Area Deprivation and the P20W Pipeline

Ari Fenn

October 2021

Abstract

This research shows the negative relationship between area deprivation and educational outcomes in the State of Utah. This study used data from a pooled cross-section for the 2017 and 2018 school years of 12 P20W metrics from the Utah State Board of Education, and Utah Systems of Higher Education merged with the Health Improvement Index from the Utah Department of Health. Descriptive statistics show strong negative relationships between area deprivation and 10 of the 12 P20W metrics. After regression analysis was used to control for individual characteristics of students there is clear evidence that area deprivation is negatively related to P20W metrics. Students from the most deprived areas are least likely to be ready for school, measured entering kindergarten, and least likely to attend any post-secondary institution; additionally students from the most deprived areas are most likely to be chronically absent in early elementary grades and high school.

Keywords

Area Deprivation, early education, post-secondary enrollment, Utah

1 | Introduction

This research shows high levels of area deprivation in Utah are associated with lower school readiness, students being less prepared for post-secondary education, and lower post-secondary enrollment. Area deprivation describes the broad social and economic health of a given geographic region. Areas with high concentrations of poor social and economic outcomes are deprived while areas with high concentrations of positive social and economic outcomes are said to be not deprived.

This study is important for Utah policymakers because education outcomes partially determine one's long-term earnings potential and labor market success (Heckman, Humphries, & Veramendi, 2018). Specifically in Utah, completing a career and technical education (CTE) program is associated with 35% to 59% higher wages (Scott, 2019). The completion of any post-secondary program is associated with higher general consumption (Scott, 2020). While it is known that there are disparities in outcomes based on social, economic, and demographic characteristics of Utahns (Utah Data Research Center, 2020), less is known about educational disparities based on the area of residence and if area deprivation shows patterns of disparities.

To address the existence of disparities in education outcomes based on place of residence in Utah this research exploits unique administrative data across several Utah State agencies. Education outcomes that are important to long-term labor market success were established by the Utah Governor's Education Excellence Commission and are known as P20W metrics (Utah Data Research Center, 2020). Several of these P20W metrics will serve as the educational outcomes to test for the existence of disparities that are associated with area deprivation. These outcomes are: school readiness, proficiency in core subjects, postsecondary readiness, and post-secondary access. To measure area deprivation the Health Improvement Index (HII) was created by the Utah Department of Health to test for the existence of health disparities based on the geography of residence (Utah Department of Health, 2020). Additionally, this paper seeks to establish a relationship between the final P20W metric, industry success, and area deprivation as measured by HII.

In Utah the HII has been used to link area deprivation to several adverse outcomes. These range from higher levels of pollution, adult asthma prevalence, and asthma emergency rooms visits (Vowles, Kerry, Ingram, & Mason, 2020) to higher infection rates of SARS-Cov-2 and a higher likelihood of hospitalization from infection (Lewis, et al., 2020). These recent studies demonstrate that HII is well suited to the study of disparities that arise from area deprivation in Utah.

This study establishes the existence of a relationship between various P20W metrics and area deprivation. Given current data limitations, it is not possible to establish a causal relationship or causal mechanisms. Despite this limitation, it is important to lay out plausible pathways through which area deprivation and P20W metrics are associated. The first occurs through the schools. In this scenario, wealthier areas have better-funded schools (Mangino and Silver, 2010). This advantage leads to all students who attend a wealthy school benefiting from the overall wealth of the area. The second potential pathway is through physical and psychological stressors to which a student is exposed. Ludwig et al. (2012) find that moving to a less deprived neighborhood leads to better self-reported mental health and better overall subjective well-being. As already noted, in Utah less deprived areas have less pollution (Vowles, et al., 2020). Higher levels of air pollution are associated with increased risk of heart attack (Shah, et al., 2020), higher incidence of type 2 diabetes (Yang et al., 2020), asthma exacerbations in children (Orellano, et al., 2017), depression (Fan, et al., 2019), increased risk of stroke (Shah, et al., 2015), and hypertension (Yang, et al., 2018). The combination of better physical and mental health contributes to better education outcomes (Suhrcke & de Paz Nieves, 2011). Finally, the networks that students have access to may be based on the area in which the student lives. Networks are the groups of people with whom an individual interacts on a regular basis. These networks have been shown to shape job referrals (Bayer, Ross, & Topa, 2008). These same networks may help to shape norms on attendance, the importance of and how to start post-secondary education, and provide resources for students who are struggling with school.

The industrial structure may be associated with area deprivation in several ways. The first again relies on the networks that exist in a given geography. It is possible that areas with lower-paying and low mobility jobs expose residents to networks that only allow access to the same low-paying and low mobility jobs. Employment has positive mental health effects (van der Noordt, IJzelenberg, Droomers, & Proper, 2014), employment is more volatile in certain industries and less volatile in others. If employment is centered on more volatile injuries in a given area, there can be mental health effects that cascade into additional negative social outcomes.

1.2 | Literature Review

There is a large and diverse body of literature on the effects of area deprivation or neighborhood effects on education. These span many disciplines and are too extensive to exhaustively cover. The literature that is reviewed serves to establish the methods and the measures used in this paper along with providing recent results for comparison. Overall, the literature reviewed here points to strong evidence that area deprivation is not only associated with poor academic performance but causes poor academic performance (Ferguson and Michaelsen, 2015; Laliberté, 2021). The literature surveyed here also provides evidence that social and economic variables together, or indices of those variables, are the appropriate measure of area deprivation rather than solely economic or social variables (Ferguson and Michaelsen, 2015; Wodtke and Parbst, 2017; Weinberg, et al., 2019; Wodtke, Yildirim, Harding, & Elwart, 2020; Laliberté, 2021).

Ferguson and Michaelsen (2015) show both the importance of area deprivation on early learning outcomes and the importance of a definition of area deprivation that includes social and economic variables. Ferguson & Michaelsen (2015) use school level pass rate for English and Mathematics standardized tests. In this study area deprivation is measured at the electoral ward level and is a weighted average of income, employment, health, education, living environment, and crime deprivation (Ferguson & Michaelsen, 2015). To control for endogeneity the measure of deprivation is instrumented by deaths from conflict between 1969 and 1994. There is a clear strong negative relationship between the author's measures of deprivation and percent of students who are proficient. Additionally, the noneconomic measures of deprivation play a significant role in the explanatory power of the models and provide evidence for a holistic approach to measuring deprivation (Ferguson & Michaelsen, 2015).

Wodtke and Parbst (2017) use the data from the United States to test how neighborhood effects are mediated to student outcomes. The authors use a principal component analysis on poverty rate, median household income, unemployment rate, proportion of female-headed households, proportion of people over 25 without a high school diploma, proportion of residents over 25 with a college degree, and the proportion of the residents that are 25 or older in managerial positions calculated at the census tract level. The outcomes measured are childhood and adolescence reading and mathematics scores. Overall, students from the least deprived neighborhoods have the highest scores on tests in childhood but the effects diminish in adolescence (Wodtke & Parbst, 2017).

In a similar study Wodtke, Yildirim, Harding, and Elwart (2020) test for neighborhood effects, mediation through schools, and the interaction of the two using data from the United States. The measure of area deprivation is a principal component analysis of seven social and economic characteristics of the census tract in which a child lives. Early exposure to high levels of area deprivation leads to lower math and reading test scores in 3rd grade (Wodtke, et al., 2020).

In the context of the Netherlands, Weinberg, et al. (2019) combine neighborhood socioeconomic status (SES) with parental education to measure: education trajectory, teacher evaluation of student ability, and cognitive ability. Neighborhood SES is a composite measure of average income, percent low income, percent "low-educated" (which is not further defined), and percent not working. There are fewer components in this measure of SES than in this study. After controlling for parental education and SES, there is no evidence of a significant link between educational attainment and neighborhood SES (Weinberg, et al., 2019).

Laliberté (2021) tests for effects of both schools and neighborhoods on secondary graduation and post-secondary attendance in Quebec. Neighborhood economic and social measures are used as the measure of SES status. There is evidence of both school and neighborhood effects on post-secondary education, though school effects dominate the neighborhood effects (Laliberté, 2021).

