Seeing is Believing: Identity, Inequality, and the Impact of Television on the Hispanic Achievement Gap*

Andrew Kao†

November 2021

Abstract

Hispanics face the lowest high school and college completion rates out of all major ethnic and racial groups in the United States. In this paper, I investigate the impact of Spanish Language Television (SLTV) on Hispanic students in public schools using a spatial regression discontinuity arising from FCC regulation. I find that SLTV improves academic outcomes and narrows the Hispanic achievement gap, increasing SAT and ACT tests taken, enrollment in calculus, and AP exams passed. However, SLTV also causes more Hispanic students to be labelled 'limited English proficiency' and bullied on the basis of their ethnicity. I dig into the mechanism driving these contradictory results and find that Hispanic students perform better academically where SLTV programming focuses more on the Hispanic identity, but not when it focus more on role models or education itself. Furthermore, Hispanics with access to SLTV visit Hispanic branded establishments more frequently. Collectively, these findings suggest that the effects of SLTV are driven by its effects on identity.

JEL Codes: I24, J15, L82, Z13.
Keywords: Hispanic, television, education, identity

1 Introduction

The Hispanic achievement gap is wide and persistent. Hispanics face the lowest high school and college completion rates out of all major ethnic and racial groups in the United States.[1] In this paper, I argue that Spanish Language Television (SLTV) has increased Hispanic educational attainment, and that moreover, these gains can be attributed to a heightened sense of a Hispanic identity.

*Many appreciated suggestions, critiques and encouragement were provided by Leonardo Bursztyn, Albert Chen, Lucas Cusimano, Andrés de Loera-Brust, Benjamin Enke, Victor Lima, Aakaash Rao, Jaya Wen, David Yang, and Kotaro Yoshida, seminar participants at EPoD, Harvard PE, and the University of Chicago Honors Thesis Workshop as well as Mark Colombo for technical advice on televisions. This is a condensed, 10 page version of the paper. The full version of this paper can be accessed at `https://andrew-kao.github.io/files/sltv_draft.pdf`

†Harvard University. Email: `andrewkao@fas.harvard.edu`

[1]See Tienda (2009). This Hispanic achievement gap encompasses a wide range of educational outcomes from kindergarten test scores to enrollment in graduate programs. Factors such as segregation (Cascio and Lewis, 2012), socioeconomic and ESL status (Carpenter, Ramirez and Severn, 2006), and immigration status (Reardon and Galindo, 2009) exacerbate the Hispanic achievement gap, whereas interventions such as providing free computers (Fairlie, 2012), detracking (Burris and Welner, 2005), or school choice, performance-based pay, and alternative teacher certification (Ladner and Burke, 2010) may help close it.

Despite the rise of the internet, broadcast Spanish Language TV remains an important fixture in Hispanic households. 78% of Spanish-dominant households watch SLTV. In 2010, every single one of the top 10 shows watched by Hispanics were Spanish language programs (Pardo and Dreas, 2011). By investigating Spanish Language TV, I take a closer look at Hispanic communities and examine how identity can affect educational outcomes.

To identify the causal effect of SLTV, I follow Velez and Newman (2019) and exploit a spatial regression discontinuity arising from a Federal Communications Commission (FCC) regulation. This regulation grants federal protection of a TV station's broadcast signal to areas within a certain distance of a station's main antenna, with a sharp cutoff in enforcement beyond this distance. Thus, households and schools just inside a TV station's coverage contour should be observably similar to those just outside the contour, except for the presence of broadcast and satellite TV. This allows me to identify the causal effect of SLTV, given several features: (1) contours are mechanically decided by a formula involving geographical features and antenna strength, (2) contours are large and their boundaries tend to cut across small towns rather than urban centers (which fall squarely within contours), (3) SLTV stations were often built before this regulation was imposed, (4) demographic and other controls across the regression discontinuity are similar, and (5) Hispanics do not differentially migrate across contours, minimizing the possibility of selection. To further dispel concerns over potential confounds, I employ a difference-in-discontinuities design, comparing outcomes for Hispanic students against Asian students in schools with and without SLTV based on a 100 kilometer cutoff to SLTV coverage contours.[2]

