
IMPROVING AND ALIGNING MEASUREMENT OF ETHNICITY IN LATIN AMERICA

A PREPRINT

Enrique Eduardo Valencia López*
Graduate School of Education
University of California-Berkeley
Berkeley, California
evalencialopez@berkeley.edu

December 9, 2019

ABSTRACT

Persistent inequality is a characteristic of many Latin American societies. Research has shown that Latin American countries remain stratified in the basis of gender, geographical location and income levels. More recently, a growing body of research has provided evidence of wide disparities between ethnic groups. But using self-identification to measure indigenous identity faces distinct challenges in the region as indigenous identity is considered fluid and multidimensional. Ethnic mobility, or the change of membership from one ethnic group to another over time, may challenge a country's ability to develop and target appropriate policies effectively.

In this study, I look at how different Latin American countries have measured indigenous identity in the census, household and international surveys. By looking at these three instruments of data collection, I find challenges in terms of comparability and limited data collection that result in a lack of comprehensive and systematic information among indigenous populations in countries within Latin America.

In a second analysis, I look at the challenges of using self-identification as the prevalent criterion to measure indigenous identity in Mexico. Using both household and longitudinal surveys, I assess the extent to which the three different criteria—self-identification, household and linguistic—measure indigenous identity and represent distinct dimensions of ethnicity. These dimensions are important in a country characterised by supporting notions of mestizaje. Longitudinal data allows me to measure the extent of indigenous fluidity for youth in Mexico and find that it is related to distinct personal characteristics.

Looking at a myriad of instruments, I conclude that while self-identification should remain the central criterion and countries should extend it to their other instruments, taking care the phrasing of the questions and unit of analyses are comparable. However, I also argue that further analyses to understand at ethnic disparities should consider the multidimensional and fluid aspects of indigenous identity.

Keywords Indigenous · Ethnic Fluidity · Ethnic Mobility

*This project was funded through a Global Education Monitoring Report Fellowship. I would like to thank Prof. Sylvia Schmelkes for comments on a previous version of this document and helpful discussions; Prof. Erin Murphy-Graham and Prof. Bruce Fuller at the University of California-Berkeley for helpful feedback. I would also like to thank Hector Robles Vasquez, Itzel Garcia Orozco, Raul Rene Rojas Olmos, Miguel Ángel Morales Hernández and Raúl Guadalupe Antonio from the Mexican Ministry of Education (The National Commission on Improvement on Education) for comments to an earlier draft of this paper. All errors remain my own.

Contents

1	Introduction	3
2	Analysis of national and international instruments that conceptualise and measure ethnicity in Latin America	6
2.1	Literature Review and Conceptual Framework	6
2.1.1	Sources of data collection	6
2.2	Diverse measures of indigenous identity	7
2.3	Indigenous identity as Fluid	10
3	Effects of alternative measurements on the identification of educational performance and inequality between indigenous peoples and other groups in Latin America.	12
3.1	Analytical Plan	12
3.2	Data sources and sample selection	12
3.3	First analysis: Census and Household Data	13
3.3.1	Descriptive results	13
4	Second analysis: Leveraging Differential Classifications of Indigenous Identity	16
4.0.1	Dependent variable	17
4.0.2	Inferential results	19
5	Third Analysis: Ethnic Fluidity in Mexico.	22
5.1	Changes of self-identified ethnic identity across time	22
5.2	Sample selection	22
5.3	Inferential results	28
6	Conclusions	31
7	Annex: Recommendations	32
8	Annex: Binary Logistic Regression with Fixed Effects	34
	References	35

1 Introduction

Persistent inequalities are a characteristic of many Latin American²societies (Telles, 2007; Bustillo, Artecona, & Perrotti, 2018; CEPAL, 2017). Statistics show that, despite progress, there are still significant disparities in the distribution of resources, opportunities, and outcomes between genders (Camou, Maubrigades, & Thorp, 2016), income levels (H. Lopez & Perry, 2008) and geographical areas (Macedo, 2012).

More recently, a growing body of literature from the past twenty years has identified that Latin America also remains stratified based on ethnicity. By practically all measures of well-being, ethnic minorities are one of the most adversely affected groups by the region's development challenges and fare much worse when compared to each country's dominant group (Telles, 2007; Telles, Flores, & Urrea-Giraldo, 2015; Psacharopoulos & Patrinos, 2004; Hall & Patrinos, 2012; INEE, 2017a; OEI, 2015).

Such increased visibility of the ethnic groups' living conditions, particularly indigenous peoples, has been the result of several factors: a persistent demand by indigenous movements to place their needs in the political agenda (Freire et al., 2015; CEPAL, 2014); greater commitment from states to sign international agreements that recognise indigenous rights, mainly the 169 ILO Convention (Freire et al., 2015; Curtis, 2009); and openness to interculturalism (Rodríguez Cruz, 2018). Progress has been visible: By the end of the 2010s, 17 countries—virtually all of continental Latin America—will have incorporated some measure of ethnicity in their latest round of census data collection. This situation contrasts with 2000, where the number of countries that had at least one ethnicity question on their census questionnaire was 13, up from six countries in the 1980s. (Del Popolo, 2008)

Despite making progress to collect information on ethnicity, Latin American countries still face significant challenges to determine the exact number, distribution and living conditions of the different ethnic minorities living in their territories, including the indigenous peoples. Several reasons underlie this challenge. First, demographic shifts have blurred ethnic boundaries and have given rise to new forms of indigenous identities. Migration, intermarriage and persistent stigmatisation has undermined language as the main marker of indigenous identity and has created different typologies of indigenous peoples (Telles & Torche, 2018). Second, elites in Latin America had previously favoured *mestizaje*, or racial and cultural mixing, as a political tool for nation-building. Under this project, societies tolerated marriages between indigenous and non-indigenous but isolated those that failed to conform to ideals of mestizo citizenship. It also meant that debates on indigenous identity were virtually non-existent and conceived as one-dimensional. Recent research, however, has challenged these notions of ethnicity by conceptualising it as a multi-dimensional construct that changes over time and place. Third, in some cases operational and legal constraints have limited the amount and type of ethnic categories that are included in the different data collection instruments, which may force individuals to misrepresent their ethnicity.

It is in this context that this paper looks at two main questions: *how do national instruments conceptualise and measure ethnicity, specifically indigenous identity, in a selected group of Latin American countries? And how do these approaches provide consistent estimates of educational indicators performance?* As Latin American countries have increasingly used a diverse set of approaches while favouring self-identification, I argue that they should develop a framework that conceives ethnicity as multidimensional and fluid. A multidimensional approach will allow researchers and practitioners to better characterise, target, expose and understand the different factors behind the indigenous peoples historical deprivations. Recognising fluidity means that looking at self-identification, while fundamental, may fail to provide consistent estimates of relevant outcomes. Research, particularly in the English-speaking countries, has found a considerable level of fluidity among ethnic and racial groups, which makes the targeting of policies more difficult³. For example, self-identified indigenous that speak Spanish face different challenges than those who speak an indigenous language—the latter implies policies prioritising textbooks and teachers that can communicate in their mother tongue with their students whereas the former potentially requires emphasising a multicultural curriculum.

Identifying and understanding the prevalence and causes of ethnic fluidity, a phenomena that groups ethnic mobility and context effects, can also contribute to effective research and monitoring of policies. Oftentimes, ethnicity is treated as a fixed, objective and context-free concept that can fail to account for new social dynamics derived from political, societal and demographic shifts (Brown, Callister, Carter, & Engler, 2010). Researchers on inequality and

²I define Latin America as the region in the Americas where Spanish, Portuguese and/or French are one of the main official languages.

³Self-identification should still constitute the central criterion, however. The United Nations has already issued a set of principles for ethnicity data collection that establishes self-identification as primordial, but analyses of ethnicity should probably be complemented with other criteria given the challenges research has found. Briefly, these principles are: 1) free self-declarations/open questions should be used; 2) respondents should be able to indicate more than one ethnic affiliation; 3) categories for 'none' or 'not declared' should be allowed; 4) instructions should be provided on determining the ethnicity of children of mixed couples; and 5) the basic criteria and classification procedures should be documented (United Nations, 2017; OECD, 2018).

ethnicity have generally addressed the fluidity of indigenous membership by adopting frameworks that sometimes exclude younger individuals from the analysis (for example children below 15) or frame how groups are to be compared (for example comparing individuals depending on whether they attained tertiary education) (Villarreal, 2014; Perez Cardenas, 2018).

If changes in ethnic self-identification are related to social conditions, then potentially, the characteristics of a group may remain stalled or improve not because of real policy actions but rather because of changing group membership. In some cases, individuals may feel the need to identify themselves differently when they move to urban areas, enter employment or gain more education in order to fit into stereotypes of what a successful individual is. However, the 'racialised' and ethnic-based structure of society remains the same (Saperstein & Penner, 2014). Recognising the fluidity of ethnicity requires societies to revise their stereotypes and continue promoting diversity. Finally, estimating the extent of indigenous ethnic fluidity may help understand the extent to which current categories to measure indigenous identity (or ethnicity in a broader sense) reflect how individuals understand these constructs in everyday social interactions⁴. (Saperstein & Penner, 2012).

Peoples' responses to ethnicity questions may reflect long-lasting changes on their ethnic affiliation over time (ethnic mobility) (Statistics New Zealand, 2005). They can also refer to specific situations derived from how the instrument is administered or the perceived objective of the survey, in which case the ethnic transition is short term (context effects). An example of context effects occurs when an individual's ethnic affiliation depends on the perceived ethnicity of the interviewer.(OECD, 2018). Both of these effects, contextual and mobility, however, show that ethnicity is fluid.

To look at the multidimensional and fluid components of ethnic identity, I will focus on indigenous peoples, as they are one of the largest ethnic groups in the Latin American region. I first proceed by looking at how countries have measured ethnicity, particularly indigenous identity, on their censuses, household and international surveys in the past twenty years. While these are not the only instruments that collect information on indigenous peoples, they are widely used to monitor progress towards attaining national and worldwide goals in education. I focus on education due to its importance as both an input and an outcome of inequality. Schooling has also historically been the main tool states have used to transfer conceptions of indigenous identity (Flores & Telles, 2012). A descriptive analysis compares educational and demographic estimates using the two main criteria countries have used in the past twenty years: language and self-identification and in the case of Mexico, an additional household criteria developed by the National Commission on Indigenous Peoples(*Comisión Nacional de los Pueblos Indígenas, CDI*, in Spanish).

An initial review of the different instruments shows that countries have adopted a diversity of criteria to measure indigenous identity both between them and in their own instruments. Looking at both International and National Household Surveys (NHS) as well as the Integrated Public Use Microdata Series (IPUMS), I find that international instruments such as the Multiple Indicator Cluster Survey (MICS) and the Demographic Household Survey (DHS) have occasionally measured indigenous identity differently. For example, the 2008 Bolivian DHS includes a question that measures ethnicity as a continuum; however, none of Bolivia's instruments adopts a similar approach, which results in incomparable statistics. Some countries have also failed to incorporate a criterion in their household surveys. Since these surveys usually provide information on the living conditions of a country's population in ways that censuses cannot, it is difficult to gain a deeper understanding of the socioeconomic disadvantage between indigenous and non-indigenous peoples. To improve the reporting of relevant outcomes and monitoring progress, countries could apply consistent ethnicity questions across their different instruments⁵.

In the second and third section, I focus on Mexico to analyse whether different criteria reflect different dimensions of indigenous identity. I select Mexico because it is one of the largest and most ethnically diverse countries in the Latin American region. By looking at the National Survey on Household Income and Expenditure (ENIGH, in Spanish), I find that consistent with previous literature, language captures additional information net of self-identification. At the same time, the effect of self-identification diminishes once the wealth of the household is taken into account for individuals aged 25 and older.

Finally, I use the National Survey on Household Living Conditions (ENNVIH, in Spanish), a longitudinal survey, to estimate how fluid is indigenous identity in Mexico. I measure this fluidity by looking at individuals' patterns of ethnic identification across the three waves of the survey. Among the advantages of using this survey is that the question of ethnicity remains comparable over time and individuals are required to answer the survey questionnaire themselves, a strategy that differs from censuses and most household surveys as these rely on a main informant. Results show

⁴While I use the term ethnic fluidity, the focus of the paper will be on indigenous peoples ethnic fluidity

⁵This is not to claim that countries should converge to a single criterion. Instead, countries should develop criteria that will allow them to understand their indigenous peoples living conditions in a manner that is consistent across the different data collection instruments and aligns with their lived experience

that individuals who identify as indigenous in the first wave (2002) exhibit high levels of fluidity whereas those that identify as non-indigenous are more stable. Among youth (15 to 24 years old) speaking an indigenous language is associated with lower likelihood of ethnic transition in 2005 whereas living in an urban area increases the likelihood of inconsistency. Attaining basic education was unrelated to self-stated indigenous identity.

2 Analysis of national and international instruments that conceptualise and measure ethnicity in Latin America

2.1 Literature Review and Conceptual Framework

2.1.1 Sources of data collection

In the past twenty years, countries in Latin America have relied on three instruments to measure indigenous identity: Censuses, household and international surveys. Among these, censuses are the prevailing source of data collection to measure the living standards of indigenous peoples. One advantage of using censuses is that countries can obtain basic statistics about their population disaggregated at very small geographical areas (such as locations) and for very particular groups. As census obtain individual records, they capture information from persons that belong to groups living in isolated locations.

Censuses have some disadvantages: due to cost and its purpose of collecting a broad range of topics, they usually do not include comprehensive information about the living conditions of individuals. Their periodicity, generally of ten years, means that sometimes any relevant information they provide is not timely for policy-making purposes.

Household surveys are used to obtain information not captured in the census or administrative records (OECD, 2018). Examples of this information in Latin America are income, perception of discrimination, and spending patterns. These instruments have the advantage of being periodically collected, which makes them suitable to monitor socioeconomic indicators. However, household surveys do not necessarily provide reliable and valid estimates to disaggregate information of disadvantaged groups, especially if these are spatially distributed in a different manner than the rest of the population. Sampling procedures of household surveys tend to exclude small areas⁶, that can be remotely located—areas where a considerable amount of indigenous peoples still inhabit (de Alba, 2017).