2 | Methods

2.1 | Data

The measure of area deprivation used in this study is the Health Improvement Index, HII, from the Utah Department of Health (UDOH). HII was calculated by UDOH for the 2017 calendar year using data from the Behavioral Risk Factor Surveillance System. The index is a composite measure of nine social, economic, and demographic characteristics that represent some of the social and economic determinants of health. These are: the percent of the population over 25 years old with fewer than nine years of education, the percent of the population over 25 years old with at least a high school diploma, median family income, income disparity, percent of housing units that are owner-occupied, unemployment rate, percent of families below the poverty level, percent of population below 150 percent of the poverty level, percent of single-parent households with children under the age of 18. The first factor from a factor analysis was centered and scaled around 100. The HII ranges from 71.89 to 160.87. Additionally, HII is divided into 5 groups: Very High HII > 120, High HII >105 and ≤ 120, Average HII > 90 and ≤ 105, Low HII > 80 and ≤ 94, and Very Low HII < 80 (Utah Department of Health, 2020). High

HII translates to high levels of area deprivation; the "Very Low" areas have the least deprivation while the "Very High" areas have the most deprivation. HII is community level measure at is the same for all residents in a given area.

HII is calculated for each Utah Small Area, geographic units created by UDOH to collect and organize data at the community level. Small Areas are determined based on political boundaries, population size, and economic similarity (Utah Department of Health, 2021). HII is connected to Small Area which is a non-standard geographic measure and requires a strategy to match subsequent elements of the data to Small Area, this strategy is discussed in the next section. There are two Small Areas that do not have demographic information available due to their recently created ZIP Codes[1]. These Small Areas are included in the analysis of education P20W metrics but not workforce outcomes.

The educational outcome data used in this study come from the Utah State Board of Education (USBE), and Utah System of Higher Education (USHE). USBE outcomes are from the 2016/17 and 2017/18 school years as those years contain the 2017 calendar year, the year for which the HII was calculated. USHE data is linked to students who graduated high school in either 2017 or 2018.

The Kindergarten Entry and Exit Profile (KEEP) measures literacy and numeracy skills entering and exiting kindergarten. Scores are measured 1 (no prerequisite knowledge), 2 (minimal prerequisite knowledge), and 3 (sufficient prerequisite knowledge). Entrance scores are used to measure the effect of area deprivation before any mediation through schools. The scores were transformed to prepared, any student that scored a three, or unprepared, students that scored a two or one.

Core subject proficiency for the third and eighth grades is measured by the Student Assessment of Growth and Excellence (SAGE) test scores. Third-grade outcomes are English Language Arts (ELA) and math, while eighth grade includes science. The students are assigned a proficiency level based on the scale score. Proficiency scores range from one to four, where scores of "1" or "2" are not proficient and scores of "3" or "4" are proficient. The outcomes were transformed to a binary measure of proficiency with 1 being proficient and 0 not proficient.

Attendance is measured for K-12 during the 2017 and 2018 school years where the outcome is if a student is chronically absent. USBE provides a measure of chronically absent. For a student to be considered chronically absent they must be enrolled for at least 60 days and miss at least 10% of the days enrolled.

1 These are "Taylorsville (WEST)" which is ZIP Code 84129 and Daybreak which is ZIP Code 84009.

This measure does not distinguish between excused and unexcused absences, and it is not a measure of truancy. Courses that are considered college preparatory courses are advanced placement (AP), International Baccalaureate (IB), and concurrent enrollment courses. If a student was enrolled in any of these courses during the 2016/17 or 2017/18 school year the student is considered to attempt a preparatory course and assigned 1, otherwise, the student is marked as not attempting a preparatory course and is assigned a 0. This study does not consider scores on an AP test; there is a fee to take an AP test and lower income students may decide to not take a test despite enrolling in the course due to the financial burden associated with the test.

The ACT is available to all Utah juniors to take for free. The ACT purports to correlate with the probability of a student succeeding in college or university and is often used in admission decisions. The ACT reports if a taker is "college ready" if a student makes a minimum score on at least one of the subject areas, from 18 for English to 26 for the STEM subjects. These scores are associated with a higher probability of success in post-secondary education (Allen & Radunzel, 2017). This binary college readiness score is used as the second measure of post-secondary readiness.

Post-secondary attendance is determined if a high school graduate appears in any post-secondary program. If a student in a program at a Utah technical college or degree-granting institution matches to a student that graduated from a Utah high school in either 2017 or 2018 that student is marked as successful post-secondary attendance. UDRC data only cover public institutions in Utah, and not private institutions such as: Brigham Young University, Westminster University, or Western Governors University.

For this study STEM majors are defined using the U.S. Department of Homeland Security's "STEM Designated Degree Program List." This list uses Classification of Instructional Program (CIP) codes to determine which majors are STEM. Four two-digit CIP codes are designated as STEM: Engineering, Biological and Biomedical Sciences, Mathematics and Statistics, and Physical Sciences. Beyond those that include all fields, there are an additional 220 six-digit CIP codes. These codes were matched to students' intended major, current major, and second current major. If any of these majors matches the designated STEM CIP codes, the student is listed as a STEM major.

Additional student demographic information comes from USBE. Gender is reported for every student, the majority are coded "F" or "M" with "U," unknown, and "O," potential data entry error, are also present in the data these are kept as is. A merged race and ethnicity category is reported with seven broad categories; Asian, Black, White, Hispanic, Indian, Multiracial, and Pacific Islander. The homeless status of the student is reported with 0 having never been homeless and 1-5 some levels of homeless. For this analysis, homeless status was transformed to a binary with a 0 representing the absence of homelessness and 1-5 assigned 1, was or is homeless. A binary variable controls for if a student is low income, which takes a value of one if a student is eligible for reduced school lunch, and zero otherwise. This is included as income partially determines HII but is also associated with test scores and controlling for student income reduces bias in the estimates of the relationship between HII and P20W metrics. Finally, English language learning (ELL) status is included with statuses English learner (Y), eligible but opted out (O), reclassified as fluent (F), not needed (N). ELL status is not transformed for this analysis.

Industry data come from the Department of Workforce Services (DWS) Quarterly Census of Employment and Wages (QCEW) program. There are two separate surveys available, March and September, only September is used for this study. The data contain the name of the firm, address, a six-digit NAICS code, and an employment range. The employment range is converted to the midpoint between the maximum and the minimum number of employees. The firms were aggregated to the two-digit NAICS, or sector, to create an employer concentration measure.

Employer concentration is used to measure the competitiveness of the labor market at the Small Area level. A Herfindahl-Hirschman index (commonly denoted HHI but to avoid confusion with the measure of deprivation is denoted CI) is constructed for each industry in each small area. Traditionally, CIs have been used for output markets though recently CIs have been used to measure employer concentration (Hershbein, Macaluso, & Yeh, 2020). The HHI for an industry in a Small Area is:

$$CI = \sum s_i^2$$

Where s_i is the employment share of the i^{th} firm in the industry of interest. CI can take a value of close to zero for the most competitive industries to 1 for a perfect monopsony, where a single firm is the only employer and is assumed to have substantial wage setting power. A market share of 0.25 or higher is associated with a highly concentrated market where firms have considerable wage setting power (Azar, Marinescu, & Steinbuam, 2020). CI is also aggregated to measure employer concentration for the top five industries in each Small Area:

$$CI = \sum \omega_j [\sum s^2_{(j,i)}]$$

In the above equation $s_{(i,j)}$ is the employment share of firm i in industry j and the inner sum is the CI of the j^{th} industry and ω_j is the weight of percent of employment in any Small Area in industry j.