I verify the relevance of this instrument's first stage by employing the difference-in-discontinuities design with the American Time Use Dataset. I find that Hispanics watch 10 minutes more TV within coverage contours. This is a plausible lower bound for the amount of extra Spanish Language TV watched if Hispanics do not substitute watching English programs with Spanish ones. I also show that Hispanics watch more TV with their children—Hispanic students, in other words. Notably, non-Hispanics do not exhibit differential TV viewership across SLTV coverage contours.

Next, I utilize the Civil Rights Data Collection to analyze the effect of SLTV on Hispanic students in public schools. The white-Hispanic achievement gap is large: 36.6% for the number of SAT and ACTs taken, 15% for the number of calculus courses taken, and 17.8% for the number of APs passed. The Asian-Hispanic gap achievement gap is larger still. I find that SLTV improves academic outcomes across the board for Hispanics: compared to Asians, Hispanics with SLTV are 16% more likely to take the SAT or ACT, 27% more likely to enroll in calculus and higher math, and pass 8% more AP exams. These gains are also present in absolute terms, extend to a variety of other academic outcomes, and remain qualitatively similar under a variety of robustness tests, establishing that SLTV reduces the Hispanic achievement gap.

However, I also find that Hispanic students are more likely to be classified as having 'limited English proficiency' in the presence of SLTV despite greater general academic achievement, a likely outcome if these students shift from English to Spanish mastery due to SLTV. Furthermore, Hispanic students are also bullied more on the basis of their ethnicity in the presence of SLTV, consistent with a more salient identity that other students may target.

Given these findings, I investigate in greater depth the mechanisms that drive these gains in Hispanic performance. I use archive.org's TV transcript database to classify the proportion of programs in each SLTV station that focus on the Hispanic identity. I show that a greater amount of SLTV programming focused

[2]I compare against Asian rather than white students because they are much less likely to identify as Hispanic.

on the Hispanic identity is associated with stronger Hispanic academic performance. However, a greater amount of programming focused on education or positive role models for children both have a null effect on Hispanic performance. This indicates that the content of these television programs matter, and that identity is a primary channel through which these gains are attained. Additionally, I use foot-traffic data from Safegraph to investigate engagement with Hispanic cultural experiences. Hispanics with SLTV are differentially more likely to visit Hispanic branded restaurants and recreation establishments. Conducting a placebo exercise, I find that Hispanics with SLTV are no more likely to visit Japanese, Brazilian, or Cajun and Creole establishments. This indicates a specific strengthening of the Hispanic identity versus a broader Latin American one. Collectively, these results suggest that identity is an important mechanism through which SLTV reduces inequality and the Hispanic achievement gap.

Layout. Following this Introduction, Section 2 presents the data sources used. Section 3 describes the difference-in-discontinuities empirical strategy and establishes the first stage. Section 4 presents evidence that SLTV narrows the Hispanic achievement gap, with two notable exceptions in 'Limited English Proficiency' and ethnicity-based bullying. Section 5 presents evidence that an identity mechanism underlies these results using SLTV transcript and foot-traffic data. Finally, Section 6 concludes with the prior literature and this paper's contribution.

2 Data

Coverage contours The central instrument used in this paper is the discontinuity in SLTV access across coverage contour boundaries introduced by FCC regulation. To build the coverage contours of SLTV stations in the United States, I combine data from TMS Media, a large provider of data on TV, movies, and other media, with the FCC's Consolidated DataBase System (CDBS) to obtain the coverage contour boundaries in 2015.

Public school data I collect data on public schools from the US Department of Education's Civil Rights Data Collection (CRDC) dataset in 2015. This data contains information on various indicators of educational performance. School addresses are geocoded using ArcGIS and coded as receiving SLTV if they fall within a coverage contour.