International surveys such as the Multiple Indicator Cluster Survey (MICS) and the Demographic Health Survey (DHS) collect data that allow for monitoring of a wide range of impact evaluation indicators in areas such as health, education, nutrition and skills. Increasingly, they have also become the main source to monitor a country's progress towards attaining the Sustainable Development Goals (SDGs).

Despite their usefulness, there are some limitations of using these instruments to assess indigenous well-being: sampling procedures oftentimes do not allow data to be disaggregated beyond subregions. Questions on ethnicity are not entirely comparable with national instruments, which may result in different conclusions about the living conditions of particular groups. For example, in Guatemala, the 2013 and 2014 National Survey on Income and Employment (ENEI) has questions related to ethnicity slightly differently from the 2014 DHS survey—while both surveys require the individual to identify with an indigenous group, only the DHS provides a predetermined list of these groups⁷. Timeliness is another challenge since some countries in the region have not collected information from this survey recently. In Bolivia, for example, the last time it was used was 2008. Ecuador collected a DHS in 1987.

Administrative data could potentially complement information from survey data on well-being. The advantages of these sources is that they can be integrated and linked systematically to a data infrastructure that can produce reliable information at the person-level (OECD, 2018). However, to monitor system-level outcomes, administrative sources need to be harmonised to avoid being affected by numerator-denominator bias (OECD, 2018). One example concerns some educational indicators that generally use figures from school records as numerators and census data in the denominator (for example, enrolment rate). But usually these estimations are not harmonised and indicators frequently reveal implausible values (for example enrolment rates usually go beyond 100%)⁸. In some cases, administrative data collects information differently from those of the census—questions are worded differently, for example. Administrative information is also prone to considerable measurement and sampling error since main informants, usually a school principal, relies on memory to register information. Likewise schools may misreport or provide information late depending on their technical capabilities (INEE, 2017a). Finally, evidence also shows that administrative records are prone to large misclassification or under-identification errors of indigenous peoples (Thompson, Woods, & Katzenellenbogen, 2012).

Despite an ample array of instruments for data collection, there are several reasons why Latin America has not provided a consistent and precise estimate of its indigenous peoples. First, countries apply different criteria to measure indigenous identity: Peru, for example, has until recently relied on language as the main marker of indigenous identity while Mexico developed an official household criterion (Table 1). Heterogeneous statistical capacities may limit a

⁶for example, Mexico's ENIGH includes as Primary Sampling Units areas that have at least 160 dwellings (INEGI, 2018)

⁷The question in ENEI is *Do you consider that you belong to the following peoples?* And excludes from the categories "Other" while the DHS asks, *Do you consider yourself: maya, ladina/mestiza, Garifuna, xinca or from other ethnic group?*

⁸This problem is ameliorated by estimating net enrolment rates (NER)

countries' capacity to collect census data efficiently and failure to meet minimum data quality standards has required in some cases a second round of censuses (Paraguay and Chile) (CEPAL, 2017). Sampling procedures sometimes result in national and regional under-coverage of certain groups such as the indigenous forest peoples (Thiede & Gray, 2020)

Second, censuses and household surveys frequently fail to account for the complete migration patterns of indigenous populations, particularly international movements because most surveys collect only information of household residents at the time of the interview. Moreover, census questionnaires which fail to incorporate open questions on ethnic identity may force migrants to misrepresent their identity by constraining them to a specific set of nationally recognised groups. Such was the case of the Chilean census in 2002 that collected only information of indigenous groups legally recognised in the country⁹ (Aravena Reyes, 2014).

Third, while most countries have included ethnicity questions in their censuses, a minority of them have included them into their household and international surveys. Argentina, for example, has not fully added ethnic variables into its main household surveys as of 2019. In Central American countries, very few countries provide updated and continuous information on ethnic groups.

Fourth, research in the region had framed differences in social and economic outcomes among ethnic groups as a social class problem partly due to pervasive notions of *mestizaje*, or racial and cultural mixing (Wickstrom & D. Young, 2014). In *mestizaje*, a project of nation-building, Latin American societies were characterised as raceless and culturally homogeneous¹⁰ (Paredes, 2018; Telles et al., 2015; Wickstrom & D. Young, 2014) and any ethnic differences were justified as being due to poverty. But studies of intergenerational mobility in Mexico, for example, have challenged the traditional notion that parents' class is the main predictor for indigenous children's outcomes and have found that skin colour or language also explain socioeconomic differences, suggesting the existence of systematic and structural discrimination¹¹ (Flores & Telles, 2012).

Finally, while most countries have converged to self-identification as the main variable to measure ethnicity, a criterion consistent with the 169 ILO Convention on Indigenous Peoples, it is likely that more information is needed to better characterise indigenous peoples in Latin America in order to understand the structure of inequality behind their living conditions. Research from Mexico, Peru and the Project on Ethnicity and Race (PERLA), for example, has shown that population estimates of indigenous peoples vary considerably depending on whether one takes into account the linguistic or self-identity criteria (Telles & Torche, 2018; INEE, 2017a). As self-identified indigenous speak an indigenous language, educational outcomes tend to worsen and are the lowest for those who are indigenous monolingual compared to indigenous who do not speak an indigenous language (INEE, 2017a). An emphasis solely in distinguishing indigenous groups on the basis of language and self-identity, while important and fundamental (Parker, Rubalcava, & Teruel, 2002), may not fully capture the complex facets of ethnic identity and the role other variables may play in preserving inequality. For example, there is evidence that indigenous groups face geographical exclusion (Del Popolo, Oyarce, Ribotta, & Rodríguez Vignoli, 2007) that may reinforce historically accumulated disadvantages. But so far, no country collects the different dimensions observed in the region in one single instrument

2.2 Diverse measures of indigenous identity

Countries in Latin America have gradually converged to self-identification as the main criterion to measure indigenous identity. Over the period of 2000 and 2010s, the amount of countries that included a question related to an individual identifying as indigenous or as a member of an indigenous group in their census increased from 13 to 17—the only exceptions being Dominican Republic and Cuba (Table 1).

The preference for self-identification is endorsed by international institutions and agreements and originates from the 169 ILO Convention on Indigenous and Tribal Peoples (Freire et al., 2015). This convention, while aligning with the United Nations (UN) decision against a formal definition of indigenous peoples, provides an operationalisation of the concept: an indigenous person is one that “descends from populations who inhabited the country or geographical region at the time of conquest, colonisation or establishment of present state boundaries” (ILO, 1989) and they retain some of the social, economic, cultural and political institutions.

Despite the prevalence for self-identification countries have also used alternative criteria to measure indigenous identity, which can be grouped into four dimensions (Del Popolo, 2008): i) Recognition of identity; ii) Common origin;

⁹The 2002 Chilean census could not collect data on international indigenous migrants as their census question restricted self-identification to a specific set of recognised groups in the country and did not have a question for "Other"

¹⁰Refer to Wickstrom (2014) for a review of how the concept of *mestizaje* changed over time in Latin America

¹¹A limited set of studies have analysed the discrimination of indigenous peoples in the labour market in Mexico (Patrinós, 2000; Canedo, 2019), Guatemala (Canelas & Gisselquist, 2018) and Bolivia (Chiswick, Patrinós, & Hurst, 2000). A recent review found that in Mexico, while a large part of the income gap between self-identified indigenous and non-indigenous is due to locality size, regional disparities and schooling, there is an unexplained component attributable to labour market discrimination (de Alba, 2017)

iii) Territoriality; and iv) Cultural-linguistic (Appendix A). Recognition of identity is the sense of belonging with ethnic identity derived from indigenous consciousness; common origin refers to having common ancestors. It involves incorporating the collective memory of the peoples and the recreation of their past. Territoriality is linked with the indigenous peoples' occupancy of lands, its use and the material as well as spiritual relations that these groups have. Finally, the cultural-linguistic criteria refer to the closeness with culture, the origin, language and social and political organisations derived from it (Del Popolo, 2008).

In this myriad of criteria, some researchers have proposed that recognition of identity remains the central criterion but prompt for the use of the rest of the dimensions to further characterise the living conditions of indigenous peoples (Del Popolo, 2008; Paredes, 2018; Telles & Torche, 2018). They point to recent studies that challenge the idea that ethnicity is a one-dimensional and fixed characteristic. Instead, demographic, political and social changes have contributed to revive or create new forms of ethnic identity that challenge the traditional notion that an indigenous person speaks an indigenous language, lives in ancestral lands and maintains certain cultural traits. For example, in some countries of the region, it is reported that individuals who identify as indigenous do not necessarily speak an indigenous language while those who speak an indigenous language do not identify as indigenous. Countries that also collect information on African ascendancy allow individuals to identify both as afro-Latin Americans and indigenous (INEE, 2017a). As a result of these changes, a growing body of research suggests that in some contexts, self-identification as indigenous may carry more of a symbolic weight (Villarreal, 2014). This symbolic weight means that indigenous identity is a cultural resource that is devoid of any social implications or attachment to a certain group, but still provides some social grounding.

Table 1: Census, National Income Household Surveys, DHS and MICS with Ethnic variable 2000-2020.

Country	Census	Household Survey	DHS	MICS
Argentina	2000, 2010, 2020*	2004-2005[1]		
Bolivia	2001, 2012, 2022	2017	2003, 2008	
Brazil	2000, 2010, 2020*	2015		
Chile	2002, 2012*, 2017*	2000-2017		
Colombia	2005, 2018*	2016-2017	2010, 2015	
Costa Rica	2000, 2011, 2020			2011, 2018*
Cuba	2012			
Dominican Republic				
Ecuador	2001, 2010, 2020*	2004, 2012*		
El Salvador	2007, 2018*			2014
Guatemala	2002, 2018*	2015, 2018	2014	
Haiti				
Honduras	2001, 2013		2011-2012	
Mexico	2000, 2010, 2015.	2008-2018		2015
Nicaragua	2005		2001	
Panama	2000, 2010, 2020*	2003, 2008		2013
Paraguay	2002, 2012*	2011-2012		2016
Peru	2007, 2017*	2003-2018	2004-2012, 2013-2014*	
Uruguay	2011	2006-2018		2012
Venezuela	2001			2000

Note [1]: This data corresponds to the Complementary Survey on Indigenous Peoples (CSIP, *Encuesta Complementaria sobre los Pueblos Indígenas*).

Note 2: [*] Data unavailable.

Note 3: Household surveys are: Argentina, Encuesta Permanente de Hogares; Bolivia: Encuesta Nacional de Hogares; Brazil, Pesquisa Nacional por Amostra de Domicílios; Chile, Encuesta de Caracterización Socioeconómica Nacional; Colombia, Encuesta Nacional de Presupuesto de los Hogares; Costa Rica, Encuesta Nacional de Hogares; Dominican Republic, Encuesta Nacional de Ingresos y Gastos en los Hogares; Ecuador, Encuesta Nacional de Ingresos y Gastos de los Hogares Urbanos; Guatemala, Encuesta Nacional de Empleos e Ingresos; Mexico, Encuesta Nacional de Ingreso y Gasto de los Hogares; Nicaragua, Encuesta de Ingresos y Gastos en los Hogares; Panama, Encuesta de Niveles de Vida; Paraguay, Encuesta de Ingresos y Gastos y Condiciones de Vida; Peru, Encuesta Nacional de Hogares sobre Condiciones de Vida y Pobreza (new methodology); Uruguay, Encuesta Continua de Hogares (new methodology);

Note 4: Data for Mexico for census in 2015 corresponds to the Inter-census survey.

Note 5: Chile had a census in 2012 but did not meet minimum data quality standards. Paraguay in 2012 conducted a Census, but failed to reach minimum coverage.

One consequence of these demographic changes is then that indigenous identity becomes more fluid than before and is recognised to be multidimensional (Villarreal, 2010; Flores & Telles, 2012; Saperstein & Penner, 2012; Paredes, 2018; Saperstein, Kizer, & Penner, 2016; Telles & Torche, 2018). By adopting this approach, it is possible to expand and provide evidence that ethnic identity reflects social dynamics related to socioeconomic inequality. Such perspective marks a shift from studies that had not sufficiently acknowledged ethnic disparities (Paredes, 2018; Carrión & Zárata, 2010) or had explained them as a result of class differences (Flores & Telles, 2012).

Studies had insufficiently recognised the role of ethnicity in creating and preserving socioeconomic disparities in Latin America partly due to the *mestizaje* ideology predominant in the region. Under *mestizaje*, politicians found a project of nation-building whereby indigenous and afro-descendants would assimilate the perceived strong and humanistic values from racial mixing (mestizos). This process of assimilation would occur mostly through education, where it played a role in over the region as a tool to integrate indigenous peoples into the state (Soto Quirós & Díaz Arias, 2006). Those that refused to comply were marginalised (Paredes, 2018), but those who conformed, obtained citizens' rights. One consequence was that Latin American societies adopted a discourse of being race-less and culturally homogeneous that made indigenous peoples invisible (Soto Quirós & Díaz Arias, 2006).

Given the prevalent view that modern Latin American societies were race-less, few studies challenged the notion that class was the most important cleavage. Socioeconomic disadvantages of indigenous peoples were expected to ameliorate once they assimilated into the mestizo culture. However, recent research in Mexico and Peru has shown persistent inequalities by skin colour, language, and in some countries, self-identification, even after controlling for class origins (Villarreal, 2014; Telles et al., 2015; Paredes, 2018).

2.3 Indigenous identity as Fluid

Despite considerable progress, research has shown that current approaches to measure ethnicity in Latin America provide inconsistent estimates of ethno-racial inequalities in the region (Flores & Telles, 2012). This inconsistency may derive from discrimination or racism, as much of this process depends on how others perceive a given individual (Telles et al., 2015; Saperstein & Penner, 2012). A study conducted among several Latin American countries comparing different criteria to measure ethnicity found that self-identification provides inconsistent estimates of the educational conditions of adult indigenous populations. In half of the countries surveyed—Colombia, Ecuador and Dominican Republic—self-identified indigenous had negative but non-statistically significant relationship with educational attainment net of class origins. This study, and subsequent others, have found that skin colour is a better predictor of years of schooling than the census criteria even after controlling for class origins (Flores & Telles, 2012).