2.2 | Geographic Matching

Three matching strategies were used to connect different measures to the Utah Small Areas. The first uses latitude and longitude data from USBE, the second uses the UDOH algorithm for small areas (Utah Department of Health, 2021), finally, ACS data uses Census Bureau crosswalks (US Census Bureau, 2021) then the UDOH algorithm. USBE provides latitude and longitude for each school. A student was assigned to a Small Area by first assigning them to the school that reported the measure of interest. The schools were then matched to a Small Area using Shapefiles from Utah SGID. The Simple Features (sf) package (Pebesma, 2018) for R was used to link schools to Small Areas using the st_join function with the st_contains link. This matched the latitude and longitude for each school to the Small Area that contains it. Figure 1 demonstrates the matching strategy. The boundary of the Salt Lake Southeast Liberty Small Area is defined by the Shapefile. Two schools, Uintah School and Clayton Middle are defined by latitudes and longitudes. The algorithm identifies that Uintah School is contained within Southeast Liberty and attaches the corresponding HII information. The algorithm identifies that Clayton is not in Southeast Liberty but is in Salt Lake City Foothill/East Bench Small Area and attaches the corresponding HII information.

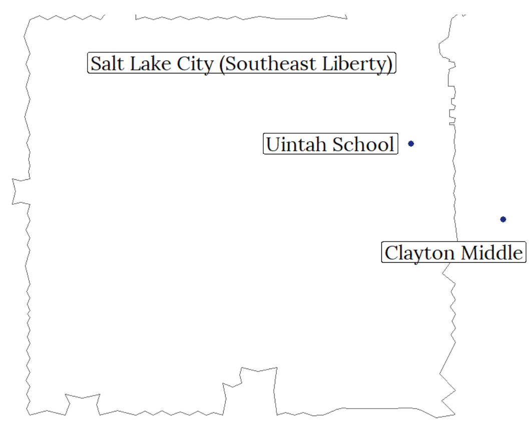

Figure 1: Geographic matching strategy

QCEW data from DWS contains the address of each firm. Using zip code and county the majority of firms were matched using the algorithm UDOH used to create the Small Areas (Utah Department of Health, 2021). The firms with incorrect counties or zip codes provided were matched by hand to the correct Small Area. ACS data was retrieved at the tract level and aggregated to zip code using the provided ZCTA with the Census Bureau "2010 ZCTA to Census Tract Relationship File." Weights for each tract were created using the ratio of tract population to ZCTA population.

2.3 | Empirical Strategy

Regression analysis was used to isolate the relationship between HII and each outcome of interest. For all P20W metrics except industry success, HII is an independent variable. Additionally for each of the regressions where HII is an independent variable a probit regression technique is used. This is due to the binary nature of the outcomes, where a linear model could produce predicted probabilities that are less than zero or greater than one. For KEEP, and SAGE proficiency the regressions take the form:

$$Probit(Y)= \beta_0 + \beta_1 Gender + \beta_2 Race/Ethnicity + \beta_3 Low\text{-}Income + \beta_4 ELLStatus + \beta_5 HomeStatus + \delta School\ Year2018 + \zeta GeographicType + \lambda HII\ Group + \epsilon$$

The first 7 regressions represented by Equation 1 control for gender, race/ethnicity, if the individual comes from a low-income family, English language learner status, if the student is homeless, if the Small Area is urban, rural, or frontier, and the HII category the student is matched to. The reference levels for each characteristic of the student except HII is the modal student in the data; a white male native English speaker, never homeless, not low income, and matched to an urban Small Area. The HII category reference group is Very Low HII while the school year reference group is 2017. Next, the models used to estimate college readiness, ACT, and prep courses, take the form:

$$Probit(Y)= \beta_0 + \beta_1 Gender + \beta_2 Race/Ethnicity + \beta_3 Low\text{-}Income + \beta_4 ELLStatus + \beta_5 HomeStatus + \beta_6 Grade\text{-}Level + \delta SchoolYear2018 + \zeta GeographicType + \lambda HIIGroup + \epsilon$$

Equation 2 includes the same controls as Equation 1 with the addition of a control for the grade that the student is in, GradeLevel. Grade is important to control for as the availability of prep courses is not the same for the different grades in secondary school. The model to estimate chronic absenteeism takes the form:

$$Probit(Y)= \beta_0 + \beta_1 Gender + \beta_2 Race/Ethnicity + \beta_3 Low\text{-}Income + \beta_4 ELLStatus + \beta_5 HomeStatus + \beta_6 Grade\text{-}Level + \beta_7 GradeLevel*HIIGroup + \delta SchoolYear2018 + \zeta GeographicType + \lambda HII\ Group + \epsilon$$

Equation 3 includes an interaction between the grade of the student and the HII group of the student.

The final model examines HII as a dependent variable. The purpose is to answer if the industrial structure of a small area relates to the deprivation of that area. This model is an ordinary least squares regression of the form:

$$HII= \beta_0 + \beta_1 GeographicType + \beta_2 \%Minority + \beta_3 \%CollegeAge + \beta_4 \%Over65 + \beta_5 Top5IndustryConcentration + \beta_6 GeographicType*\%Minority + \epsilon$$

In Equation 4 there are controls for if a Small Area is urban, rural, or frontier. Controls for the demographics of the Small Area are the percent of residents that are a minority, percent of residents that are college age, and percent of residents that are over the age of 65. The control for employer concentration of the top five industries is measured by HHI. The model also includes an interaction between the percent minority and the geography type to control for potentially different relationships between urban and rural or frontier areas and the reasons minorities are concentrated in the areas.

All analysis was performed using R v3.6.0 (R Core Development Team, 2019), in RStudio (RStudio Team, 2020), with the following packages: Tidyverse (Wickham et al., 2019), Broom (Robinson et al., 2020), Flextable (Gohel, 2021), Rstatix (Kassambara, 2020) and Multicomp (Hothorn et al., 2008).

3 | Results

3.1 | Descriptive Statistics

Area deprivation and P20W outcomes have a negative relationship. The percent of students who have a positive outcome by HII group and the state average is shown in Figure 2 with measures of each P20W metric and significance tests reported in Appendix A Table 1. For the majority of outcomes, the Very Low HII group has the highest percentage of students who are proficient while the "Very High" HII group has the lowest percentage of students that tested proficient. For kindergarten and third grade standardized tests the percent of students that test as proficient monotonically decreases from the lowest to the highest HII groups. This pattern also holds for ACT College Ready and attending any post-secondary institution. Additionally, the proportions for the "Very High" and "Very Low" groups are significantly different from the state average as shown in Appendix A Table 1. The opposite pattern emerges for chronic absenteeism where the percent of students chronically absent increases moving from the lowest to the highest HII areas. Proficiency in eighth-grade standardized tests shows a different pattern. Moving from the lowest to the highest HII group a lower percentage of students are proficient but levels off between the "High" and "Very High" HII groups. Students from the "Very High" HII group are least likely to attempt a college prep course while enrolled in high school, as opposed to other outcomes the "Very Low" and "Low" HII Groups switch as do the High and Average HII groups. There does not appear to be an association between STEM majors and HII. Appendix A Table 2 contains pair-wise comparisons of proportions between HII groups. For all outcomes, the "Very High" and the "Very Low" areas are significantly different from each other. Taken as a whole the unconditional trends point to area deprivation as having a clear strong and negative relationship with P20W metrics.

The major industry does not systematically vary by HII group. Table 1 shows the top five industries by employment volume. In "Very Low," "Low," and "Average" HII areas Health Care and Social Assistance employs the most people, for "High" HII areas this industry represents the second largest employing industry and the third largest industry for "Very High" areas. Educational services are the largest employers in the "High" and "Very High" regions, the second largest in the "Very Low," the third in "Low" areas, and the fourth in "Average" areas. The Accommodation and Food Services industry employs the second most people in the "Low" HII areas, the third most in "Average," and "High" HII areas, the fourth most in "Very High," and the fifth most in "Very Low" HII areas. The constructions industry is in