Television transcript data To code the content of programs broadcasted by SLTV stations, I make use of archive.org's television transcript database covering the years 2005-2015. Because transcript data is available at the television network level, I assign all affiliate stations data from their parent network. For each network in the database, I code the fraction of television programs whose transcripts contain keywords related to the mechanisms that I study: identity, education, and role models.

3 Empirical strategy

To isolate the causal effect of Spanish language television, I adapt the technique used in Velez and Newman (2019) and extend it from two counties to the entirety of the United States.

Digital and satellite TV stations operate by broadcasting signals from a central antenna, and the antenna's field strength at any given location is a mechanical product of several geographical and technical factors. This signal declines in strength with the square of distance, making it subject to interference and general loss of signal. When one gets far enough away from a TV station, this interference becomes widespread and meaningfully impedes TV viewership. To safeguard TV signals, the FCC passed in 1997 a series of regulations to protect signals for commercial TV stations from interference. These established coverage contours inside of which sufficiently strong interfering signals are banned.[3]

This regulation creates a natural spatial regression discontinuity. Combined with the decaying strength of a TV signal due to distance, this cutoff in broadcast protection creates a split among households and schools just inside and outside of these coverage contours that should be ex ante comparable save for their access to broadcast TV. This is operationalized as a 100 kilometer cutoff from the coverage contour border in my baseline specifications. Even with this discontinuity, one may still worriy that unobserved variation across coverage contours could drive the observed results. Therefore, I follow the recent literature on difference-in-discontinuities (Casas-Arce and Saiz (2015), Grembi, Nannicini and Troiano (2016)) and employ a design that, in addition to the regression discontinuity, also compares outcomes for Hispanic students against Asian students. I compare against Asian rather than white students because they are much less likely to identify as Hispanic. Thus, any alternative explanation for these results would need to differentially affect only Hispanics across these coverage contours. The main specification is:

$$y_{i,j} = \beta \mathbb{I}[InsideContour_{i,j}] \times \mathbb{I}[Hispanic_{i,j}] + \gamma_k + \delta X_i + \epsilon_{i,j}$$

where $y_{i,j}$ is an outcome for observation i (which may be an individual, school, or establishment) under demographic category $j \in \{\text{Hispanic, not Hispanic}\}$, γ_k is fixed effect for school district k (included when relevant), and X is a vector of controls for the observation. The main coefficient of interest is β, and in particular, the interaction term between the Inside Contour and Hispanic indicators.

3.1 First stage evidence: do Hispanics in SLTV coverage contours watch more TV?

I test for the amount of television watched across these contour boundaries with data from the American Time Use Survey. Figure 1 graphs the minutes of television watched against the distance to the contour boundary: it is clear that Hispanic television viewership increases inside the boundary while non-Hispanic viewership remains flat. Running this as a regression following the main specification, Hispanics watch an average of 10 minutes more television when within the contour, whereas the effect of the contour on non-Hispanics is insignificant. If one reasonably assumes that English and Spanish Language TV are not complements, then this estimate is a lower bound for the increase in SLTV watched by Hispanics.

[3]The relevant sections of federal law are 47 C.F.R. 73.622, 73.623, and 74.704. The FCC's OET Bulletin No. 69 most clearly summarizes and provides guidance on the salient features in this law. These contour interference protection lines are constructed following the Longley-Rice methodology also adopted in 1997 and are termed $F(50, 10)$ lines.

4 The impact of Spanish language television on Hispanic educational performance

To determine the effect of SLTV on the Hispanic achievement gap, I apply the main difference-in-discontinuity specification at the school-ethnicity level and compare differential outcomes between Hispanics and Asians while varying the presence of SLTV.