Another challenge to self-identification is that even if individuals are asked about their indigenous affiliation, it is unclear how it could help estimate the potential demand for certain educational services. For example, some countries collect information on self-identification starting at age three. As countries begin to prioritise early childhood education, they may have to adopt a household approach to identify the number of self-identified indigenous children that are likely to enrol. However as children are in a period of identity formation, these estimates may be subject to considerable fluctuations (Kiang & Baldelomar, 2016).

To better measure those inequalities, countries could probably adopt a multidimensional and fluid perspective to the study of ethnicity¹². Research has shown that the fluidity results from: 1) the ambiguity of census categories, 2) changes in how people identify over time (ethnic mobility) and 3) the context or circumstances in which individuals are asked about their indigenous identity (context effects) (Statistics New Zealand, 2005; Saperstein & Penner, 2012).

First, the ambiguity of census categories. It is widely acknowledged that census categories reflect technical and political considerations that are idiosyncratic and may differ from how individuals understand, use and experience indigenous identification in daily life (Telles et al., 2015, 2015; Angosto-Ferrández & Kradolfer, 2014; Del Popolo, 2008). These categories can also make certain groups of indigenous peoples invisible. For example, Chile in 2002 only allowed indigenous peoples to identify their ethnicity based on a list of legally recognised groups, which excluded migrants who self-identified as indigenous (INE, 2012). Bolivia, in its most recent censuses, decided against including a mestizo (mixed) category, leaving “non-indigenous” without a category to select. The government argued that including mestizo would imply endorsing the “whitening” of Bolivian society as the term had such connotation in the country (Chirino, Almeida Garcia, & Morales Torrez, 2017). The census agency also changed the phrasing of the question in the 2012 exercise—it coalesced both indigenous and Bolivians of African-descent—which resulted in a reduction of Bolivians who identified as indigenous by 20 percentage points, a change unexplained by demographic change alone (Chirino et al., 2017).

Second, ethnic mobility. A large body of anthropological and sociological studies has shown that Indigenous identity can shift over time and place. Research has shown that individuals change their ethno-racial identity purposefully due to a variety of factors (Montgomery, 2011), which occur at both the macro and micro-level. At the macro level, greater recognition of indigenous rights as well as the emergence of indigenous movements may have contributed to de-stigmatise people’s perceptions of indigeneity. As a result, individuals may feel more confident to preserve their indigenous heritage or re-establish contact with it. On the other hand, social stigma and discrimination still persist in many areas and this situation may prevent individuals from stating their own self-indigenous identity (Carter, Hayward, Blakely, & Shaw, 2009).

These macro processes affected several policy domains. One of these was education, which has been both a policy instrument to hinder and promote indigenous identity. On one hand the once predominant mestizo ideology promoted, through school’s curriculum, a homogenous set of behaviours and attitudes of what constituted a citizen, which implicitly glorified “white” stereotypes and aimed to create a “race-less” and culturally homogeneous society at the expense of indigenous culture (de la Cadena, 2005; Freire et al., 2015). Moreover, linguistic policies promoted Spanish as the main language and many countries required proficiency in this language as a condition to incorporate indigenous peoples into this homogeneous and dominant mestizo culture (Cortina, 2014).

Starting the second half of the XX century, there was an educational shift in some countries in Latin America. As a result of greater political recognition, higher democratisation levels and collaboration between different sectors, Bilingual education (EIB) became the standard response to meet the needs of indigenous peoples. This change of paradigm contributed to societies to become more open and appreciative of indigenous cultures (H. Lopez & Perry, 2008). During the 1990s, more educational systems expanded the conceptual underpinning EIB to adopt a rights

¹²as opposed to converging to just one way to measure indigenous identity. Equally important, countries could also consult the indigenous communities about the planned categories to include in the different data collection census

perspective where interculturalism played a central role for all students. Interculturalism in education presupposes a respectful relationship between cultures where power dynamics are symmetric (Schmelkes, 2013).

Intercultural education also aims to enrich the understanding of non-indigenous students as they become aware of their culture, traditions and language (L. E. Lopez, 2017). If well implemented, this pedagogical model may lead to some students with indigenous ancestry to re-orient their identity and recognise as indigenous. Despite this progress, however, countries face the challenge of adopting an intercultural approach that is appropriate and effective and can translate into real improvement of the indigenous peoples' socioeconomic conditions (Schmelkes, 2013).

At the micro-level, migration to urban areas and loss of indigenous language proficiency over generations has also triggered complex processes of negotiation of identity. Anthropological research has documented how the presence of indigenous peoples in urban areas has led to ethno-genesis and new forms to express indigenous identity (Gomez Murillo, 2008). In some cases, for example Mexico and Guatemala, migration has resulted in the loss of indigenous languages, which used to be considered the main marker of indigenous identity (Telles & Torche, 2018; Yoshioka, 2010). One reason for such loss is that indigenous peoples seek access to the local labour market, which is predominantly Spanish speaking. This pattern is not uniform across the region. In Peru, for example, urban bilinguals who do not consider themselves ethnically indigenous may commonly speak Quechua or Aymara (Howard, 2011).

Third, contextual effects refer to changes in ethnicity derived from the social environment or circumstances in which individuals are asked about their indigenous identity. Common causes of these changes refer to how respondents understand the purpose of the instrument whether positive (identify cultural needs) or negative (discrimination) and the race or ethnicity of the interviewer (Statistics New Zealand, 2005; OECD, 2018).

Context effects may also occur more predominantly in some areas than others. For example, in the United States, an increase in Amerindian self-identification was common in "cosmopolitan areas" where organisations have managed to promote a sense of pride in locations with weak and largely dispersed indigenous ties (Nagel, 1995). In this situation, indigenous identity plays a symbolic role and becomes a cultural resource in a context where social stigma towards indigenous peoples has diminished (Alba, 1990). On the other hand, rural to urban migration of indigenous peoples may result in a decrease of indigenous identity, specially if grave discrimination persists. In Latin America, some anthropological studies have found that indigenous people that migrate from rural to urban areas are likely to mention they are non-indigenous due to several reasons: discrimination and stigma as well as weaker ties with their communities of origin and adaptation to urban culture (Trujano, 2008).

Research on ethnic fluidity has used three analytical approaches to assess the extent to which societies are ethnically mobile and the factors driving this change. The first approach looks at inter-generational transmission of identity between parents and children (Villarreal, 2014). Studies of this strand have shown how relevant are household and parental factors, such as education and intermarriage, in transmitting language, and potentially, a sense of indigenous identity. A limitation of this approach is that it relies on data collected through a main informant, which is generally an adult. In consequence, it fails to capture whether the child actually appropriates the indigenous identity of his or her parents and assumes that indigenous identity becomes stable during adulthood.

The second approach looks at the added value of different criteria to identify indigenous identity (Telles, 2007; Paredes, 2018; Saperstein et al., 2016). This research assesses the extent to which self-indigenous identity, the central criterion that countries have adopted, captures all the components of what constitutes indigenous identity. Researchers from this strand argue that indigenous identity is not fixed and one-dimensional, since self-identification does not capture all the phenotypic and cultural variation of more precisely defined categories, such as language or skin colour. One limitation of this approach is that it often fails to adjust for the endogeneity present in some variables. For example language is often correlated with other characteristics that are also thought to affect relevant socioeconomic indicators.

Finally, a third strand looks at the dynamic and individual changes of ethnic identity (Saperstein et al., 2016) By looking at longitudinal data, it is possible to explore whether changes in social status drive changes in self-identification and vice versa. Descriptive data from this research can also describe the patterns of ethnic transitions of individuals over time. One limitation is the small sample sizes of indigenous peoples and high attrition rates.

3 Effects of alternative measurements on the identification of educational performance and inequality between indigenous peoples and other groups in Latin America.

3.1 Analytical Plan

Changes in ethnic identity, or ethnic mobility, have the potential to make targeting of policies more difficult. In countries where there is low concordance of self-identification and indigenous language, policy responses in education also vary. For example, Mexico is required to adopt an intercultural indigenous curriculum in schools that have high levels of self-identified indigenous groups. Yet, whether this curriculum requires a linguistic component depends on how many children speak an indigenous language, which as discussed before, may not necessarily reflect those that recognise themselves as indigenous.

One potential solution in light of ethnic mobility is to incorporate in the different instruments' information about other dimensions of ethnic identity, for example, self-identification and ancestry, criteria that are currently used in New Zealand (Kukutai, 2004). But to identify relevant categories, it is necessary to first compare the effects of two different measures of indigenous identity in Latin America on socioeconomic outcomes.

Consistent with the literature on race in the United States, the most common approach to assess the different measures involves looking at one category of indigenous identity while estimating the differences in outcomes on the other dimension. For example, comparing speakers of indigenous languages versus non-speakers restricting the sample to those that self-identify as indigenous (INEE, 2017a).

Another approach, which I use in this paper, consists in looking at gross and net effects of different criteria on educational outcomes, following Sapperstein, Kizer & Penner's (2016) approach. The method consists of isolating the effect of speaking an indigenous language among people who do not self-identify as indigenous as well as people who self-identify as indigenous but does not speak an indigenous language. While Sapperstein, Kizer & Penner (2016) used it to compare self-identification and interviewer classification, their approach can extend to other dimensions. Finally, looking at longitudinal data provides information of changes in ethnic identity through time.

In the following section, I analyse how Latin American countries have measured indigenous identity to argue why the self-identification criterion, while useful and important, does not need to be the only variable to measure indigenous identity. Research has shown that when self-identification is used as the only measure of indigenous identity, it can provide inconsistent estimates of educational inequality in the region. Part of this inconsistency is due to the ambiguity of the categories, the mestizo context and, I argue, the fluidity of indigenous identity, where individuals change their ethnic affiliation, which in turn, may depend on sociodemographic characteristics.

3.2 Data sources and sample selection

I divide the data analyses into three. In the first section, I compare descriptive statistics of educational outcomes based on different criteria to identify indigenous peoples for the five countries that have the highest amount of indigenous populations—Mexico, Bolivia, Peru, Ecuador, Peru—and Paraguay. I include Paraguay because it has one of the higher proportions of a country that speaks an indigenous language in Latin America (over 70%). The main source is the census data, which we obtain it from the Integrated Public Use Microdata Series (IPUMS) otherwise noted. Household and International Surveys are listed in Table 1. We selected the data sources based on whether they were the most current; whether they included at least two criteria to identify indigenous populations and if the sampling design allowed for population inferences.

For the second section, I focus mainly in Mexico. I select this country for two reasons. First, it is one of the most diverse countries in the region with over 68 recognised native languages. It also has one of the largest populations of self-identified indigenous in Latin America: 36.5 million, or an estimated 30 percent of the Mexican population aged 3 and older in 2018. To conduct our analyses, I use the 2018 data sample from the National Household Survey on Income and Expenditure (ENIGH, in Spanish).

Finally, we use the first three waves (2005, 2009 and 2012) of the National Survey on Living Conditions (*Encuesta Nacional sobre Niveles de Vida en los Hogares*, ENNVIH, in Spanish) in Mexico. One advantage of using this survey is that individuals aged 12 years or older responded the question of ethnicity and therefore provide how they perceived themselves independent of how the head of household identifies. Another advantage of this survey is that the question remained comparable across the different waves. One limitation is that the survey does not capture data from the interviewer. This information is important to assess potential "social desirability bias" in the sense that the individual's responses may depend on the interviewer's personal characteristics.

3.3 First analysis: Census and Household Data

3.3.1 Descriptive results

Using De Popolo's approach (2008), we group the different criteria into four categories. Results show convergence to self-identification as the main criterion to identify indigenous peoples, except Peru which until recently relied entirely on language to measure indigenous identity. In more recent years, countries have also included additional questions of indigenous identity (Peru and Guatemala, for example) that are related to ancestry or cultural-linguistic components. But despite prevalence of ethnic questions in the censuses, it is still not a widespread practice to incorporate these questions into household surveys, particularly in Central American countries.

The phrasings of the questions also differ among countries, even those related to self-identification. On one hand, Paraguay and Argentina use as the main unit of analysis the household whereas for most countries it is the individual (Table 2). Phrasing of the question in terms of the household may be more consistent with data collection practices, which often rely on a main informant, but fail to recognise the individual's right to self-identify as member of an ethnic group (Table 2).

Table 2: Example of Criteria available in the latest census questionnaire available (2000-2018)

Recognition of identity		
Country	Year	Question
Bolivia	2012	29. As a Bolivian, do you belong to any nation or indigenous people who are originally farmers or Afro Bolivians?
Brazil	2010	6.05 Do you consider yourself indigenous?
Chile	2017	16. Do you consider yourself to belong to any indigenous or original population?
Colombia	2018	37. According to her or his culture, people or physical features, does [NAME] consider himself or herself as indigenous?
Costa Rica	2011	7. ¿(NAME) considers indigenous?
Ecuador	2010	16. How do you identify (NAME) according to your culture and customs?
El Salvador	2007	6. b. If you are indigenous, which group do you belong to?
Guatemala	2018	PCP12. According to your origin or history, how do you considered or self-identify?
Honduras	2013	6. To which indigenous people do you belong?
Mexico	2015	10. According to his or her culture, how does (NAME) identify?
Nicaragua	2005	7. To which of the following indigenous or ethnic groups do you belong?
Panama	2010	8. To which indigenous groups do you belong?
Paraguay	2012	42. Of the people who usually inhabit this household, do any of them considers indigenous?
Peru	2017	26. Based on your customs and ancestry, do you consider yourself to be?
Common Origin		
Argentina	2010	4.Does any person in this household is indigenous or descendant from an indigenous person?
Peru	2017	26. Based on your customs and ancestry, do you consider yourself to be?
Uruguay	2011	Do you think you have ancestry (of the following groups)?
Territoriality		
Colombia	2018	Q. 13. Is the household inside an ethnic territory?
Venezuela	2011	Indigenous community
Cultural-Linguistic		
Bolivia	2012	P. 31 What languages do you speak?
Guatemala	2018	P13. What linguistic community do you belong? P14. Do you usually wear mayan, Garifuna, afrodescendent or xinka clothing?

