but with different accents and are therefore combined to form the "Banyankore/kiga" ethnic group. Another basis for combining two or more ethnicities is historical. For example, the Tooro kingdom (Batoro) was historically part of the Bunyoro kingdom (Banyoro) until the early 19th century (Turyahikayo, 1976). Therefore, Batoro and Banyoro form one ethnicity (language group). Combining groups that are mutually intelligible and similar reduces the ethnic groups to 16 groups. Another concern for the performance of the classification algorithm is how segregated ethnicities are. As Figure1 portrays, ethnicities within Uganda are geographically segregated.

[15] UNCHR ranks Uganda as the fifth largest refugee host nation. See this link: accessed 4/14/23

Table A2: Ethnicity/Language Group Composition

Ethnicity/language group	**Composition**	**Number**
Alur_Jonam	Alur, Jonam	2
SW	Banyankore, Bakiga, Bafumbira, Banyaruguru, Banyarwanda, Batagwenda, Barundi, Bahororo	8
Ganda	Baganda	1
Gisu	Bagisu and Babukus	2
Iteso	Iteso	1
Jopadhola	Jopadhola	1
Kakwa	Kakwa	1
Kelenjin	Pokot and Sabiny	2
Karimojong	Karamoja, Jie, Dodoth, Napore, Nyagia	5
Madi	Madi	1
Northern Luo	Acholi, Lango, Kumam, and Ethur	4
Nyoro	Batuku, Bunyoro, Batoro, Bagungu, Babwisi	5
Rwenzori	Bakonzo, Baamba	1
Samia_nyole_gwe	Banyole, Basmia, Bagwe	3
Soga	Basoga, Bagwere, Bakenyi	3
West Nile	Lugbara, Aringa	2
Extremely small	Vonoma(.008%), SoTopeth(.007%), Shana(.003%), Reli (.025%), Chope(.102%), Nube(.086%), Ngikutio(.017%), Mvuba(.009%), Mening(.008%), Lendu(.056%), Kuku(.140%), Kebuokebu(.161%), Bahehe(.012%), Gimar(.03%), Ikteuso(.041%), Batwa(.018%), Baruli(.565%), Banyabutumbi(.03%), Banyabindi(.049%), Aliba(.006%), Banyara(.142%), Nyangia(0.028%), Non-Ugandan(1.4%)	24
All		66

Notes: Source is the Uganda population and housing census of 2014. Groupings were informed using several sources as this section mentions.

8.3 Deriving the Reduced-Form Peer Effect

As described in the main text, this paper estimate the reduced-form peer effect based on random dorm assignment. In this section, I derive and discuss the relationship between this reduced-form estimate and the true underlying peer effect. Starting with equation (4) and simplifying subscripts, we can write the individual specific effect of 'actual' high-ability share, $\widetilde{S}_i$ on student i's grade as

$$Y_i = \rho X_i + \phi \widetilde{S}_{iG} + e_i \tag{A2}$$

where ϕ is the effect of the share of high-ability peers in a student's peer group on her academic performance. If I observed both random dorm assignment and actual (endogenous) dorm residence, it would be natural to use an IV approach to estimate the local average treatment effect of peers on academic performance, using dorm assignment to instrument for dorm residence as follows:

$$\widetilde{S}_{iG} = \kappa_{10} X_i + \kappa_{11} S^H_{iG} + e_1 i \tag{A3a}$$

$$Y_i = \kappa_{20} X_i + \kappa_{21} S^H_{iG} + e_2 i \tag{A3b}$$

where S^H_{iG} is the share of high-ability peers computed from peer groups as the result of the dorm assignment as in equation (4) that may not be equal to $\widetilde{S}_{iG}$ because some students do not live in dorms. Equation (A3a) as the first stage capturing the effect S^H_{iG} on $\widetilde{S}_{iG}$, while κ_{211} captures the reduced form of the high-ability share due to dorm assignment. Substituting equation (A3a) into equations A2 will give:

$$\kappa_{20} \equiv \rho + \kappa_{10} \tag{A4a}$$

$$\kappa_{21} \equiv \phi \kappa_{11} \tag{A4b}$$

$$e_{2i} \equiv \phi e_{1i} + e_i \tag{A4c}$$

Thus, the true high-ability peer effect (ϕ) is equal to $\frac{\kappa_{21}}{\kappa_{11}}$. That is, the IV estimate weights the reduced-form effect by the inverse of the first stage. Since I only observe dorm assignment, not residence, I am unable to recover this structural peer effects coefficient, so estimates captured in equation (4) are reduced-form estimates of peer effects based on dorm assignment.