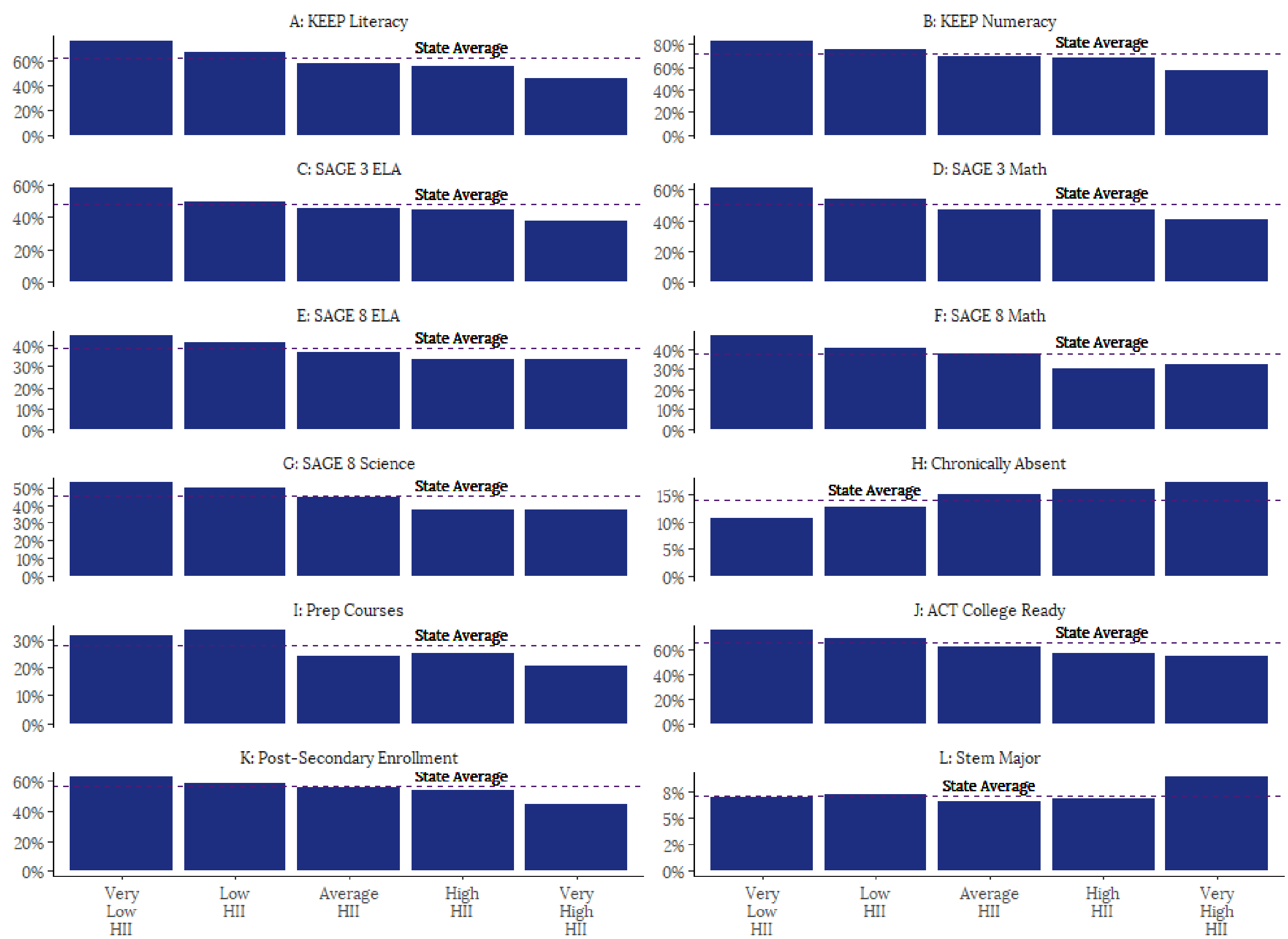

Figure 2: P20W metrics by HII group with state average

Table 1: Top five ndustries NAICS by employment volume by HII group

	1st	2nd	3rd	4th	5th
Very Low HII	62 - Health Care & Social Assistance	61 - Educational Services	23 - Construction	54 – Professional, Scientific & Technical Services	72 - Accommodation & Food Service
Low HII	62 - Health Care & Social Assistance	72 – Accommodation & Food Service	61 - Educational Services	23 - Construction	54 - Professional, Scientific & Technical Services
Average HII	62 - Health Care & Social Assistance	44 – Retail Trade	72 - Accommodation & Food Service	61 - Educational Services	23 - Construction
High HII	61 – Educational Services	62 - Health Care & Social Assistance	72 - Accommodation & Food Service	54 - Professional, Scientific & Technical Services	44 - Retail Trade

the top five for "Average," "Low," and "Very Low" areas; due to the nature of construction work not having a fixed job site and the large incidence of small scale contractors this should be taken that most contractors live in these small areas. Professional Scientific and Technical services are also present in the top five industries in "Very Low," "High," and "Very Low" areas. In the "Very High" areas, Administrative and Support, and Waste Management, and Remediation Services are in the top five employers, this points to potential environmental effects.

Demographics by Small Area systematically vary with HII, shown in Table 2. The "Very Low" HII areas are whitest while the Very High HII areas have the highest percentage of nonwhite residents. On average the "Very High" HII areas have 77% more minority residents than the "High" HII areas and 260% percent more than the "Very Low" HII areas. The percent of residents whose first language is not English increases as area deprivation increases. The "Very High" HII areas have 270% the amount of residents whose first language is not English, additionally the "Very High" HII areas have 54% more residents whose first language is not English.

Area deprivation is not equally distributed by geographic region type. The majority, 75%, of frontier areas are either "High" or "Very High" HII status additionally no frontier areas are "Very Low" HII status. The distribution of the HII group is more even for rural areas; 31% are "Low" HII and 31% are "High" HII while every other HII level each contributes 12.5%. In the Small Areas that are considered Urban 24% are "Very Low," 25% are "Low" and 20% are "Average." Urban areas have fewer "High," 16%, and "Very High," 15%, than Frontier Small Areas. This points to potentially different causes of area deprivation between urban, rural, and frontier areas.

Table 2: Student demographics by HII group.

	Very Low HII	Low HII	Average HII	High HII	Very High HII
Percent Minority	15.41%	21.69%	26.36%	31.67%	49.07%
Percent ELL	2.53%	5.43%	8.32%	10.41%	22.46%
Percent Low Income	17.82%	27.83%	41.61%	47.80%	60.44%
Percent Ever Homeless	1.28%	1.82%	2.62%	3.44%	4.16%

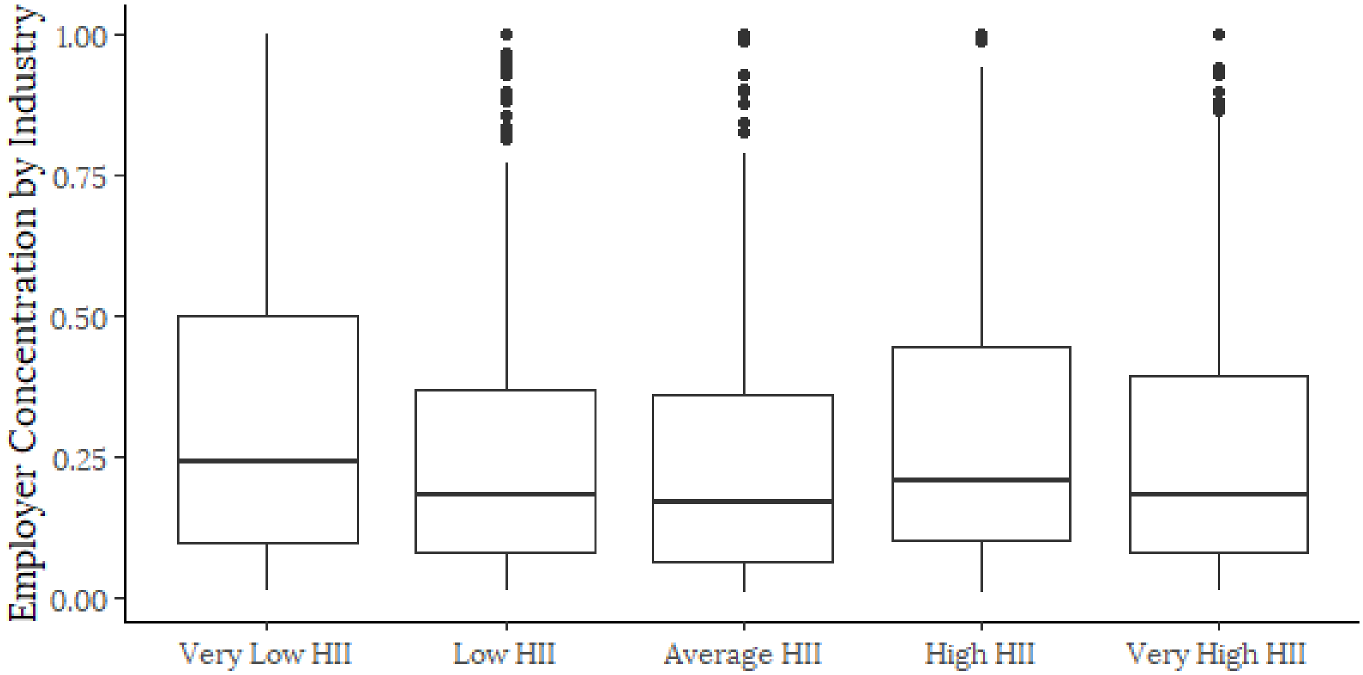

Figure 3: Industry concentration by HII group.