Table 1, Panel A presents results on the IHS transformed number of SAT and ACT tests taken, Panel B presents results on the IHS transformed number of calculus courses taken, and Panel C presents results on the IHS transformed number of AP exams passed. Results are statistically and economically significant. SLTV differentially increases the number of Hispanics taking the SAT or ACT by 16% (Panel A), the number of Hispanics taking calculus by 27% (Panel B), and the number of Hispanics passing an AP exam by 10% (Panel C). The degree to which this decreases the achievement gap is presented in Appendix Table A.3, Column 3. These results also hold in absolute terms (see Appendix Table A.4) and hold for a variety of other outcomes, including the number of gifted students, advanced math courses, and college preparatory science courses taken (see Appendix Table A.5). The results are also robust to a variety of specifications and I rule out alternative hypotheses in Appendix Table A.6.

Taken as a whole, these results suggest that SLTV is a meaningful force that can improve Hispanic student performance in public schools and reduce the achievement gap. However, academic brilliancy is not typically associated with the banal, mindless enjoyment of lazing on a couch before a flatscreen. So what might drive these results?

4.1 Identity within schools

I identify two outcomes in public schools that speak to the strength of one's Hispanic identity: (1) classification as a 'Limited English Proficient' (LEP) student, and (2) harassment or bullying on the basis of race, color, or national origin. Table 1, Panel D shows that Hispanics with access to SLTV are 30% more likely to be classified as a 'Limited English Proficient' (LEP) student when compared to their peers. This decrease in educational performance contrasts with all other academic outcomes studied thus far, suggesting that it is not a difference in general intelligence or work ethic that drives this result, but rather an idiosyncratic decline in English speaking ability. Panel E shows that Hispanics with access to SLTV are also differentially more likely to be bullied on the basis of their ethnicity. Appendix Table A.8 shows that these results hold in absolute terms as well. Any mechanism through which SLTV increases the propensity for Hispanic students to be bullied (such as students picking on strong academic performers) would need to reconcile the null result for bullying based on sex. My preferred explanation is that Hispanics watching SLTV make their ethnic identity more salient, making them a greater target for bullying along this specific dimension.

Thus, though it is impossible to rule out all other stories that may drive this set of results, there are not many which can explain the reversal in academic ability for English proficiency and the increased bullying on the basis of ethnicity but not sex among Hispanic students. A strengthened Hispanic identity through SLTV, wherein Hispanics feel a stronger affinity towards Hispanic cultural practices, people, and countries of origin, neatly fits these facts.

5 Zeroing in on the identity mechanism

I turn to the content of these SLTV programs in order to assess the potential mechanisms through which SLTV could increase educational performance: (1) SLTV programs strengthening the Hispanic identity, (2) SLTV programs stressing the importance of education, and (3) SLTV programs providing good role models for students. I infer the content of SLTV programs using keyword matching of terms related to each mechanism in the archive.org TV transcript database, described in Table A.1. To evaluate the strength of each mechanism, I modify the main specification by interacting the difference-in-discontinuity with the salience of each mechanism:

$$y_{i,j} = \beta \mathbb{I}[InsideContour_{i,j}] \times \mathbb{I}[Hispanic_{i,j}] \times Mechanism_s + \gamma_k + \delta X_i + \epsilon_{i,j}$$

where s indexes a given SLTV station, and $Mechanism_s$ is the percentage of programs from the nearest station focused on a given mechanism. The triple interaction term between Inside Contour, Hispanic, and Mechanism yields the elasticity of the outcome with respect to the mechanism, and is thus the coefficient of interest. Table 2 displays results for regressions following this specification. I focus first on column 1, which measures the strength of the identity mechanism. The triple interaction term is positive for all academic outcomes. These results indicate that SLTV stations increase Hispanic academic performance more when they focus more on the Hispanic identity. Thus, if these identity-focused shows impart a stronger sense of identity for Hispanics, then this would suggest that identity can be mobilized to close the Hispanic achievement gap.

However, the results do not support other mechanisms driving the increases in educational attainment. Table 2, Column 2 suggests that a greater emphasis on education in SLTV programs does not necessarily translate into stronger academic performance among Hispanics when compared to their peers. Table 2, Column 3 looks at the percentage of programs that contain good role models for students and children. As in the preceding case of education, the sign on the triple interaction term is mixed and never significant.