Table A3: Differential effect by Ethnic Salience (Nonmajority) Coethnic vs High-ability Share

	Year One	Year Two	Year Three
Coethnic share	0.813*	1.037**	0.422
	(0.47)	(0.48)	(0.46)
High-ability share	0.605**	0.735**	1.000***
	(0.28)	(0.29)	(0.29)
High ethnic salience (nonmajority)	−1.394***	−1.043***	−0.644**
	(0.27)	(0.27)	(0.26)
Coethnic share × high ethnic salience (nonmajority)	3.254*	2.243	1.286
	(1.71)	(1.70)	(1.66)
High-ability share × high ethnic salience (nonmajority)	1.355**	1.095*	0.972*
	(0.56)	(0.59)	(0.52)
p-val Coethnic share (nonmajority): high ethnic salience	0.016	0.054	0.298
p-val High-ability share (nonmajority): high ethnic salience	0.026	0.084	0.173
R-squared	0.328	0.384	0.380
N	321,375	343,761	330,158
Dorm FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Group controls	Yes	Yes	Yes

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group comprises students admitted to majors within the same school and assigned to the same dorm. Each column is an independent regression, but the outcome is course grades in all regressions. All regressions control for own ethnicity, own ability, and major, HS subject combination, and classroom FE. Individual controls include age, religious indicators, and graduating from the district of origin. Group controls include the leave-me-out averages of individual controls in addition to peer group size. SEs are parentheses and are clustered at the peer group level. Nonmajority ethnicities exclude the largest two groups (Banyankore/Kiga and Baganda). The table also reports p-values for coethnic share and high-ability share of arts degree. These tests correspond to $\phi_1 + \varphi_1$ and $\phi_2 + \varphi_2$ in equation (6).
* p<0.10, ** p<0.05, *** p<0.01

8.4 Additional Results

8.4.1 Differential Impacts by Ethnic Salience

The results presented in this section should be interpreted in conjunction with the effects in Section 6.3.4. I proxy high ethnic salience as graduating high school from one's district of origin. As illustrated in Figure 1, non-majority groups are even more segregated and might consequently encounter greater diversity shock when they relocate to the capital for university education. This is especially true since they are also the most underrepresented group at MUK. I present the differential effect by diversity shock in Table A3.