Looking at the two most common approaches to measure indigenous identity, it is possible to assess differences in population estimates. Among the countries included, only Paraguay and Bolivia have a higher proportion of indigenous by the linguistic criterion than through self-identification. For countries surveyed in 2018, Peru is the country with the highest proportion of self-identified indigenous (one out of three individuals recognise as such) while in Paraguay only

1.7% would declare herself as indigenous. On the other hand, approximately three out of four individuals in Paraguay speak an indigenous language—the highest proportion among in the region. Ecuador has the lowest proportion among the countries surveyed with 4.8%.

Mexico, besides the other two approaches, also has an official criterion, which is defined by the National Commission on Indigenous Peoples (*Comisión Nacional de los Pueblos Indígenas*, CDI). Under this criterion, an indigenous person is one that inhabits an indigenous household, which is defined as one where the head of household, the spouse or the ascendants of the household speaks an indigenous language (OEI, 2015)

Table 3: Population estimates by criteria.

Country	Year	Self-identification	Linguistic	Official Criteria	Estimated Total Population	Source
Bolivia	2018	24.9%	31.4%	N/A	11,370,090	NHS
Paraguay	2002	1.7%	77.2%	N/A	5,185,297	IPUMS
Mexico	2018	30.5%	5.9%	9.5%	119,561,904	NHS
Peru1	2018	63.8%	21.1%	N/A	31,169,370	NHS
Ecuador	2010	7.0%	4.8%	N/A	14,482,330	IPUMS
Guatemala	2018	37.1%	27.4%	N/A	17,483,697	NHS

Educational indicators also show variations across different criteria. In terms of gross attendance rate, we divide the age ranges from 3 to 5, 6 to 11, 12 to 14 and 15 to 17. These ages correspond to ideal ages to enter a given educational level—pre-school, primary, lower secondary and upper secondary, respectively—in most of the countries analysed (Paraguay, Mexico, Ecuador) or are very close, as in the case of Peru (OEI, 2015).

Table 4: Gross school attendance by criteria per country

School Attendance 3 to 5 years old by country					
Country	Year	Self-identification	Linguistic	National Criteria	Source
Bolivia	2018	27.2%	31.7%	N/A	NHS
Paraguay	2002	11.1%	48.5%	N/A	IPUMS
Mexico	2018	75.9%	67.4%	73.3%	NHS
Peru	2018	N/A	61.0%	N/A	NHS
Ecuador	2010	30.0%	30.5%	N/A	IPUMS
Guatemala	2018	N/A	N/A	N/A	NHS
School Attendance 6 to 11 years old by country					
Country	Year	Self-identification	Linguistic	National Criteria	Source
Bolivia	2018	89.8%	86.7%	N/A	NHS
Paraguay	2002	61.4%	91.1%	N/A	IPUMS
Mexico	2018	98.3%	96.5%	97.8%	NHS
Peru	2018	N/A	72.0%	N/A	NHS
Ecuador	2010	95.9%	95.6%	N/A	IPUMS
Guatemala	2018	94.4%	94.3%	N/A	NHS
School Attendance 12 to 14 years old by country					
Country	Year	Self-identification	Linguistic	National Criteria	Source
Bolivia	2018	87.4%	79.2%	N/A	NHS
Paraguay	2002	57.8%	86.3%	N/A	IPUMS
Mexico	2018	91.5%	82.0%	87.7%	NHS
Peru	2018	N/A	70.0%	N/A	NHS
Ecuador	2010	86.2%	86.0%	N/A	IPUMS
Guatemala	2018	80.1%	78.6%	N/A	NHS
School Attendance 15 to 17 years old by country					
Country	Year	Self-identification	Linguistic	National Criteria	Source
Bolivia	2018	79.9%	71.9%	N/A	NHS
Paraguay	2002	24.2%	64.8%	N/A	IPUMS
Mexico	2018	69.9%	50.3%	64.2%	NHS
Peru	2018	61.7%	58.6%	N/A	NHS
Ecuador	2010	67.2%	66.2%	N/A	IPUMS
Guatemala	2018	46.9%	41.5%	N/A	NHS

Most of the countries surveyed show that a higher proportion of self-identified indigenous go to schools at each given school age. The differences are in Ecuador, where the proportions are practically the same and Paraguay, where speakers of an indigenous language fare much better.

In terms of illiteracy rate, except Bolivia and Paraguay, speakers of an indigenous language have a higher proportion of illiterate people aged 15 years or older. These rates have a wide range with the highest is Guatemala in 2018, where approximately 1 out of 3 indigenous speakers did not know how to read and write. In terms of self-identification, Paraguay in 2002 reported that almost half of the self-identified indigenous was illiterate.

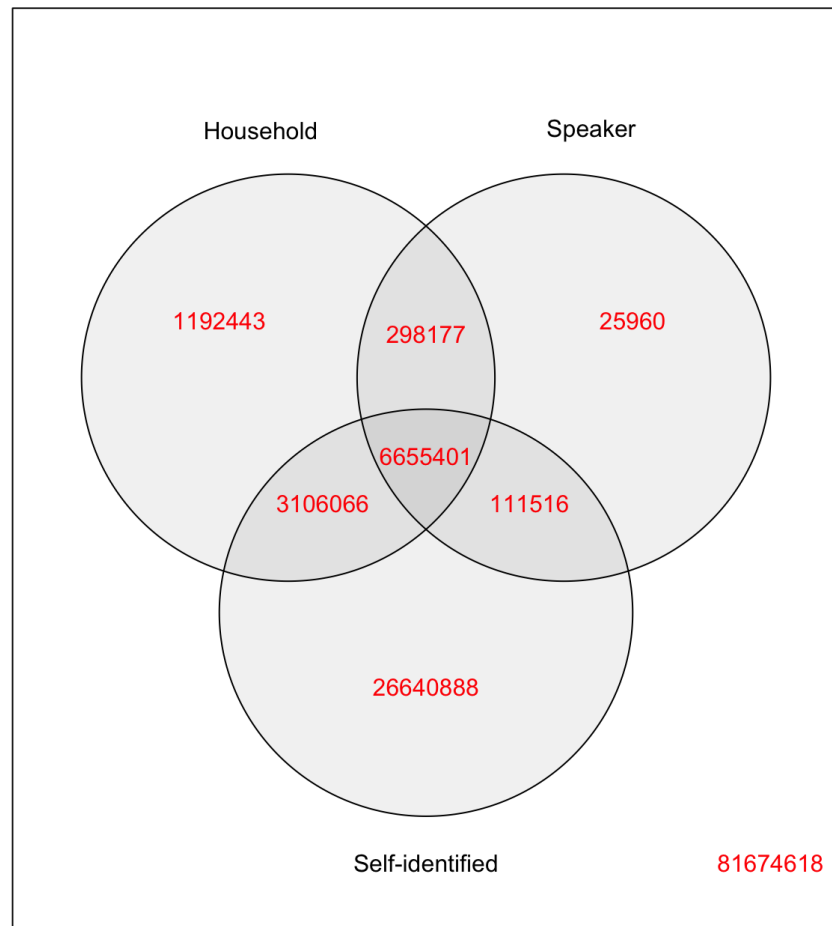
Table 5: Illiteracy Rate: Population 15 years or older by country

Country	Year	Self-identification	Linguistic	National Criteria	Source
Bolivia	2018	14.7%	11.8%	N/A	NHS
Paraguay	2002	44.6%	4.8%	N/A	IPUMS
Mexico	2018	9.6%	23.5%	17.9%	NHS
Peru	2018	5.77%	15.5%	N/A	IPUMS
Ecuador	2010	20.6%	23.6%	N/A	IPUMS
Guatemala	2018	28.6%	31.8%	N/A	NHS

4 Second analysis: Leveraging Differential Classifications of Indigenous Identity

In this section, I conduct multivariate analyses to look at the extent to which different criteria predict educational attainment. The country of analysis is Mexico. In 2018, in Mexico, 30.5 percent of the population aged 3 or older self-identified as indigenous, an increase of nine percentage points compared to 2015, whereas 5.9 percent mentioned they spoke an indigenous language—a stable percentage compared to 6.2 percent in 2015. Considering the intersections, we find that the majority of those that self-identify as indigenous do not speak an indigenous language (81.5 percent) whereas most of those that speak an indigenous language do consider themselves as indigenous (95.4 percent).

Figure 1: Estimated number of individuals classified as indigenous per criteria and corresponding combinations



Note 1: Data from INEGI (2018) *Encuesta Nacional de Ingresos en los Hogares 2018*

Note 2: People aged 3 years or older

Note 3: Number outside the Venn diagram represents individuals not identified as indigenous under any criteria. Household=The CDI's Indigenous criteria; Speaker=If the individual speaks an indigenous language; and Self-identified=if the individual declares to be indigenous.

In Mexico, an estimated 5.6% of individuals would be considered indigenous under the three criteria. 2.6 % of individuals would consider themselves indigenous and be considered as such under the household criteria. 22.3% are indigenous who do not speak an indigenous language nor live in a household where the head of household, partner of the head or any of their ascendants speaks an indigenous language. Finally, 1.0 % of individuals would be considered indigenous per the household criteria but would not identify as such and would not be able to speak an indigenous language. All of these show potential typologies of what constitutes an indigenous person (Figure 1).

Table 6: Mexico: Comparison of speaker of indigenous language and self-identification by type of household (2018)

Belongs to an indigenous household			
Speaker of Indigenous Language	Indigenous (Self-identification)		
	Non-indigenous	Indigenous	Total
Does not speak an indigenous language	1,192,443	3,106,066	4,298,509
Speaks an indigenous language	298,177	6,655,401	6,953,578
Total	1,490,620	9,761,467	11,252,087
Does not belong to an indigenous household			
Speaker of Indigenous Language	Indigenous (Self-identification)		
	Non-indigenous	Indigenous	Total
Does not speak an indigenous language	81,674,618	26,640,888	108,315,506
Speaks an indigenous language	25,960	111,516	137,476
Total	81,700,578	26,752,404	108,452,982
Total			
Speaker of Indigenous Language	Indigenous (Self-identification)		
	Non-indigenous	Indigenous	Total
Does not speak an indigenous language	82,867,061	29,746,954	112,614,015
Speaks an indigenous language	324,137	6,766,917	7,091,054
Total	83,191,198	36,513,871	119,705,069

Source: INEGI (2018) *Encuesta Nacional de Ingreso y Gastos en los Hogares, ENIGH, 2018*.

4.0.1 Dependent variable

To conduct our inferential analyses, we use educational attainment as our dependent variable and create a binary indicator where the reference category is those who have not completed upper secondary. We selected upper secondary in light of recent reforms that make this educational level mandatory and expected to be universal by 2024. We select people aged 25 years or older as they would be expected to enter the labour market and unlikely to return to schooling, which implies that we are looking at individuals at the end of their educational career. People in this age are also likely to live away from their parents and have developed at this age a sense of identity (Villarreal, 2014). In this part, consistent with previous studies, we treat indigenous identity as fixed.

Using Sapperstein, Kizer & Penner's approach (2016), we find that the use of different criteria offers a more nuanced understanding of ethnic inequalities in Latin America, particularly in Mexico—one of the countries with the largest population of indigenous groups. In this method, we conduct logistic regressions and compare individuals who are speakers of an indigenous language, but do not identify as indigenous against self-identified indigenous that do not speak an indigenous language (table 7)¹³. The size of the coefficients should be interpreted as gross inequalities before accounting for any confounding factors¹⁴.

The results show that indigenous who do not speak an indigenous language have higher odds of attaining upper secondary compared to those that speak a similar language but failed to recognise themselves as indigenous. This provides suggestive evidence that language is a stronger predictor of inequality in Mexico than self-identification.

¹³We obtained the same results from a logistic regression by adding two additional groups of comparison: individuals who do not identify as indigenous by any criterion and those that are indigenous and speak an indigenous language.

¹⁴Table presents odds-ratios controlling for age. We also run analyses where we did not control for age and results were very similar (0.374 for the first group and 2.673 for the second group. Both results remained statistically significant at $p < 0.001$).

Table 7: Leveraging Differential Classification in Logistic Regression Models Predicting Upper Secondary Educational Attainment Among 25 years and older

	Restricted to Non-indigenous (self-identified)	Restricted to Non-speakers of indigenous language
Speaker of an indigenous language	0.405*** (0.024)	
Self-identified as indigenous		2.470*** (0.015)
N	685	64,312

I also run multinomial logistic regressions where I account for the sampling design of ENIGH, which uses a probabilistic two-stage sampling design that is nationally representative. I achieve this by using the command `svy` in Stata 16, which reports standard errors clustered at household level. Strata with single sampling primary units were treated as certainty units. To correct for the correlation between individuals living in the same household, I estimate standard errors clustered at the household level.

In the models, I include separately the three different criteria used in the country to identify indigenous peoples. The first is about self-identification; the second is whether the individual belonged to an indigenous household and the third is language. We distinguish between Maya and Nahuatl speakers as they are the two largest linguistic groups in Mexico with a third category grouping the rest of the languages. The reference category is those that do not declare to speak an indigenous language¹⁵.

Among the control variables, I created an ordinal variable for size of location with the reference being rural area (smaller than 2,500 inhabitants), suburban (larger than 2,500 but smaller than 100,000 inhabitants) and urban (larger than 100,000 inhabitants). I also use age as a continuous variable and sex of the individual as a binary response.

Finally, I control for wealth by creating an index following the methodology of the Multiple Indicator Cluster (MICS) for Mexico in 2015. The variable is ordinal and reflects the quintiles of wealth in the household. Wealth reflects intergenerational transitions and earned income (Asadullah, 2012), as such we use it as a proxy for class origins since ENIGH does not collect occupation of the individuals retrospectively, which means that it is impossible to identify the occupational history of the parents of the sampled individuals.