Figure 3 shows boxplots of different industry concentrations by the HII group. Employer concentration does not systematically vary with HII. All HII groups have a relatively equal mix of industries that are be considered monopsonies, and industries that are considered perfectly competitive. Additionally, about half the industries in each HII group are considered highly concentrated.

3.2 | Regression Results

Early education, post-secondary outcomes, and chronic absenteeism are negatively related to area deprivation; students living in the least deprived areas are most likely to have positive outcomes while those living in the most deprived areas are least likely to have positive outcomes. Figure 4 shows the predicted probabilities of outcomes for which the relationship between HII group and outcome is significant and monotonic for the modal Utahn. Full regression results are reported in Appendix B Tables 1 and 2, pairwise coefficient tests between HII groups are reported in Appendix B Table 3, and chronic absenteeism results in Appendix B Table 4.

Table 3 shows the percent difference in predicted probabilities for each outcome for every HII group compared to the "Very Low" group. The first two models are KEEP entry literacy and numeracy. Students that reside in the "Very Low" HII areas are 12% more likely than students in the "Very High" HII areas to be proficient in literacy entering kindergarten and 8% more likely to be prepared in numeracy. The overall pattern of the coefficients for each HII group shows the same monotonically decreasing relationship with HII as the unconditional proportions.

The probability of attending any post-secondary institution has the same negative relationship that appeared in the unconditional proportions. A resident of a "Very High" HII area is 15% less likely to attend either a degree-granting institution or technical college than a resident of a "Very High" HII area.

The relationship between chronic absenteeism and area deprivation is more complex than other outcomes. For students in the early grade levels and later grade levels, the highest HII areas have the highest probability of being chronically absent. In fourth, fifth, and sixth grade the relationship changes with the: "Very High" HII areas are less likely to be chronically absent than the "Very Low" areas. In kindergarten, a student from a "Very Low" HII area is 28% less likely to be chronically absent than a similar student in a "Very High" HII area. In eleventh grade, a student in a "Very High" HII is 25% more likely to be chronically absent than similar students in the "Very Low" HII Areas.

For all other outcomes, the trend between HII and outcome is not as straightforward. For some, the students from the "Very Low" HII group remain the most likely to have a positive outcome while for

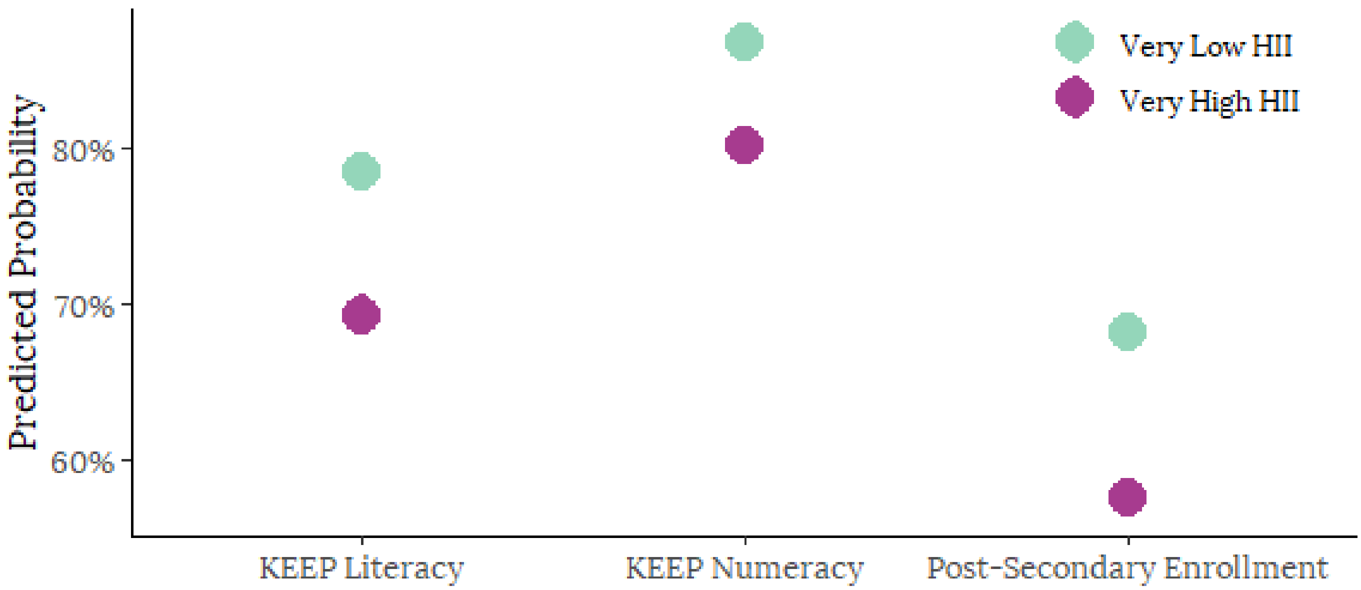

Figure 4: Predicted probabilities for P20W metrics that display a monotoinc reationship between HII and successHII group. Point estimate and 95% CI.

Table 3: Percent difference in predicted probablity of success (failure) for each P20W metric between Very Low HII and every other HII group.

	Low HII	Average HII	High HII	Very High HII
KEEP Literacy	-6%	-9%	-11%	-12%
KEEP Numeracy	-4%	-5%	-6%	-8%
SAGE 3 ELA	-9%	-7%	-6%	-6%
SAGE 3 Math	-6%	-7%	-7%	-5%
SAGE 8 ELA	-1%	-7%	-7%	9%
SAGE 8 Math	-7%	-7%	-17%	2%
SAGE 8 Science	0%	-4%	-8%	3%
Attempt Prep Courses	11%	-15%	-9%	-14%
ACT College Ready	-7%	-18%	-21%	-17%
Post-Secondary Enrollment	-3%	-4%	-4%	-15%