In the full version of this paper, I also examine foot-traffic data and find that Hispanics more frequently visit Hispanic branded establishments when in the presence of SLTV. Collectively, these results suggest that an identity mechanism drives these results.

6 Conclusion

Americans spend an average of three hours a day watching TV—more than any other activity but sleep! Accordingly, a large literature has examined the impact that television has on education. Prior work has frequently been correlational and findings remain conflicted: one line of research contends that TV is as a distraction which 'rots' the mind and harms student outcomes (Zavodny, 2006),[4] while another line of inquiry has found consistent null effects.[5] Gentzkow and Shapiro (2008) are closest to this paper in using

[4]See also Aksoy and Link (2000), Hornik (1981), and Keith et al. (1986). This theory enjoys popular support (see Winn (2002) or Gentile (2004) which finds broad support for the theory among paediatricians). Huang and Lee (2010) and Nakamuro et al. (2015) use more sophisticated panel data approaches and also find negative (but smaller) effects.

[5]Gaddy (1986), Gortmaker et al. (1990), and Hu et al. (2020) take correlational approaches, while Munasib and Bhattacharya (2010) and Kureishi and Yoshida (2013) use self-reportedly weak instruments that may generate the null.

a difference-in-difference strategy to find that TV improves student test scores—particularly among non-white students and English language learners. I contribute to this literature by taking a quasi-experimental approach and examining mechanisms.

Others have studied the effect of television on Hispanic communities. Oberholzer-Gee and Waldfogel (2009) demonstrate that the presence of Spanish language local news increases Hispanic voter turnout, whereas Velez and Newman (2019) (who develop the instrument used in this paper) find that SLTV depresses Hispanic voter turnout. I extend on this literature by moving beyond the political realm, arguing that the consequences of SLTV are large in educational settings, and also provide the first evidence on a mechanism through which SLTV operates: identity.

There is a growing literature that looks at how identity can influence behaviour. This has been studied through theory, in the lab, and the field.[6] However, the underlying forces that construct identity (rather than simply triggering them via priming or other short-term interventions) are less well understood. Bisin et al. (2010), Atkin, Colson-Sihra and Shayo (2019), and Bazzi et al. (2019) encompass some recent studies on this topic, and all come to the conclusion that intergroup tensions or differences lead to a strengthening of identity. I contribute to this literature by proposing a media-based channel through which the Hispanic identity may be strengthened and influence action. This is closest to work such as Jensen and Oster (2009) and Gentzkow and Shapiro (2004), which establish a link between media & gender norms and media & anti-Americanism respectively.[7]