There are two limitations of this analysis. First, it does not include a variable of skin colour. As mentioned before, skin colour has become an important characteristic to assess socioeconomic inequalities in the region, but thus far no country has incorporated this dimension into their official national instruments. Second, I do not correct for potential endogeneity bias in this analysis, which can potentially affect the estimates of self-identification since as mentioned before, education may reinforce a sense of *mestizaje*, but similarly structural inequalities and discrimination may prevent an individual from accessing education. The results in this section are therefore predictive and should not be interpreted as causal evidence.

Given that the official and linguistic criteria are estimated similarly, we check the variance inflation factors (VIF) of independent variables in these models. The values of the VIF indicate that multicollinearity is not a problem.

¹⁵an alternative model categorises between monolingual and bilingual indigenous speakers, but fit statistics suggested the model was inadequate, perhaps due to the low counts of monolinguals attaining upper secondary education. Results of this model show that monolinguals, even after controlling for current wealth of the household are amongst the most disadvantaged groups

Table 8: Descriptive Statistics

Variables	Percentage (Mean)
Self-identification	
Indigenous	30.4
Non-Indigenous	69.6
Language spoken	
Non speaker of indigenous language	93.36
Maya	1.04
Nahuatl	1.42
Other indigenous language	4.18
Indigenous Household	
Indigenous	8.7
Non-indigenous	91.3
Size of location	
Rural	22.80
Semi-urban	28.58
Urban	48.61
Quintile of Wealth	
1	18.46
2	19.92
3	19.27
4	20.33
5	20.33
Age	46.9
Female	53.1
Male	46.9
N	150,018

Source: INEGI (2018) *Encuesta Nacional de Ingreso y Gastos en los Hogares, ENIGH, 2018*.

Wealth quintiles estimated using the entire sample.

Sample consists of individuals aged 3 or older

In the table, the weighted estimates show that in the population, 3 out of 10 individuals aged 25 or older consider themselves as indigenous, but less than 10% of them speak an indigenous language. Nahuatl speakers constitute 1.4% of the total individuals aged 3 or older with Maya speakers sharing a similar proportion. Most of the individuals reside in urban areas followed by semi-urban and rural. The mean age is 47 years old with the majority of the individuals being female.

4.0.2 Inferential results

Table 9 shows the result of binary logistic regressions models with coefficients representing odds ratios predicting upper secondary attainment¹⁶. As expected from the descriptive analyses in section 1, self-identification, speaking an indigenous language and living in an indigenous household are associated with lower likelihood of educational attainment compare to their references groups. But we should expect that those who are speakers of an indigenous language fare much worse than self-identified indigenous. In the model, these associations remain negative when they are included alone (Model 1 to 3) and when they are jointly considered, controlling for location size (Model 4). When controlling for other variables, there are heterogenous effects depending on language as the relationship of speaking Maya is not statistically significant compared to the reference group (non-indigenous language speakers). On the other hand, Nahuatl and other speakers of indigenous languages have lower odds of completing upper secondary compared to those who do not speak an indigenous language, controlling for other variables (Model 6).

¹⁶Multilevel binary logistic analysis accounting for clustering at the household level provide estimates consistent with the findings in this model

Table 9: Coefficients (Odds Ratios) of Binary Logistic Regression Models Predicting Educational Attainment.

	(1)	(2)	(3)	(4)	(5)	(6)
	Completed Upper Secondary	Completed Upper Secondary	Completed Upper Secondary	Completed Upper Secondary	Completed Upper Secondary	Completed Upper Secondary
Completed Upper Secondary Non-indigenous	1 [1,1]			1 [1,1]	1 [1,1]	1 [1,1]
Indigenous	0.425*** [0.366,0.494]			0.633*** [0.573,0.698]	0.755*** [0.697,0.818]	0.916* [0.846,0.992]
Male	1 [1,1]	1 [1,1]	1 [1,1]	1 [1,1]	1 [1,1]	1 [1,1]
Female	0.712*** [0.670,0.757]	0.691*** [0.640,0.747]	0.691*** [0.639,0.747]	0.687*** [0.635,0.743]	0.662*** [0.608,0.720]	0.643*** [0.589,0.702]
Non-indigenous speaker		1 [1,1]		1 [1,1]	1 [1,1]	1 [1,1]
Maya		0.371*** [0.281,0.489]		0.893 [0.688,1.159]	0.859 [0.645,1.144]	0.776 [0.601,1.003]
Nahuatl		0.151*** [0.0990,0.229]		0.358*** [0.234,0.549]	0.430*** [0.290,0.637]	0.512*** [0.347,0.754]
Other indigenous language		0.141*** [0.107,0.184]		0.339*** [0.268,0.430]	0.394*** [0.311,0.500]	0.469*** [0.376,0.585]
Age		0.936*** [0.933,0.940]	0.935*** [0.932,0.939]	0.936*** [0.932,0.939]	0.935*** [0.932,0.938]	0.933*** [0.930,0.936]
Non-indigenous household			1 [1,1]	1 [1,1]	1 [1,1]	1 [1,1]
Indigenous Household			0.185*** [0.150,0.230]	0.549*** [0.459,0.657]	0.622*** [0.525,0.737]	0.707*** [0.592,0.844]
Rural					1 [1,1]	1 [1,1]
Semiurban					1.767*** [1.427,2.188]	1.174 [0.995,1.386]
Urban					3.740*** [2.861,4.890]	1.484*** [1.258,1.750]
5 quintiles=1						1 [1,1]
5 quintiles=2						1.952*** [1.795,2.123]
5 quintiles=3						3.405*** [3.031,3.825]
5 quintiles=4						6.635*** [5.751,7.654]
5 quintiles=5						16.79*** [14.51,19.43]
Observations	150018	150018	150018	150018	150018	150018
Goodness-of-fit	p=0.281	p<0.001	p<0.001	p<0.001	p<0.001	p=0.115

Exponentiated coefficients; 95% confidence intervals in brackets. Standard errors are clustered using svy command.

Strata with single PSU were treated as certainty units. Goodness-of-fit test fit for survey design (Archer & Lemeshow, 2006)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The effect of self-identification, while still statistically significant, is low compared to indigenous language speakers and individuals identified under the household criteria (Model 6). This finding suggests that self-identification can be largely explained by the wealth of the household, in contrast to the the effect of language and the household criteria where the effects still remain large. In that same model, indigenous household is statistically significant even after controlling for the individual's language condition. This finding suggests that there is potentially a dimension unexplained by language or self-identification.

Consistent with previous studies (Flores & Telles, 2012; Villarreal, 2014), the relationship between socioeconomic conditions and educational attainment is strong. Individuals from the 5th quintile have approximately 17 times the odds of attaining upper secondary than one from the first quintile, after controlling for other variables.

The results from these models suggest that educational inequalities between the criteria are the result of different processes: promoting economic well-being may increase educational attainment of self-identified indigenous peoples. In the case of speakers of indigenous language wealth deprivation contributes to low educational attainment but it does not explain all the observed differences between the groups. In terms of research, one implication of the analysis is that by just looking at self-identification, one could reinforce the narrative that ethnic differences are the result of class disparities and ignore other determinants required to leverage the different groups' socioeconomic conditions.

5 Third Analysis: Ethnic Fluidity in Mexico.

5.1 Changes of self-identified ethnic identity across time

We now turn to the ENNVIIH survey (*Encuesta Nacional sobre Niveles de Vida en los Hogares*) to identify and describe patterns of ethnic fluidity in Mexico. In this section, I aim to measure the extent to which ethnicity remains fixed for individuals aged 15 or older, a central assumption of studies that identify inequality between indigenous and non-indigenous peoples.

The ENNVIIH is a nationally representative survey of the Mexican population and the data can be disaggregated at the country, urban, rural and regional level. The survey follows individuals across three waves (2002, 2005-2006 and 2009-2012)¹⁷. Due to the extensive data collection efforts, ENNVIIH has a high response rate: approximately 90% of the original household sample was recontacted successfully (Rubalcava & Teruel, 2007). ENNVIIH collects information through face-to-face interviews on a wide range of variables, which include demographics, income, migration, health and violence. If an individual was not present at the time of the interview, a main informant provided some basic details about him or her and the data was collected on a proxy book.

While the ENNVIIH was not designed to measure ethnic mobility, it can provide useful information to describe this phenomenon in Mexico. First, the survey allows the same individuals to repeatedly declare their indigenous identity over several years. Second, the consistent sampling design and wording of the question indicates that we can rule out that changes in indigenous self-identification are due to these types of measurement error. Finally, ENNVIIH, unlike other household surveys or the census, does not rely on a main informant to obtain individual characteristics of the members of the household. Instead, individuals themselves aged 15 or over are asked about their ethnic identification.

To describe indigenous ethnic fluidity, we will proceed first by looking at the number of changes among individuals based on their original indigenous identification. Second, using only individuals present during the three waves, we identify the main patterns of mobility and finally, we run logistic regressions examining whether individuals report a consistent ethnicity during the first two waves (2002 and 2005). An advantage of using these two periods is that we can observe the immediate effects of certain socioeconomic changes on indigenous self-identification (Saperstein & Penner, 2012).

5.2 Sample selection

We restrict the initial sample to those individuals that answered the question "Do you recognise yourself as member of an indigenous group?" in 2002. We follow them through the other two waves by linking the datasets using the procedure described in Rubalcava & Teruel (2007). In this process, we merged the datasets by the `pid_link` variable. To estimate the ethnic transitions over time, we distinguished between item non-response and survey non-response. If an individual had information on the proxy booklet, we treated the ethnicity variable as item non-response. However, if there was no matching code during a specific wave, it was treated as survey non-response. In this category, we also include deaths and refusals. The total sample size for 2002 is 19,756. The flows and inflows for each individual in the sample are shown in Figure 1. Weighted percentages and counts are derived using the 2002 weight from the education booklet, which accounts for that section's non-response. ^f

We identify a considerable degree of fluidity between waves: 50 percent of the respondents who declared themselves indigenous in the first wave changed their ethnic identity at least once during the period of data collection (2002-2009). Even if we incorporate the information from those that dropped out of the first wave, the change is still considerable for the individuals that identified as indigenous.

Those that identify as non-indigenous had a lower fluidity rate: 7.7% declared to change their ethnicity at least once. This change cannot be attributed to measurement error from the survey collector since other categories, such as sex, had a 0.19% transition rate, consistent with analyses of racial mobility in the United States (Saperstein & Penner, 2012).

¹⁷Additional waves of ENNVIIH will be collected until 2022

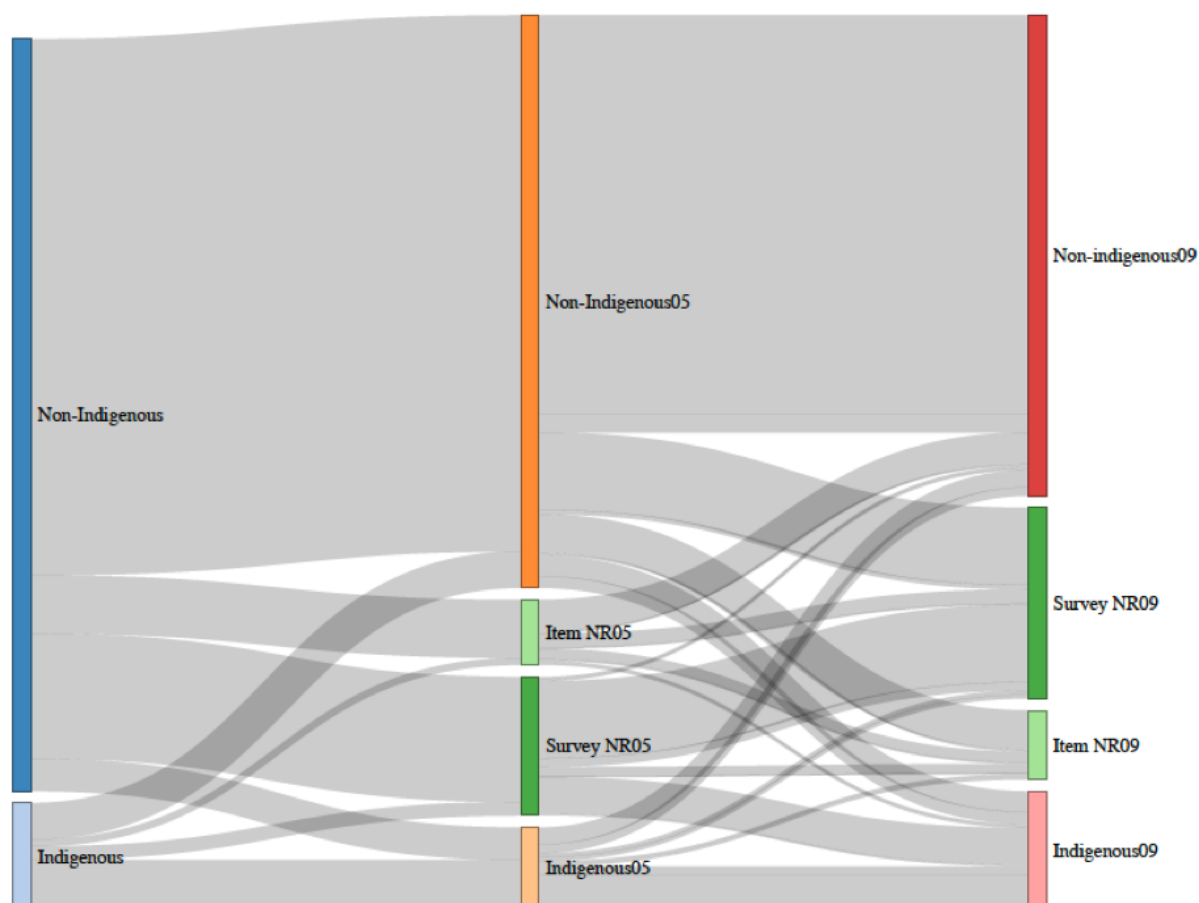


Figure 2: Ethnic Transition for the Population 15 years or older (2002-2012)

Table 10: Ethnic Transitions Across Waves (15 year or older)

Number of changes in self-perceived ethnicity 2002-2009					
Unweighted					
Origin (2002)	0	1	2	Missing	Total
Indigenous	35.8%	35.0%	17.6%	11.6%	2,246
Non-Indigenous	77.6%	5.4%	2.3%	14.8%	17,510
Total	72.4%	9.0%	4.2%	14.4%	100.0%
N	14,310	1,784	820	2,842	19,756
Changes in self-perceived ethnicity					
Weighted					
Origin (2002)	0	1	2	Missing	Total
Indigenous	39.5%	32.5%	15.0%	13.1%	7,337,124
Non-Indigenous	75.0%	5.4%	2.5%	17.1%	60,950,507
Total	71.2%	8.3%	3.9%	16.7%	100.0%
N	48,595,719	5,662,922	2,632,126	11,396,864	68,287,631

Source: Data from MxFLS 2002, MxFLS 2005 & MxFLS 2009.