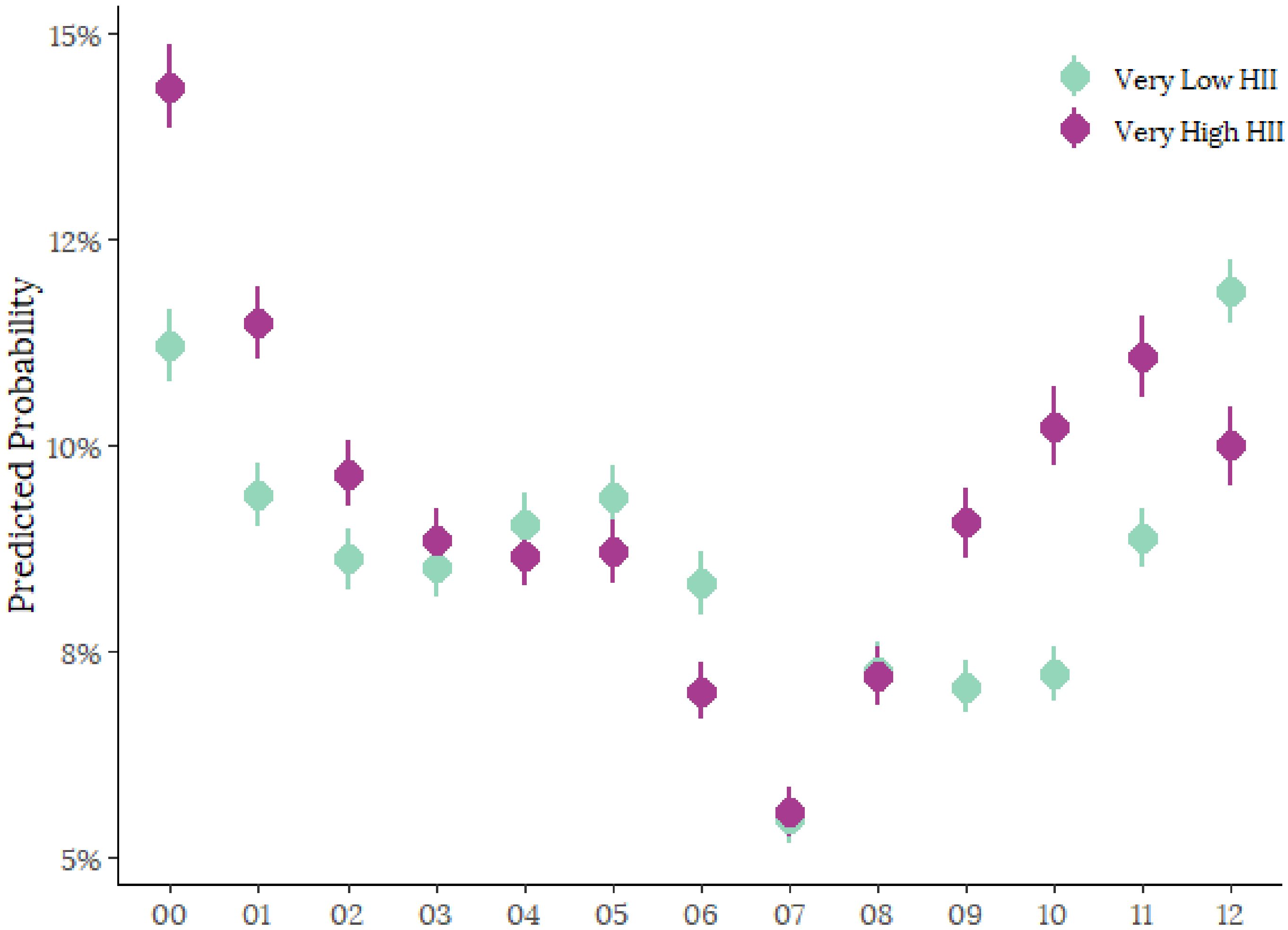

Figure 5: Predicted probability of chornic absenteeism by grade. Point estimate and 95% CI.

others students from the "Very High" HII group become most likely to have a positive outcome. Finally for taking any college preparatory course the students from the "Low" HII group are most likely to attempt a preparatory course.

There is no evidence of a relationship between employer concentration and area deprivation. The results of the regression are shown in Appendix B Table 4. The only variables with a significant relationship to area deprivation are demographics.

4 | Discussion

There is a clear relationship between area deprivation, as measured by HII, and kindergarten and post-secondary outcomes. This relationship is shown by a decreasing proportion of students testing proficient in KEEP scores and a lower proportion of graduates attending any post-secondary institution. This relationship is also evidenced by the probability of proficiency in KEEP scores and post-secondary attendance monotonically decreasing from the "Very Low" HII areas to the "Very High" HII areas. Additionally, chronic absenteeism in kindergarten through third grade and ninth through eleventh grade follows the same pattern after controlling for individual characteristics.

The findings on KEEP scores are consistent with the literature; students from the most deprived areas are least likely to be prepared for kindergarten. The magnitude found in this paper, 13% more likely to be proficient in literacy for a student from a "Very Low" compared to a "Very High" HII area is smaller than the findings of Wodtke and Parbst (2017) who find that a movement from the 20th percentile to 80th percentile of advantage leads to a 20% increase in the probability of proficiency. These are remarkably similar estimates given different measures of both advantage/deprivation and standardized tests.

By third grade, the relationship between area deprivation and both English and math proficiency is less pronounced. While students from the least deprived areas are most likely to be proficient on the SAGE tests the monotonic relationship seen with KEEP proficiency disappears. These results differ from previous research which finds a monotonic relationship where students from the most deprived area/s are more likely to have lower scores (Wodtke, Yildirim, Harding, & Elwart, 2020). This result is interesting and coincides with the switch in chronic absenteeism, where students from the "Very High" HII areas are least likely to be chronically absent.

The relationship between HII and post-secondary attendance is strongly negative and monotonic. This pattern is partially consistent with previous literature. These results are contradictory to Weinberg et al. (2019) who do not find a relationship but the results are qualitatively similar, though not directly comparable, to the results of Laliberté (2021). These results are interesting as the implication is that despite having a higher likelihood of being college ready based on the ACT they are less likely to attend college. This scenario points to the potential role that networks play in setting norms. It also points to affordability or perceived affordability as playing an important role. The effects of attending post-secondary institutions outside of USHE and the effects of ecclesiastical missions are discussed in the Limitations section.

The relationship between chronic absenteeism and area deprivation is more complex than other outcomes. In kindergarten, first grade, and ninth through eleventh grades students from the highest HII areas are most likely to be chronically absent. For several years in elementary school and junior high students from the highest HII areas are least likely to be chronically absent. There is not an a priori explanation for this phenomenon. It is possible that parents in the "Very High" HII areas have less access to child care and need to use schools and after-school activities as child care to be able to work longer hours without the added cost of childcare. It is also possible that in the "High" and "Very High" HII areas high school-aged students take on wage-earning activities or at-home care work which can interfere with attendance. The measure of chronically absent is all absences, not only unexcused. It is possible that the highest income families are willing to take their children out of school for various reasons and have access to outside tutoring and do not have to worry about their children falling behind.

The switch in those most likely to be chronically absent is not driven by an outlier school district. Alpine School District has the highest rate of chronic absenteeism in the State of Utah (U.S. Department of Education, 2019). Alpine school district also contains all or part of 21 different Small Areas. Of these 13 are "Very Low," or "Low" while eight are "Average," "High," or "Very High." Additional regressions that controlled for the Alpine SD had similar results to the reported regression. Additional regressions that only included students who did not switch schools during the school years were similar to the reported results. Due to the pooled cross-section nature of the data, it is not possible to determine if the switching is a general occurrence or if it is an artifact of the two years for which data was available.

The overall pattern of results suggests that schools in Utah can partially mitigate area deprivation for outcomes that involve a baseline level of skill and knowledge acquisition. This observation is evidenced with KEEP entrance scores which reflect zero mediation through schools and have a strong and negative relationship with HII. The pattern does not hold for eighth-grade SAGE proficiency. This phenomenon coincides with several years in which students from the "Very High" HII group are least likely to be chronically absent. Additionally despite the switch in chronic absenteeism students from the "Very High" HII areas are the third most likely to be college ready but the least likely to attend college. This finding points to the networks in Small Areas as a possible mediator. The relationship with Small Area does appear to dominate schools based on attempting any prep courses and post-secondary attendance. Given the cross-section nature of this data, it is not possible to know if this represents a real relationship or is an artifact of this cohort and data.

It was not possible to establish a relationship between the industrial structure of a Small Area and HII, the only significant association was between Small Area demographics and HII. The main implication of these results is that the driving force behind area deprivation is larger than just economics but is part of the political-economic-social ecology of the state.

4.1 | Limitations

There are several main limitations to this study. These limitations stem from what data are available, leading to imperfect matching to Small Areas, incomplete information about the student, and imperfect information about Small Areas. In the Salt Lake Metropolitan area, high school-age students who live in the Glendale Small Area, "Very High" HII, attend East High School which is in the Salt Lake Downtown Small Area, "High" HII. The matching method assigns these students to the incorrect Small Area and by extension, they are associated with the wrong HII. If increasing area deprivation is associated with worse outcomes, these students can bias the results of the area. Additionally, bussing information is not available at this time.

Beyond matching, the limited information about each student presented additional limitations. It is possible to control for low-income students with the provided binary variable but it is not possible to include family income or even high income. The lack of better income controls presents two limitations; first, it forces the use of binary outcome variables. Second, there is a higher proportion of high income families in "Very Low" and "Low" HII areas. High income is associated with higher test scores, and without being able to control for high income it will bias the size of the relationship between "Very Low" and "Low" HII areas and test scores.