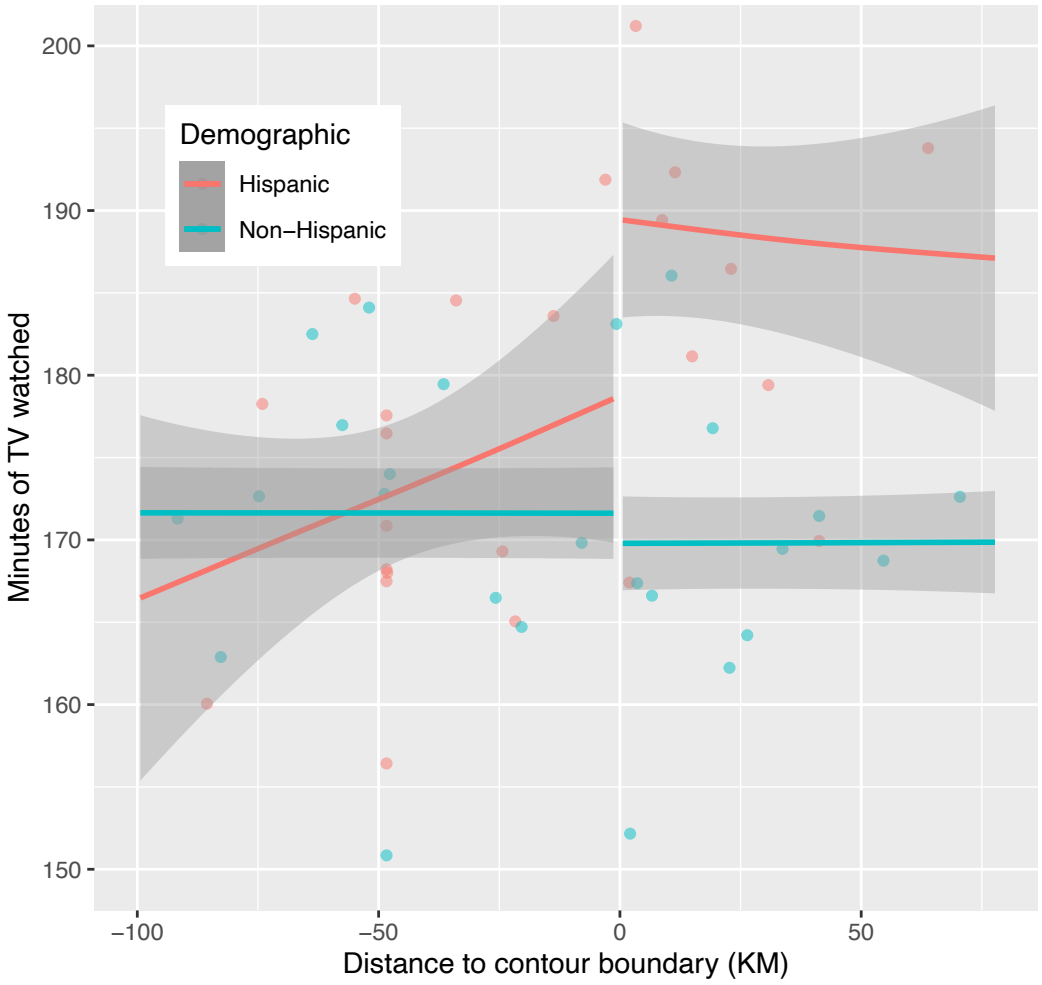
Finally, in the education and psychology literature, stereotype threat is a phenomenon that pinpoints minority identities as a root cause of achievement gaps (Appel and Kronberger (2012), Spencer, Logel and Davies (2016)). This has led to the rise of methods such as "situational disengagement" to avoid the negative stigma of identity (Nussbaum and Steele, 2007). This paper argues that a stronger sense of identity may not have uniformly negative consequences on Hispanic students, creating space for a more positive conception of identity.

[6]See Akerlof and Kranton (2000), Benjamin, Choi and Strickland (2007), Benjamin, Choi and Fisher (2010), and Bursztyn et al. (2019), among others. Alesina, Giuliano and Nunn (2013) take the long view and show how gender norms can be traced back to early agricultural practices.

[7]Other related work on the impact of mass media on social outcomes include Ferrara, Chong and Duryea (2012), Kearney and Levine (2015), Olken (2009), DellaVigna and Kaplan (2007), Yanagizawa-Drott (2014), and Putnam (2001). For an overview, see DellaVigna and La Ferrara (2015).

Figures and Tables

Figure 1: Minutes of TV watched across the coverage contour



Lowess smoothed (lines) and binscattered (points) minutes of television watched by distance to SLTV coverage contour boundary. Negative values indicate individuals outside of a contour (no SLTV). Hispanic viewership is in red, non-Hispanic viewership is in blue. Minutes of TV watched are residualized by individual level age, age^2, sex, and county level controls for log(income), log(population) and percent Hispanic.