To clarify the patterns of change, 32.5% of those that self-identified as indigenous in 2002 changed their ethnic identification once during the three waves. The patterns could be IMN, INM, and INN, where I is indigenous, M is missing and N as non-indigenous. In other words, if the respondents provided their ethnic identification in at least any two years, then these changes were included in the table. Due to survey non-response and item non-response, it is possible that the change of ethnicity is underestimated.

We will focus on those aged 15 to 24 years old to assess the extent to which completing basic education (or lower secondary level) related to changes in indigenous identification. Since the group is in schooling age, we can obtain evidence of the immediate relationship of attaining an educational level on self-perceived ethnicity.

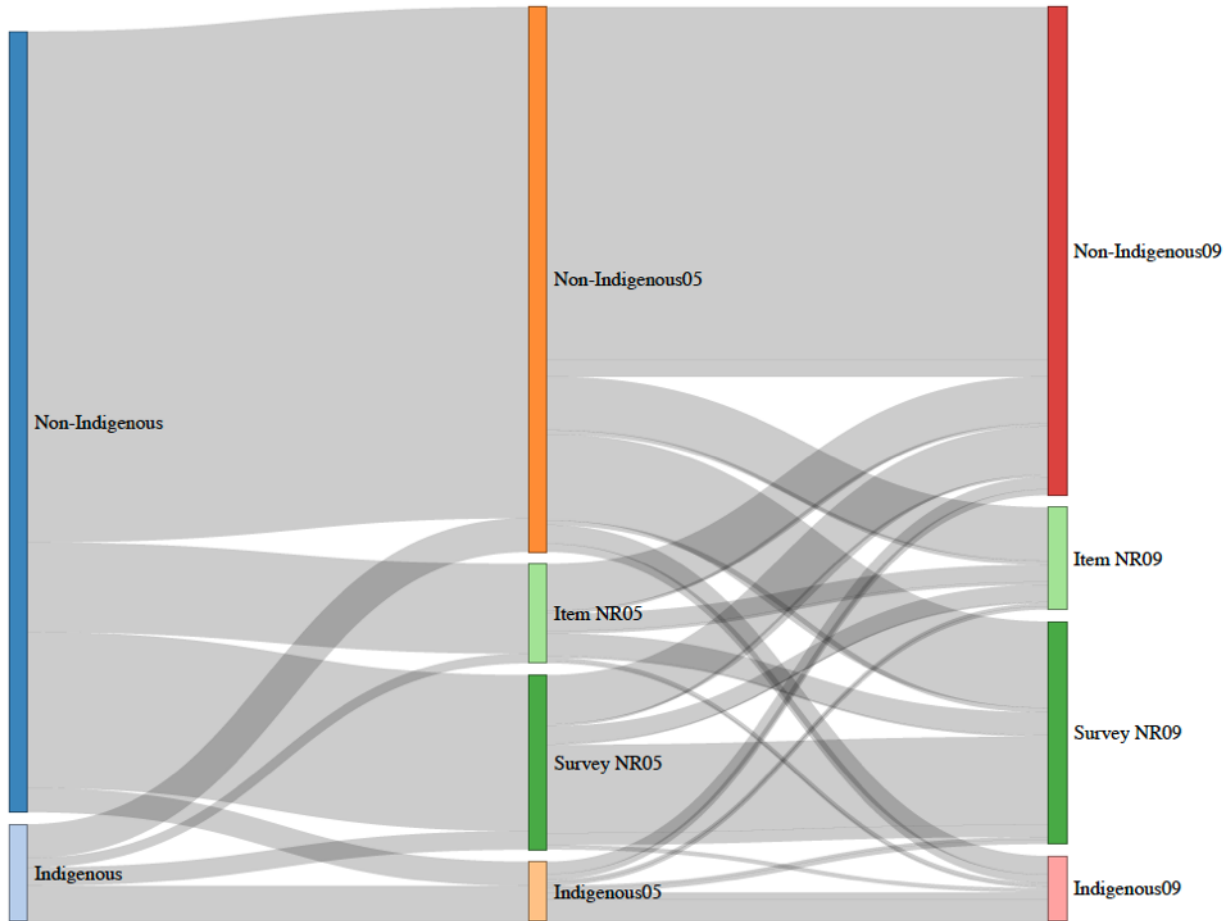


Figure 3: Ethnic Transition for the Population 15 to 24 years(2002-2012)

Consistent with the results from the previous group, self-declared indigenous youth in 2002 had high transition rates—approximately 6 out of 10 reported a different ethnicity at least once during the three waves. However, non-indigenous youth were more consistent in their ethnic origin—approximately 7.3% of them changed ethnicity during the three waves, even after accounting for the missing individuals¹⁸

¹⁸An analysis of missing values (not included here) shows that most of those that remained in the sample do not significantly differ from those that left it entirely except in gender and size location (attrition rates for males and city dwellers in 2002 were higher).

Table 11: Ethnic Transitions Across Waves (15 to 24 years old)

Changes in self-perceived ethnicity 2002-2009					
Unweighted					
Origin (2002)	0	1	2	Missing	Total
Indigenous	28.2%	39.4%	17.1%	15.3%	528
Non-Indigenous	80.9%	5.0%	1.7%	12.4%	3,941
Total	74.7%	9.1%	3.5%	12.8%	100.0%
N	3,337	405	156	571	4,469
Changes in self-perceived ethnicity					
Weighted					
Origin (2002)	0	1	2	Missing	Total
Indigenous	30.3%	36.5%	17.4%	15.8%	1,595,440
Non-Indigenous	80.6%	5.2%	2.1%	12.1%	13,297,511
Total	75.2%	8.6%	3.7%	12.5%	100.0%
N	11,199,744	1,273,964	556,858	1,862,385	14,892,951

Source: Data from MxFLS 2002, MxFLS 2005 & MxFLS 2009.

There are some limitations to these comparisons. First, it is possible that changes are due to some types of measurement error—the respondent may have not understood the question, or the interviewer recorded the answer incorrectly. But if that were the case, we would expect little association between relevant variables (Saperstein & Penner, 2012). Moreover, there is already some evidence that indigenous individuals do not necessarily identify their own children with the same ethnicity as theirs, reflecting that there is some fluidity across generations that may affect how individuals also identify in the long-term over the course of their life (Villarreal, 2014).

The survey also fails to collect information on the interviewer, which may influence the respondent's responses. However, research in the region about the influence of the interviewer has been shown to matter for skin colour and not necessarily ethnic self-identification (Cernat, Sakshaug, & Castillo, 2019).

Another limitation is that the survey does not distinguish language as a separate criterion, but rather as a subset of self-identification. This limits the potential discordance between language and self-identification that could compare other dimensions of indigenous identity. It also does not allow me to test the extent to which speakers of indigenous languages do not identify as indigenous¹⁹. Due to these constraints, the analyses presented in this section remain exploratory.

It is also possible that most of these changes are short term and arguably, any real change on how an individual identifies require that they are long lasting (Statistics New Zealand, 2005). Thus, we estimate the proportion of individuals that only shift once their declared ethnicity during the three waves. Among those for which we have full information during the three waves, 5.8% of those declared a different ethnicity in the second wave and did not change their indigenous (or non-indigenous) identification subsequently. 5.7% of the individuals also declared a consistent ethnicity between wave 1 and 2, but changed it in wave 3. These proportions should be interpreted with caution as they are a lower threshold of potential long term changes in self-identified indigenous identity, considering that individuals with missing information were not included.

¹⁹Given the way indigenous language is collected in ENNVIIH, the number of speakers of indigenous languages is underestimated

Table 12: Ethnic Transitions Across Waves 15 years or older. Complete cases sample

2002	2005	2009	Unweighted persons	Percentage	Weighted Total	Weighted Percentage
Non-Indigenous	Non-Indigenous	Non-indigenous	9,197	77.1%	29,828,803	77.6%
Indigenous	Indigenous	Indigenous	777	6.5%	2,540,716	6.6%
Non-Indigenous	Non-Indigenous	Indigenous	475	4.0%	1,560,834	4.1%
Indigenous	Non-Indigenous	Non-indigenous	408	3.4%	1,095,477	2.9%
Non-Indigenous	Indigenous	Non-indigenous	390	3.3%	1,534,983	4.0%
Indigenous	Non-Indigenous	Indigenous	271	2.3%	637,702	1.7%
Non-Indigenous	Indigenous	Indigenous	211	1.8%	717,849	1.9%
Indigenous	Indigenous	Non-indigenous	197	1.7%	499,878	1.3%
Total sample (2002)			11,926	100.0%	38,416,242	100.0%

To assess whether changes in ethnic identification in the survey affect educational indicators, I estimate selected indicators accounting for the indigenous fluidity. Concretely, I reclassified indigenous individuals who declared to be non-indigenous in 2005 as well as non-indigenous individuals that declared to be indigenous in 2005. These changes were made for individuals in 2002 to control for time variant factors (such as entering adult learning during after the 2002 survey was collected). Consistent with official statistics, the selected indicators are illiteracy rates for people 15 years older and completed basic education, or lower secondary, for those aged 25 and 64 years old.

I find that among the group that remained in the sample, change of ethnic identification accounts for approximately two percentage points of the difference over time. In other words, if all individuals in 2005 had remained in the same ethnic group they declared in 2002, the differences in the indicators would be approximately of 1.6 percentage points for indigenous peoples. The change is less pronounced among non-indigenous.

Table 13: Table 11: Effect of ethnic transitions on selected indicators (2002)

Changes of illiteracy rate 15 years and older by criteria			
Group	2002	Accounting for ethnicity changes	Estimated Difference
Indigenous people	27.7%	26.2%	-1.6%
Non-indigenous	8.78%	8.75%	-0.03%
National	10.94%	10.94%	0.0%
Changes of basic education attainment for 25 to 64 year old by criteria			
Indigenous people	28.30%	26.49%	-1.8%
Non-indigenous	47.40%	47.90%	+0.5%
National	45.2%	45.20%	0.0%
Estimated Population Changes by criteria (15 years and older)			
Indigenous people	5,731,025	6,344,367	+10.0%
Non-indigenous	44,749,855	44,136,513	-1.3%
National	50,480,880	50,480,880	0.0%

Source: Data from MxFLS 2002 & MxFLS 2005.

Sample includes those that reported ethnicity in both 2002 and 2005.

Rates of attrition were 19.4% for indigenous and 24.3% for non-indigenous.

Estimates are weighted using the 2002 factor.

In terms of population, the entire shift represented a change of 613,342 individuals that accounted for an increase of 10% for indigenous peoples and a decrease of 1.4% for non-indigenous individuals.

To test how different socioeconomic changes relate to an individual's sense of ethnic identity, we modelled a multinomial logistic analysis. Our dependent variable is categorical where the reference is those indigenous that are consistent. Consistent refers to individuals who mentioned they belonged to the same ethnicity during waves 1 and 2. Inconsistent are those that changed during these two waves and a missing category for individuals who appeared in subsequent waves but had failed to respond the ethnicity question during wave 2 and had information on the proxy book. Individuals with survey non-response were excluded from the analysis. We use first two waves since we are interested in looking at individuals who were of school age during at least one additional wave to test the potential immediate

relationship between education and probability of reporting consistently their ethnicity. We restrict our sample among indigenous youth who declared their ethnicity in the first wave, which results in 499 individuals and represent 81% of the total indigenous youth.

Table 14: Descriptive statistics youth (15 to 24 years old).

Variable	Indigenous Consistent	Indigenous Inconsistent	Non-Indigenous inconsistent	Non-indigenous consistent
Sex (Female)	69.8%	60.5%	56.2%	58.3%
Age in 2002 (mean)	18.58	18.65	18.98	19.06
Working 2002	23.8%	31.7%	35.0%	38.5%
Working 2005	36.4%	44.0%	44.8%	42.2%
Attending school (2002)	33.1%	41.8%	41.9%	37.0%
Attending school (2005)	12.5%	17.9%	19.2%	21.5%
Basic education (2002)	40.4%	68.9%	57.2%	69.3%
Basic education (2005)	47.1%	81.5%	70.3%	76.7%
Migrated in the last two years (2002)	3.9%	4.6%	5.3%	4.7%
Migrated in the last two years (2005)	6.1%	1.4%	5.7%	4.4%
Rural location at the time of interview (2002)	79.7%	32.8%	20.2%	21.1%
Lives under the poverty line (2002)	91.2%	79.6%	71.9%	69.7%
Lives under the poverty line (2005)	80.5%	77.7%	75.2%	65.4%
N (unweighted)	226	213	153	3,208

Source: Data from MxFLS 2002 & MxFLS 2005.

Note 1: Results are unweighted.

5.3 Inferential results

Table 15: Descriptive statistics youth (15 to 24 years old).

Variable	Inconsistent indigenous	Consistent indigenous	Missing
Ever lived in a city	44.6%	16.4%	51.7%
Ever completed lower secondary	61.5%	40.3%	51.7%
Ever under poverty line	93.9%	96.5%	91.7%
Female	45.1%	38.1%	58.3%
Age	18.8	18.7	18.8
Indigenous speaker	30.0%	80.5%	53.3%
Total	213	226	60

Our independent variables are as follows. We use a proxy variable to measure increases in education by identifying whether the individual finished lower (or basic) secondary education at some point during the two waves. Higher educational attainment has been documented to revitalise interest and pride in ethnic origins (Doyle & Kao, 2007; Villarreal, 2014). However, for the case of Mexico, we hypothesise that indigenous students that get more schooling are likelier to change their ethnicity as they are immersed in the mestizo ideology prevailing at the time. During the first wave of data collection, the Mexican government aimed to transition from an “incongruent model” that focus solely on teaching on indigenous languages for a particular group (farmworkers) to a more comprehensive approach (Martínez Buenabad, 2015). Before 2003, those changes focused exclusively on elementary education. As a result, individuals in the sample were exposed to the traditional model that limitedly recognised the multicultural background of the country when they studied in lower secondary (Martínez Buenabad, 2015).