USHE only governs public technical colleges and degree-granting institutions in Utah. This means that the data is not available for private universities or technical colleges in Utah or any out of state post-secondary institution. The main private post-secondary institutions in Utah are: Brigham Young University, Westminster College, Western Governors University, and Ensign College. If there is also a relationship between attending a private post-secondary institution or any out of state post-secondary institution and area deprivation, the inability to include students at private post-secondary institution may bias the results of this study.

In Utah the existence of a tight-knit religious majority, The Church of Jesus Christ of Latter-day Saints, complicates and potentially biases this research. Wards are assigned to members based on where those members live. The size of Small Areas may be physically larger than ward boundaries and ward boundaries may cross Small Area boundaries. The additional resources and networks that are available to members of the LDS Church may help mitigate certain aspects of living in the most deprived areas. The potential positive association between ward and outcome may mask part of the relationship between HII and P20W outcomes. Religious networks have been shown to mitigate the relationship between disadvantage and post-secondary education (Dehejia et al, 2009). It is also possible that wards can have a compounding effect between HII and P20W outcomes. In this case, the networks that exist in the Small Area are amplified in Wards increasing the positive association of living in the lowest HII areas and increasing the negative relationship between the highest HII areas and P20W metrics.

Additionally many members of the LDS Church serve a two-year ecclesiastic mission in early adulthood. This coincides with the time in which most high school graduates start college. It is possible that areas with a large concentration of LDS members can bias downwards post-secondary attendance immediately after high school if there is a relationship between area deprivation and the adherence to the LDS faith. For the 2018 cohort, this timeframe is especially important as there was only one school year between high school graduation and the measurement of post-secondary attendance. Additional Utah Specific studies that focus on the area relationships that are unable to control for the confounding effects of the dominant religion may suffer from biases.

The second part of this analysis sought to establish a relationship between the industrial structure and area deprivation. Neither the major sectors of employment nor the concentration of employers has a relationship with HII. This finding is despite employer concentration being negatively associated with wages, and wages are a component of HII. Small Area may be simultaneously too large and too small an area to calculate employer concentration. In Utah the average commute is 21.9 minutes (U. S. Census Bureau, 2020); with this average commute time for rural and frontier areas, it may not be possible to "commute" from one town to another for work in the same Small Area, while in urban areas it is entirely possible to commute across several Small Areas for work.

There are additional limitations from the QCEW data. First, in this analysis, QCEW is aggregated and the two-digit NAICS codes do not have the same level of detail as six-digit codes but allow for easier comparison across Small Areas. Second, using September data may overstate seasonal summer work. Finally, every firm is assigned a single NAICS though it is possible that a firm employs people for a core job and others for support positions that are in different industries. In the case of a construction firm, the firm may have a support employee who does administrative work but the employee is still considered in the NAICS two-digit code 23 despite not doing any construction work.

4.2 | Future Research

A relationship has been established between area deprivation and several education outcomes. This relationship only serves as a baseline as there are several methodological and data limitations that further studies should address. Previous research has been able to control for parental SES, which was not possible given the data available for this study, further research on area deprivation in Utah should control for parental SES. As more data becomes available to the UDRC it will be possible to address parental SES. Additionally, this study was unable to control for any sorting into Small Area where individual families sort into Small Area based on unobservable characteristics of the Small Area and the family; there may be "high achieving families" that sort into areas with other "high achieving" families. In this scenario, the relationships suffer from endogeneity and the estimation of the relationship between area deprivation and the outcomes of interest is inconsistent. To address this future studies should attempt to find an appropriate identification strategy for HII. This effort will allow for an interpretation that is closer to causal than the current paper.

5 | Conclusion

The evidence presented shows a clear negative relationship between HII and school readiness and post-secondary P20W metrics. This relationship disappears, becomes less pronounced, or partially flips for some measures during adolescence. These results coincide with shifts in the relationship between HII and chronic absenteeism. There is no clear relationship between HII and the industries and the employer concentration in Small Areas.

This research used a cross-section of testing records from USBE from the 2016/2017 and 2017/2018 school years along with registration records from USHE linked to UDOH Small Areas and HII to measure area deprivation. School readiness was defined as KEEP literacy and numeracy scores. Post-secondary attendance was defined as any graduate from a USBE high school in 2017 or 2018 who subsequently registered at a USHE technical college or degree-granting institution. Students from the "Very Low" HII areas are 13% more likely to be prepared in literacy and 8% more likely to be prepared for numeracy than students from the "Very High" HII areas. Students from the "Very Low" HII areas are 18% more likely than students from the "Very High" HII areas to attend any public post-secondary institution in Utah. Due to the nature of the data and methods used in this study these results do not represent a causal relationship between area deprivation and P20W outcomes.

These results add important information for Utah policymakers about areas for potential gains in understanding educational disparities. Despite the limitations of this research these results generally compare well to recent literature that establishes causal relations between areas of deprivation and various educational outcomes. Additionally, this study lays the groundwork for future research between P20W outcomes and the Health Improvement Index.

Data Partners & Acknowledgements

The author of this research would like to thank Karen Tao and Jeremias Solari for reviewing this research. The author would also like to thank Stephanie Su at the Utah State Board of Eduation, Srimoyee Bose, and Navina Forsythe at the Utah Department of Health, and Zachary Barrus at the Utah System of Higher Education for their reviews of and helpful feedback for this study. Finally the author would like to thank Aaron Brough at the Utah State Board of Education, Laura Zemp at the Utah System of Higher Education, Dulce Diez and Michael Friedricks at the Utah Department of Health and Stephen Matherly at Utah's Early Childhood Integrated Data System for early feedback on the project.

References

Allen, J., & Radunzel, J. (2017). What are the ACT College Readiness Benchmarks? ACT Research & Policy. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/pdfs/R1670-college-readiness-benchmarks-2017-11.pdf

Azar, J., Marinescu, I., & Steinbuam, M. (2020). Labor Market Concentration. Journal of Human Resources. doi:10.3368/jhr.monopsony.1218-9914R1

Bayer, P., Ross, S. L., & Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. Journal of Poltical Economy, 116(6), 1150-1196.

Dehejia, R., DeLeire, T., Luttmer, E. F., & Mitchell, J. (2009). 8. The Role of Religious and Social Organizations in the Lives of Disadvantaged Youth (pp. 237-274). University of Chicago Press.

Fan, S. J., Heinrich, J., Bloom, M. S., Zhao, T. Y., Shi, T. X., Feng, W. R., ... & Dong, G. H. (2020). Ambient air pollution and depression: a systematic review with meta-analysis up to 2019. Science of The Total Environment, 701, 134721.

Ferguson, N. T., & Michaelsen, M. M. (2015). Money changes everything? Education and regional deprivation revisted. Economics of Education Review, 48, 129-147.

Gohel, D. (2021). flextable: Functions for Tabular Reporting. https://CRAN.R-project.org/package=flextable

Heckman, J. J., Humphries, J. E., & Veramendi, G. (2018). Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking. Journal of Political Economy, 126(S1), S197-S246. doi:10.1086/698760

Hershbein, B., Macaluso, C., & Yeh, C. (2020). Monopsony in the US labor market. Federal Reserve Bank of Richmond Working Paper..

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. Biometrical Journal, 50(3), 346–363.

Kassambara, A. (2020). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. https://CRAN.R-project.org/package=rstatix.

Laliberté, J.-W. (2021). Long-term contextual effects in education: Schools and neighborhoods. American Economic Journal: Economic Policy, 13(2), 336-77.

Lewis, N. M., Friedrichs, M., Wagstaff, S., Sage, K., LaCross, N., Bui, D., . . . Dunn, A. (2020). Disparities in COVID-19 Incidence, Hospitalization, and Testing, by Area-Level Deprivation--- Utah, March 3-July 9, 2020. Morbidity and Mortality Weekly Report, 69(38), 1369-1373.

Ludwig, J., Duncan Greg J., G. L., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2012). Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults. Science, 337(6101), 1505-1510.