8.4.2 More on Robustness checks

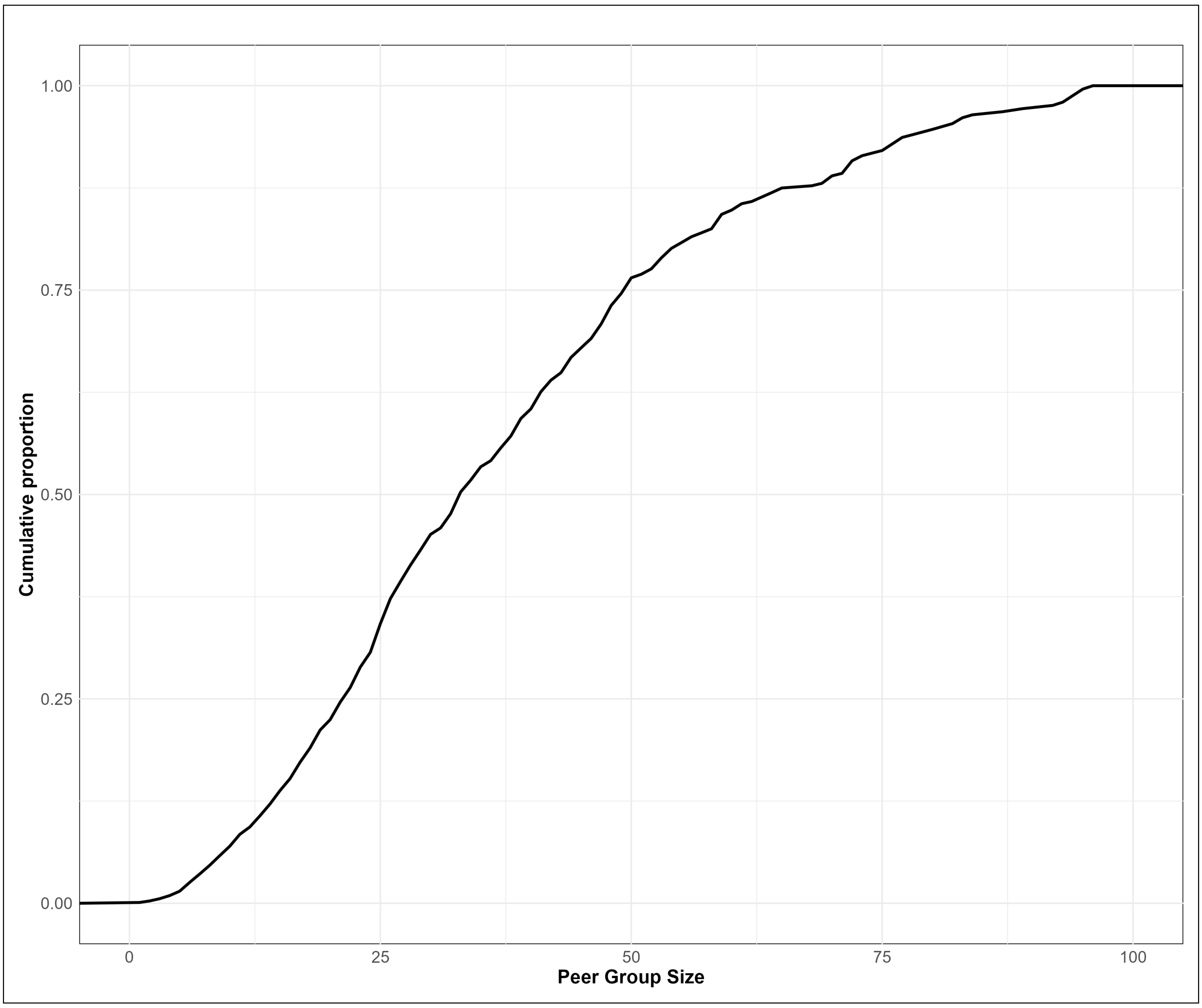
Table A4: More Evidence against Selection

	Coethnic share	High-ability share
Age	0.000	0.000
	(0.00)	(0.00)
Anglican	-0.000	-0.002
	(0.00)	(0.00)
Catholic	0.001	-0.002
	(0.00)	(0.00)
Muslim	0.001	-0.003
	(0.00)	(0.00)
Seventh Day Adventist	-0.003	-0.004
	(0.00)	(0.01)
Pentecostal	-0.002	0.002
	(0.00)	(0.00)
High ethnic salience	0.001	0.002
	(0.00)	(0.00)
Other Religions	0.001	0.007
	(0.01)	(0.01)
High-ability	0.000	0.011***
	(0.00)	(0.00)
Peer group Size	0.000**	-0.000
	(0.00)	(0.00)
R-squared	0.420	0.263
N	25,323	25,323
Joint Fstat	0.84	2.15

Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. Each column is an independent regression that regresses either the coethnic share or high-ability share on all pre-university characteristics. All regressions include school-by-year FE (not classroom), ethnicity, and dorm FE. SEs clustered at the peer group level.
*p<0.1, **p<0.5, ***p<0.001

8.5 List of Figures

Figure 6: Distribution of Peer Group Sizes.



Notes: Data are from MUK and are restricted to students admitted to non-extension day majors from six colleges for 2009-2017, excluding 2010. A peer group includes students admitted to majors within the same school and assigned to the same dorm.