We include sex of the child since gender norms may relate to who carries a community’s traditions and is therefore, likely to develop a stronger sense of indigenous identity (Villarreal, 2014). A child’s age may reflect whether they move away from their home, and as a result of their communities’ culture and traditions (Villarreal, 2014). Additional controls are whether the individual is a speaker of an indigenous languages, as they tend to have a stronger sense of ethnic identification (?). Research has also found that size of locality matters as rural inhabitants are likelier to identify as indigenous than urban dwellers (de la Cadena, 2000). Finally, we measure socioeconomic status by looking at whether individuals ever lived in a household under the poverty threshold estimated by the National Council

for Evaluation (CONEVAL, in Spanish). We constructed the poverty line by estimating a weighted monthly average of the preceding year of the wave (as the survey included income from the previous year that contributed to the household expenses). Estimations of this line vary between rural and urban locations. To account for the potential correlation in the data, we estimate standard errors clustered by household.

Table 16: Coefficients (Odds-Ratio) of Multinomial Logistic Regression Models Predicting Consistency of Self-identification.

	(1) Consistency of indigenous self-identification
Inconsistent_indigenous	
Consistent_indigenous	
Ever lived in a city=0	1 [1,1]
Ever lived in a city=1	0.432** [0.247,0.756]
Ever completed basic education	0.841 [0.531,1.331]
Ever been under poverty line	0.768 [0.251,2.346]
Female	0.762 [0.488,1.191]
Age	0.955 [0.884,1.031]
Indigenous Language Speaker	7.716*** [4.714,12.63]
Missing Response	
Ever lived in a city=0	1 [1,1]
Ever lived in a city=1	1.876 [0.980,3.591]
Ever completed basic education	0.819 [0.443,1.515]
Ever been under poverty line	0.588 [0.200,1.733]
Female	1.724 [0.947,3.139]
Age	0.980 [0.884,1.088]
Indigenous Language Speaker	3.152*** [1.623,6.120]
Observations	499
Hosmer-Lemeshow test of goodness-of-fit	p<0.632
Exponentiated coefficients; 95% confidence intervals in brackets	
Source: ENNVIIH 2002 & ENNVIIH 2005	
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$	

The results show that attaining basic (lower secondary) education during the period of the two waves is not statistically significant. Potentially, this may be because either the effects of education on an individual's decision to change their ethnicity are long term or perhaps they are related as long as they mediate other variables such as occupation. The model, however, does show that being a speaker of an indigenous language is related to reporting a

consistent indigenous identity. These results are consistent with the theory that speakers of indigenous languages may have a stronger sense of indigenous identity than those that do not speak any.

Living in an urban area decreased the likelihood of individuals reporting consistent indigenous identification over time. Potentially, this situation could reflect weaker ties to the communities of origin and a greater adaptation to urban values. If that is the case, intercultural education could continue expanding into urban areas, where limited culturally-pertinent services are offered to indigenous children.(INEE, 2017b).

We also ran a binary logistic regression using the entire sample (age 15 years or older) and applied random effects for the household (see Annex). However, we did not report these results in the main section because we could not include the language variable as ENNVIH only asks for it if the respondents have previously identified as indigenous in a given wave. The results show that living below poverty increase the odds of being classified as indigenous in a subsequent wave while attaining basic education (lower secondary) decreases them, controlling for the other variables. The result of ρ derived from the household random effect indicate a large intra-household correlation of the outcome—members of the same household are likely to identify as indigenous as opposed to selecting persons at random from different households.

6 Conclusions

This paper described the diverse types of criteria that Latin American countries have used to measure indigenous identity. The evidence shows that the region has gradually converged to using self-identification as their central criteria, first in their census instruments and gradually in household surveys. Countries in the Central American area remain a priority since only Guatemala collects data that can provide frequent and reliable estimates of their indigenous peoples.

Despite increased availability of information on indigenous identity, some challenges persist. Particularly, household and international surveys still need to adopt comparable criteria between each other and the census. One way to proceed would be to reach consensus on the wording and the categories to include in each instrument. By doing so, countries can increase the comparability of information and obtain a more comprehensive view of the indigenous peoples living conditions. In this process, countries should consult with indigenous peoples to identify both relevant categories to measure indigenous identity and groups that are currently uncounted. In this paper, I mentioned two where researchers are currently working to estimate demographic information: indigenous migrants and indigenous forest peoples.

While countries should retain self-identification as the main criterion, they can also explore other ways to identify indigenous peoples. Consistent with previous literature, the results presented here are suggestive that self-identification does not completely reflect the extent to which inequalities are prevalent among indigenous peoples in Mexico. Once we account for wealth, the effect of self-identification to predict educational attainment decreases. On the other hand, the household and linguistic criterion still have larger remaining effects.

Another reason to include additional criteria is because indigenous self-identification is fluid—a higher proportion of self-identified indigenous individuals change their ethnicity than non-indigenous individuals. Potentially, these changes can overstate the extent to which policies are successful in improving or worsening the living conditions of different groups. Data presented here shows that while in the first two waves, ethnic transitions can account for approximately 2% of change in illiteracy and educational attainment rates, further research will need to measure the long term effects of these transitions.

Language is the main predictor of consistency of indigenous identification across the two waves, perhaps because they have a stronger sense of identity whereas living in an urban area is associated with lower consistency. One reason individuals in urban areas are more likely to be inconsistent could be that they have weaker links to their communities of origin. If that is the case, intercultural education could continue expanding into urban areas, where limited culturally-pertinent services are offered to indigenous children.

Further research on ethnic fluidity needs to better estimate changes on identity derived from contextual effects. One way to do that is to collect long term longitudinal surveys that also include demographic information about the interviewer and allow for the interaction of different criteria. For example, everyone should answer the question of whether an individual speaks an indigenous language and not only those that initially self-identify as indigenous.

7 Annex: Recommendations

Based on the research, the World Inequality Database on Education (WIDE) that contains educational indicators estimated from datasets used in this paper, could highlight the multidimensional aspects of ethnicity and provide information on how ethnic fluidity can affect their estimates. The following are a set of recommendations:

- One step would be to report the sampling method and the purpose of the survey when reporting ethnicity data and education indicators. Some surveys in Latin America, due to their design, also fail to include certain groups
- Identify the official approaches, and the alternatives, that countries have used to measure ethnicity and report educational indicators. For example, some reports in Mexico and Colombia include information about indigenous municipalities
- Analyse the intersections derived from grouping different criteria, where information is available
- Provide confidence intervals or a coefficient of variation so that readers understand the uncertainty surrounding an estimate. It is possible that some indicators can be reported, but only for larger groups
- Identify the extent to which ethnic fluidity occurs in countries where information is available and the extent to which these can affect potential indicators.
- Consult with countries and indigenous peoples what relevant additional criteria could be helpful to know and monitor through the WIDE dataset.
- Incorporate research, potentially as a linked document in the WIDE database, that has attempted to create demographic portraits of invisible groups not reported in official statistics but who are relevant. For example, indigenous migrants and indigenous forest peoples
- Explore the possibility of linking individual-data with municipal or district level data
- Explore the suitability of including information from longitudinal surveys that add information on educational trajectories of individuals

Table 17: (Annex) Surveys and Census Data with Ethnicity Criteria (2000-2018)

Country	Source of Data	Year	Type of Criteria
Argentina	CSIP	2005	Self-identification and common origin
	Census	2000	Self-identification
	Census	2010	Common origin
Bolivia	DHS	2003-2008	Self-identification and Linguistic
	ENH	2005-2017	Self-identification and Linguistic
	Census	2001	Linguistic
Brazil	Census	2012	Self-identification and Linguistic
	PNAD	2015	Race
Chile	Census	2010	Self-identification and Linguistic
	CASEN	2000-2017	Self-identification and Linguistic
Colombia	Census	2002-2017	Self-identification
	Census	2005	Self-identification and Linguistic
	Census	2018	Territoriality, Self-identification and Linguistic
	ENPH	2006-2007	N/A
Costa Rica	ENPH	2016-2017	Self-identification
	DHS	2010-2015	Self-identification
	MICS	2011	Common origin
Cuba	Census	2000	Self-identification
	Census	2011	Self-identification and Linguistic
Dominican Republic	Census	2012	Race
	N/A	N/A	N/A
Ecuador	Census	2001	Self-identification and Linguistic
	Census	2010	Common origin, Self-identification and Linguistic
	ENIGHU	2004	Common origin, Self-identification and Linguistic
El Salvador	ENIGHU	2012	Self-identification and Linguistic
	MICS	2014	Common origin & Self-identification
Guatemala	Census	2007	Self-identification
	Census	2002	Self-identification and Linguistic
	Census	2018	Self-identification, Linguistic and cultural
	ENEI	2002	Territoriality, Self-identification and Linguistic
Honduras	ENEI	2018	Self-identification and Linguistic
	DHS	2014	Self-identification.
	Census	2001-2013	Self-identification
Haiti	DHS	2011-2012	Self-identification
	N/A	N/A	N/A
Mexico	Census	2000-2010	Self-identification, linguistic, common origin
	Inter-census	2015	Self-identification, linguistic, territoriality, common origin
	MICS	2015	Self-identification, linguistic, common origin
Nicaragua	ENIGH	2008-2016	Self-identification, linguistic, common origin
	DHS	2011	Linguistic
Panama	Census	2005	Self-identification
	MICS	2013	Territoriality, common origin, linguistic & self-identification
	Census	2000-2010	Self-identification & common origin
Paraguay	ENV	2003, 2008	Linguistic
	MICS	2016	Common origin
	Census	2002	Self-identification and Linguistic
	Census	2012	Self-identification, Territoriality and linguistic
Peru	EIGCV	2011-2012	Linguistic
	DHS	2004-2006	Linguistic
	DHS	2007-2014	Self-identification and linguistic
	Census	2007-2017	Self-identification and linguistic
Uruguay	ENHCVP	2003-2009	Linguistic
	ENHCVP	2010-2018	Linguistic, common ancestry, self-identification
	MICS	2012	Common origin
Venezuela	Census	2011	Common origin
	ECH	2006-2018	Common origin
Venezuela	Census	2001	Self-identification
	Census	2011	Territoriality, self-identification, linguistic.

8 Annex: Binary Logistic Regression with Fixed Effects

Table 18: Descriptive statistics

Variable	Person-Years (2002-2005)
Indigenous	12.8%
Completed basic education	41.9%
Lived under poverty line	79.2%
Lived in a city	56.5%
Age	42.4%
Female	41.3%
Total	29,898

Table 19: Coefficients (Odds Ratios) of Binary Logistic Regression Models Predicting Indigenous Identification.

	(1) Indigenous
Indigenous	
Completed basic education	0.416*** [0.352,0.492]
Lived under poverty line	2.548*** [2.106,3.082]
Lived in a city	0.209*** [0.177,0.247]
Age	1.001 [0.997,1.005]
Female	1.167 [0.997,1.367]
/	
lnsig2u	9.221*** [8.340,10.19]
Observations	29898

Exponentiated coefficients; 95% confidence intervals in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

- Alba, R. D. (1990). *Ethnic identity: The transformation of white america*. Yale University Press. Retrieved from <http://www.jstor.org/stable/j.ctt3zbp8p>
- Angosto-Ferrández, L. F., & Kradolfer, S. (2014). The politics of identity in latin american censuses: Introduction to a special issue. *Journal of Iberian and Latin American Research*, 20(3), 321-327. Retrieved from <https://doi.org/10.1080/13260219.2014.995871> doi: 10.1080/13260219.2014.995871
- Aravena Reyes, A. (2014). Identidad indígena en Chile en contexto de migración, urbanización y globalización. *Amérique Latine Histoire et Mémoire. Les Cahiers ALHIM*(27).
- Archer, K. J., & Lemeshow, S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The Stata Journal*, 6(1), 97-105. Retrieved from <https://doi.org/10.1177/1536867X0600600106> doi: 10.1177/1536867X0600600106
- Asadullah, M. N. (2012). Intergenerational wealth mobility in rural bangladesh. *The Journal of Development Studies*, 48(9), 1193-1208. Retrieved from <https://doi.org/10.1080/00220388.2011.646988> doi: 10.1080/00220388.2011.646988
- Brown, P., Callister, P., Carter, K., & Engler, R. (2010). Ethnic mobility: is it important for research and policy analysis? *Policy Quarterly*, 6(3). Retrieved from <https://ojs.victoria.ac.nz/pq/article/view/4343> doi: 10.26686/pq.v6i3.4343
- Bustillo, I., Artecona, R., & Perrotti, D. E. (2018, October). *Inequality and growth in latin america: Achievements and challenges* (Tech. Rep.). The Group of 24 and Friedrich-Ebert-Stiftung New York.
- Camou, M. M., Maubrigades, S., & Thorp, R. (2016). *Gender inequalities and development in latin america during the twentieth century* (1st ed.). Routledge.
- Canedo, A. (2019, 03). Labor market discrimination against indigenous peoples in Mexico: A decomposition analysis of wage differentials. *Iberoamericana: Nordic Journal of Latin American and Caribbean Studies*, 48, 12-27. doi: 10.16993/iberoamericana.433
- Canelas, C., & Gisselquist, R. M. (2018). Human capital, labour market outcomes, and horizontal inequality in Guatemala. *Oxford Development Studies*, 46(3), 378-397. Retrieved from <https://doi.org/10.1080/13600818.2017.1388360> doi: 10.1080/13600818.2017.1388360
- Carrión, J., & Zárate, P. (2010). *Cultura política de la democracia en el Perú, 2010*. Instituto de Estudios Peruanos and Vanderbilt University.
- Carter, K., Hayward, M., Blakely, T., & Shaw, C. (2009, 08). How much and for whom does self-identified ethnicity change over time in New Zealand? results from a longitudinal study. *Soc Policy J N Z*, 36.
- CEPAL. (2014, Noviembre). *Los pueblos indígenas en América Latina. avances en el último decenio y retos pendientes para la garantía de sus derechos* (Tech. Rep.). Santiago: Author.
- CEPAL. (2017). Los censos de la ronda 2020: desafíos ante la agenda 2030 para el desarrollo sostenible, los objetivos de desarrollo sostenible y el consenso de Montevideo sobre población y desarrollo. *Población y Desarrollo*. Retrieved from <https://libproxy.berkeley.edu/login?qurl=http%3a%2f%2fsearch.ebscohost.com%2fflogin.aspx%3fdirect%3dtrue%26db%3deddsrep%26AN%3deddsrep.p.ecr.colo45.42394%26site%3dedds-live>
- Cernat, A., Sakshaug, J. W., & Castillo, J. (2019, 02). The Impact of Interviewer Effects on Skin Color Assessment in a Cross-National Context. *International Journal of Public Opinion Research*, 31(4), 779-793. Retrieved from <https://doi.org/10.1093/ijpor/edy030> doi: 10.1093/ijpor/edy030
- Chirino, A., Almeida García, M., & Morales Torrez, B. (2017). ¿reducción de la auto pertenencia indígena en Bolivia? una aplicación experimental al diseño de cuestionarios. *Revista Compendium: Cuadernos de Economía y Administración*, 4(9).
- Chiswick, B. R., Patrinos, H. A., & Hurst, M. E. (2000). Indigenous language skills and the labor market in a developing economy: Bolivia. *Economic Development and Cultural Change*, 48(2), 349-367. Retrieved from <http://www.jstor.org/stable/10.1086/452462>
- Courtis, C. (2009). Notes on the implementation by Latin American courts of the ILO Convention 169 on indigenous peoples. *SUR*, 10.
- de Alba, I. G. G. (2017). *Poverty, remoteness and social mobility of the indigenous population in Mexico* (Unpublished doctoral dissertation). University of Oxford.
- de la Cadena, M. (2000). *Indigenous mestizos the politics of race and culture in Cuzco, Peru, 1919-1991*. Durham, N.C: Duke University Press.
- de la Cadena, M. (2005). Are "mestizos" hybrids? the conceptual politics of Andean identities. *Journal of Latin American Studies*, 37(2), 259-284. Retrieved from <http://www.jstor.org/stable/3875686>
- Del Popolo, F. (2008). Los pueblos indígenas y afrodescendientes en las fuentes de datos: experiencias en América Latina. *Documentos de Proyectos*. Retrieved from <https://libproxy.berkeley.edu/>