Table 1: Effect of Spanish language TV on Hispanic vs. Asian academic achievement

	(1)	(2)	(3)
Panel A: IHS(SAT/ACTs taken)			
TV dummy × Hispanic	0.1598***	0.1598***	0.1598***
	(0.0264)	(0.0264)	(0.0264)
N	21,610	21,610	21,610
Panel B: IHS(calculus taken)			
TV dummy × Hispanic	0.2718***	0.2718***	0.2718***
	(0.0369)	(0.0369)	(0.0369)
N	11,460	11,460	11,460
Panel C: IHS(APs passed)			
TV dummy × Hispanic	0.0964***	0.0966***	0.0972***
	(0.0346)	(0.0353)	(0.0360)
N	3,757	3,757	3,757
Panel D: IHS(limited English proficiency)			
TV dummy × Hispanic	0.3042***	0.3042***	0.3042***
	(0.0379)	(0.0379)	(0.0379)
N	83,004	83,004	83,004
Panel E: IHS(bullied based on ethnicity)			
TV dummy × Hispanic	0.0015*	0.0015*	0.0015*
	(0.0009)	(0.0009)	(0.0009)
School district FE	Yes	Yes	Yes
# Hispanic, Asian students	Yes	Yes	Yes
School size controls	No	Yes	Yes
School type controls	No	No	Yes

Notes: The table presents coefficient estimates from regressions at the school-ethnicity level, only keeping schools within 100 KM of a Spanish language TV contour boundary. The dependent variables are inverse hyperbolic sine transformed counts of the number of students taking the SAT or ACT in Panel A, the number of students enrolled in calculus in Panel B, the number of Advanced Placement exams passed in Panel C, the number of students labelled as having limited English proficiency in Panel D, and the number of students bullied on the basis of their ethnicity in Panel E. TV dummy is an indicator variable for a school with access to Spanish language television, which is interacted with an indicator for whether the demographic is Hispanic (the omitted group is Asians). Columns 1-3 control for the number of Hispanic and Asian students enrolled. Columns 2-3 control for the number of teachers and total number of students at the school. Column 3 controls for indicators denoting whether the school contains a primary, middle, and high school division. School district fixed effects are always included. Standard errors are clustered at the school district level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 2: Differential effect of Spanish language TV by program content on Hispanic vs. Asian academic achievement

	(1)	(2)	(3)
Panel A: IHS(SAT/ACTs taken)			
TV × Hispanic × % programs on identity	2.313**		
	(0.943)		
TV × Hispanic × % programs on education		−0.516	
		(0.626)	
TV × Hispanic × % programs with role models			−2.085
			(2.151)
N	21,610	21,610	21,610
Panel B: IHS(calculus taken)			
TV × Hispanic × % programs on identity	2.788***		
	(1.034)		
TV × Hispanic × % programs on education		0.829	
		(0.666)	
TV × Hispanic × % programs with role models			1.616
			(2.463)
N	7,112	7,112	7,112
Panel C: IHS(APs passed)			
TV × Hispanic × % programs on identity	1.721		
	(1.280)		
TV × Hispanic × % programs on education		0.903	
		(0.922)	
TV × Hispanic × % programs with role models			−1.184
			(2.989)
N	3,168	3,168	3,168
School district FE	Yes	Yes	Yes
# Hispanic, Asian students	Yes	Yes	Yes
School size controls	No	Yes	Yes
School type controls	No	No	Yes

Notes: The table presents coefficient estimates from regressions at the school-ethnicity level, only keeping schools within 100 KM of a Spanish language TV contour boundary. The dependent variables are inverse hyperbolic sine transformed counts of the number of students taking the SAT or ACT in Panel A, the number of students enrolled in calculus in Panel B, and the number of Advanced Placement exams passed in Panel C. % programs on identity, education, and role models are coded based on TV channel network transcripts. TV dummy is an indicator variable for a school with access to Spanish language television, which is interacted with an indicator for whether the demographic is Hispanic (the omitted group is Asians) and the % of programs on identity, education, and role models. Columns 1-3 control for the number of Hispanic and Asian students enrolled. Columns 2-3 control for the number of teachers and total number of students at the school. Column 3 controls for indicators denoting whether the school contains a primary, middle, and high school division. School district fixed effects are always included. Standard errors are clustered at the school district level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.