- login?url=http%3a%2f%2fsearch.ebscohost.com%2flogin.aspx%3fdirect%3dtrue%26db%3dedssrep%26AN%3dedssrep.p.e.cr.coloz2.3616%26site%3dedss-live
- Del Popolo, F., Oyarce, A. M., Ribotta, B., & Rodríguez Vignoli, J. (2007, July). *Indigenous peoples and urban settlements: spatial distribution, internal migration and living conditions* (Población y Desarrollo No. 78). Naciones Unidas Comisión Económica para América Latina y el Caribe (CEPAL). Retrieved from <https://ideas.repec.org/p/ecr/colo45/7219.html>
- Doyle, J. M., & Kao, G. (2007). Are racial identities of multiracials stable? changing self-identification among single and multiple race individuals. *Social Psychology Quarterly*, 70(4), 405-423. Retrieved from <https://doi.org/10.1177/019027250707000409> (PMID: 19823596) doi: 10.1177/019027250707000409
- Flores, R., & Telles, E. (2012). Social stratification in Mexico: Disentangling color, ethnicity, and class. *American Sociological Review*, 77(3), 486-494. Retrieved from <https://doi.org/10.1177/0003122412444720> (PMID: 25382861) doi: 10.1177/0003122412444720
- Freire, G., Schwartz Orellana, S. D., Zumaeta Aurazo, M., Costa, D. C., Lundvall, J. M., Viveros Mendoza, M. C., ... Sousa, L. D. C. (2015). *Indigenous Latin America in the twenty-first century: the first decade* (Tech. Rep.). Washington, D.C.: World Bank Group.
- Gomez Murillo, A. R. (2008). *Indígena urbano en América Latina: Etnogénesis en el distrito metropolitano de Quito (Spanish edition)*. Editorial Académica Española.
- Hall, G. H., & Patrinos, H. (Eds.). (2012). *Indigenous peoples, poverty, and development*. Cambridge University Press. Retrieved from <https://EconPapers.repec.org/RePEc:cup:books:9781107020573>
- Howard, R. (2011). The quechua language in the Andes today: Between statistics, the state, and daily life. In P. Heggarty & A. J. Pearce (Eds.), *History and language in the Andes* (pp. 189-213). New York: Palgrave Macmillan US. Retrieved from https://doi.org/10.1057/9780230370579_9 doi: 10.1057/9780230370579_9
- ILO. (1989). *Ilo convention on indigenous and tribal peoples, 1989 (no. 169)*. Author. Retrieved November 25, 2019, from https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_ILO_CODE:C169
- INE. (2012). *Cuestionario censal censo 2002*. online.
- INEE. (2017a). *Breve panorama educativo de la población indígena*. Mexico City: Author.
- INEE. (2017b). *Directrices para mejorar la atención educativa de niñas, niños y adolescentes indígenas*. México: Author.
- INEGI. (2018). *Encuesta nacional de ingresos y gastos de los hogares 2018. enigh. nueva serie. diseño muestralñ*. Author.
- Kiang, L., & Baldelomar, O. A. (2016). Ethnic identity formation. In R. J. Levesque (Ed.), *Encyclopedia of adolescence* (pp. 1-15). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-32132-5_298-2 doi: 10.1007/978-3-319-32132-5_298-2
- Kukutai, T. (2004). The problem of defining an ethnic group for public policy: who is Maori and why does it matter? *Social Policy Journal of New Zealand*(23).
- Lopez, H., & Perry, G. (2008). Inequality in Latin America: determinants and consequences. *Policy Research Working Paper Series*. Retrieved from <https://libproxy.berkeley.edu/login?url=http%3a%2f%2fsearch.ebscohost.com%2flogin.aspx%3fdirect%3dtrue%26db%3dedssrep%26AN%3dedssrep.p.wbk.wbrwps.4504%26site%3dedss-live>
- Lopez, L. E. (2017). Decolonization and bilingual/intercultural education. *Language Policy and Political Issues in Education*, 297. Retrieved from <https://libproxy.berkeley.edu/login?url=http%3a%2f%2fsearch.ebscohost.com%2flogin.aspx%3fdirect%3dtrue%26db%3dedssrep%26AN%3dedssrep.978.3.319.02344.1.24%26site%3dedss-live>
- Macedo, J. (2012, 4). Introduction: The urban divide in Latin America: Challenges and strategies for social inclusion. *Bulletin of Latin American Research*, 31(2), 139-141. Retrieved from <https://doi.org/10.1111/j.1470-9856.2011.00653.x> doi: 10.1111/j.1470-9856.2011.00653.x
- Martínez Buenabad, E. (2015). La educación intercultural y bilingüe (EIB) en México. ¿el camino hacia la construcción de una ciudadanía democrática? *Relaciones Estudios de Historia y Sociedad*, 36(141), 103-131.
- Montgomery, J. D. (2011). The population dynamics of black-white-mulatto racial systems. *American Journal of Sociology*, 117(1), 46-89. Retrieved from <https://doi.org/10.1086/660744> doi: 10.1086/660744
- Nagel, J. (1995). American Indian ethnic renewal: Politics and the resurgence of identity. *American Sociological Review*, 60(6), 947-965. Retrieved from <http://www.jstor.org/stable/2096434>
- OECD. (2018). *Diversity statistics in the OECD: How do OECD countries collect data on ethnic, racial and indigenous identity?* Paris: Author.
- OEI. (2015). *Miradas sobre la educación en Iberoamérica 2015. educación de los pueblos y comunidades indígenas (originarios) y afrodescendientes* (Tech. Rep.). Organización de Estados Iberoamericanos.
- Paredes, C. L. (2018). Multidimensional ethno-racial status in contexts of mestizaje: Ethno-racial stratification in contemporary Peru. *Socius*, 4, 2378023118762002. Retrieved from <https://doi.org/10.1177/2378023118762002>

- doi: 10.1177/2378023118762002
- Parker, S. W., Rubalcava, L., & Teruel, G. (2002, January). *Schooling inequality among the indigenous: A problem of resources or language barriers?* (Research Department Publications No. 3134). Inter-American Development Bank, Research Department. Retrieved from <https://ideas.repec.org/p/idb/wpaper/3134.html>
- Patrinos, H. A. (2000, Jun 01). The cost of discrimination in latin america. *Studies in Comparative International Development*, 35(2), 3–17. Retrieved from <https://doi.org/10.1007/BF02687467> doi: 10.1007/BF02687467
- Perez Cardenas, E. (2018). Participación política de mujeres indígenas en tiempos de la revolución ciudadana. *Alteridades*, 28(55), 61-72.
- Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: a further update. *Education Economics*, 12(2), 111-134. Retrieved from <https://doi.org/10.1080/0964529042000239140> doi: 10.1080/0964529042000239140
- Rodríguez Cruz, M. (2018). Construir la interculturalidad. políticas educativas, diversidad cultural y desigualdad en ecuador. *Íconos. Revista de Ciencias Sociales*(60), 217 - 236. Retrieved from <https://libproxy.berkeley.edu/login?url=http%3a%2f%2fsearch.ebscohost.com%2flogin.aspx%3fdirect%3dtrue%26db%3dagh%26AN%3d127479249%26site%3ded-live>
- Rubalcava, L., & Teruel, G. (2007). *Guía de usuario: Encuesta nacional sobre niveles de vida de los hogares tercer ronda* (Tech. Rep.). Centro de Investigación y Docencia Económicas A.C. and Universidad Iberoamericana.
- Saperstein, A., Kizer, J. M., & Penner, A. M. (2016). Making the most of multiple measures: Disentangling the effects of different dimensions of race in survey research. *American Behavioral Scientist*, 60(4), 519-537. Retrieved from <https://doi.org/10.1177/0002764215613399> doi: 10.1177/0002764215613399
- Saperstein, A., & Penner, A. M. (2012). Racial fluidity and inequality in the united states. *American Journal of Sociology*, 118(3), 676–727. Retrieved from <http://www.jstor.org/stable/10.1086/667722>
- Saperstein, A., & Penner, A. M. (2014). Beyond the looking glass: Exploring fluidity in racial self-identification and interviewer classification. *Sociological Perspectives*, 57(2), 186-207. Retrieved from <https://doi.org/10.1177/0731121414523732> doi: 10.1177/0731121414523732
- Schmelkes, S. (2013, Junio). Educacion para un mexico intercultural. *Sinectica*(40), 01 - 12. Retrieved from http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-109X2013000100002&nrm=iso
- Soto Quirós, R., & Díaz Arias, D. (2006). *Mestizaje, indígenas e identidad nacional en centroamérica : de la colonia a las repúblicas liberales*. Sede Académica, Costa Rica, Facultad Latinoamericana de Ciencias Sociales (FLACSO). Retrieved from <https://libproxy.berkeley.edu/login?url=http%3a%2f%2fsearch.ebscohost.com%2flogin.aspx%3fdirect%3dtrue%26db%3dcato4202a%26AN%3ducub.b18067113%26site%3ded-live>
- Statistics New Zealand. (2005). *Understanding and working with ethnicity data. a technical paper* (Tech. Rep.). Wellington: Statistics New Zealand.
- Telles, E. (2007). Race and ethnicity and latin america's united nations millennium development goals. *Latin American and Caribbean Ethnic Studies*, 2(2), 185-200. Retrieved from <https://doi.org/10.1080/17442220701489571> doi: 10.1080/17442220701489571
- Telles, E., Flores, R. D., & Urrea-Giraldo, F. (2015). Pigmentocracies: Educational inequality, skin color and census ethnoracial identification in eight latin american countries. *Research in Social Stratification and Mobility*, 40, 39 - 58. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0276562415000177> doi: <https://doi.org/10.1016/j.rssm.2015.02.002>
- Telles, E., & Torche, F. (2018, 10). Varieties of Indigeneity in the Americas. *Social Forces*, 97(4), 1543-1570. Retrieved from <https://doi.org/10.1093/sf/soy091> doi: 10.1093/sf/soy091
- Thiede, B. C., & Gray, C. (2020). Characterizing the indigenous forest peoples of latin america: Results from census data. *World Development*, 125, 104685. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0305750X1930333X> doi: <https://doi.org/10.1016/j.worlddev.2019.104685>
- Thompson, S. C., Woods, J. A., & Katzenellenbogen, J. M. (2012, 11). The quality of indigenous identification in administrative health data in australia: insights from studies using data linkage. *BMC medical informatics and decision making*, 12, 133–133. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/23157943> doi: 10.1186/1472-6947-12-133
- Trujano, C. (2008). *Indigenous routes : a framework for understanding indigenous migration*. Geneva, Switzerland: International Organization for Migration.
- Villarreal, A. (2010). Stratification by skin color in contemporary mexico. *American Sociological Review*, 75(5), 652-678. Retrieved from <https://doi.org/10.1177/0003122410378232> doi: 10.1177/0003122410378232
- Villarreal, A. (2014). Ethnic identification and its consequences for measuring inequality in mexico. *American Sociological Review*, 79(4), 775-806. Retrieved from <https://doi.org/10.1177/0003122414541960> (PMID: 25999600) doi: 10.1177/0003122414541960
- Wickstrom, S., & D. Young, P. (Eds.). (2014). *Mestizaje and globalization: Transformations of identity and power*. University of Arizona Press.

Yoshioka, H. (2010). Indigenous language use and maintenance patterns among indigenous people in the era of neoliberal multiculturalism in Mexico and Guatemala. *Latin American Research Review*, 45(3), 5–34. Retrieved from <http://www.jstor.org/stable/40926268>