Orellano, P., Quaranta, N., Reynoso, J., Balbi, B., & Vasquez, J. (2017). Effect of outdoor air pollution on asthma exacerbations in children and adults: systematic review and multilevel meta-analysis. PloS one, 12(3), e0174050.

R Core Development Team. (2019). R: A Language and Environment for Statistical Computing (3.6.0). R Foundation for Statistical Computing. URL https://www.R-project.org/.
Robinson, D., Hayes, A., & Couch, S. (2020). broom: Convert Statistical Objects into Tidy Tibbles. https://CRAN.R-project.org/package=broom
Scott, S. (2019). Education Appropriations' Return on Investment of Career and Technical Education Provided by the Utah System of Technical Colleges. Utah Data Research Center.
Scott, S. (2020). Increased Spending from Post-Secondary Graduates in Utah (2011-2018). Utah Data Research Center.
Shah, A. S., Langrish, J. P., Nair, H., McAllister, D. A., Hunter, A. L., Donaldson, K., ... & Mills, N. L. (2013). Global association of air pollution and heart failure: a systematic review and meta-analysis. The Lancet, 382(9897), 1039-1048.
Shah, A. S., Lee, K. K., McAllister, D. A., Hunter, A., Nair, H., Whiteley, W., ... & Mills, N. L. (2015). Short term exposure to air pollution and stroke: systematic review and meta-analysis. bmj, 350.
Suhrcke, M., & de Paz Nieves, C. (2011). The impact of health and health behaviours on educational outcomes.
U. S. Census Bureau. (2020). Utah. Retrieved from QuickFacts: https://www.census.gov/quickfacts/UT
U.S. Department of Education. (2019). CHRONIC ABSENTEEISM IN THE NATION'S SCHOOLS: A hidden educational crisis. Retrieved from https://www2.ed.gov/datastory/chronicabsenteeism.html#four
US Census Bureau. (2021). Relationship Files. Retrieved from Census Reference Files: https://www2.census.gov/geo/docs/maps-data/data/rel/zcta_tract_rel_10.txt.
Utah Data Research Center. (2020, March). P-20 Metrics: Utah's Education to Workforce Pipeline. Retrieved from Utah Data Research Center: https://udrc.utah.gov/p20/Index.html
Utah Department of Health. (2020). Complete Health Indicator Report of Utah Health Improvement Index (HII). Retrieved from Utah Department of Health: https://ibis.health.utah.gov/ibisph-view/indicator/complete_profile/HII.html
Utah Department of Health. (2021). Utah Small Area Boundary Definitions. Retrieved from IBIS-PH Methodology Guidelines: https://ibis.health.utah.gov/ibisph-view/pdf/resource/UtahSmallAreaInfo.pdf
van der Noordt, M., IJzelenberg, H., Droomers, M., & Proper, K. I. (2014). Health effects of employment: a systematic review of prospective studies. Occupational and environmental medicine, 71(10), 730-736. doi:10.1136/oemed-2014-102143
Vowles, M., Kerry, R., Ingram, B., & Mason, L. (2020). Investigation of the Environmental and Socio-Economic Characteristics of Counties with a High Asthma Burden to Focus Asthma Action in Utah. International Journal of Environmental Research and Public Health, 17(14). doi:10.3390/ijerph17145251.
Weinberg, D., Gonneke, S. W., Finkenauer, C., Brunekreef, B., & Smit, H. A. (2019). The pathways from parental and neighbourhood socioeconomic status to adolescent educational attainment: An examination of the role of cognitive ability, teacher assessment, and educational expectations. PLOS ONE, 14(5). doi:10.1371/journal.pone.0216803.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686.
Wodtke, G. T., Yildirim, U., Harding, D. J., & Elwart, F. (2020). Are Neighborhood Effects Explained by Differences in School Quality? IRLE Working Paper No 102-20. Retrieved from http://irle.berkeley.edu/files/2020/05/Are-Neighborhood-Effects-Explained-by-Differences-in-School-Quality.pdf.
Wodtke, G., & Parbst, M. (2017). Neighborhoods, Schools, and Academic Achievement: A Formal Mediation Analysis of Contextual Effects on Reading and Mathematics Abilities. Demography, 54, 1653-1676. doi:10.1007/s13524-017-0603-1.
Yang, B. Y., Fan, S., Thiering, E., Seissler, J., Nowak, D., Dong, G. H., & Heinrich, J. (2020). Ambient air pollution and diabetes: a systematic review and meta-analysis. Environmental research, 180, 108817.
Yang, B. Y., Qian, Z., Howard, S. W., Vaughn, M. G., Fan, S. J., Liu, K. K., & Dong, G. H. (2018). Global association between ambient air pollution and blood pressure: a systematic review and meta-analysis. Environmental pollution, 235, 576-588.

Appendix A | Additional Descriptive Tables

Table A1: Comparison of propotions of P20Wmetrics between HII groups and the state average. Asterisks indicate significant p-values levels. (* p<0.05, ** p<0.01, *** p<0.001).

	Very Low HII	Low HII	Average HII	High HII	Very High HII	State Average
KEEP Literacy	75.23***	66.06***	57.36***	54.68***	45.06***	61.28
KEEP Numeracy	82.88***	75.78***	69.33***	67.45***	56.88***	71.82
SAGE 3 ELA	57.62***	49.55	45.34***	44.27***	37.76***	47.62
SAGE 3 Math	60.75***	53.57	47.07***	46.63***	40.09***	50.51
SAGE 8ELA	44.95***	41.67	37.13***	33.36***	33.66***	38.72
SAGE 8 MATH	46.99***	40.62**	38.05***	30.15***	32.21***	38.24
SAGE 8 Science	52.49***	49.81	44.18***	37.6***	37.2***	45.05
Attempt Prep Course	31.46***	33.3***	24.16***	25.12***	20.62***	27.77
ACT College Ready	75.92***	69.54***	62.58***	56.57***	54.06***	65.42
Chronic Absenteeism	11.75***	13.59**	14.01***	14.76***	15.18***	13.99
Post-Secondary Enrollment	62.74***	58.52***	55.54*	53.8***	44.37***	56.3
STEM Major	6.99	7.19	6.61	6.85	8.89***	7.11

Table A2: Pairwise comparison of difference of propotions of P20Wmetrics between HII groups. Asterisks indicate significant p-values levels. (* p<0.05, ** p<0.01, *** p<0.001).

	KEEP Literacy	KEEP Numeracy	SAGE 3 ELA	SAGE 3 Math	SAGE 8 ELA	SAGE 8 Math	SAGE 8 Science	Attempt Prep Course	ACT College Ready	Chronic Absen-teeism	Post-Sec-ondary Enroll-ment	STEM Major
Very Low HII-Low HII	0.092***	0.071***	0.081***	0.072***	0.033*	0.064***	0.027*	-0.018***	0.064***	-0.019***	0.042***	-0.002
Very Low HII-Average HII	0.179***	0.136***	0.123***	0.137***	0.078***	0.089***	0.083***	0.073***	0.133***	-0.042***	0.072***	0.004
Low HII-Average HII	0.087***	0.065***	0.042***	0.065***	0.045***	0.026**	0.056***	0.091***	0.07***	-0.023***	0.03***	0.006
Very Low HII-High HII	0.205***	0.154***	0.134***	0.141***	0.116***	0.168***	0.149***	0.063***	0.194***	-0.052***	0.089***	0.001
Low HII-High HII	0.114***	0.083***	0.053***	0.069***	0.083***	0.105***	0.122***	0.082***	0.13***	-0.033***	0.047***	0.003
Average HII-High HII	0.027	0.019	0.011	0.004	0.038	0.079***	0.066***	-0.01***	0.06**	-0.01***	0.017*	-0.002
Very Low HII-Very High HII	0.302***	0.26***	0.199***	0.207***	0.113***	0.148***	0.153***	0.108***	0.219***	-0.064***	0.184***	-0.019**
Low HII-Very High HII	0.21***	0.189***	0.118***	0.135***	0.08***	0.084***	0.126***	0.127***	0.155***	-0.045***	0.141***	-0.017**
Average HII-Very High HII	0.123***	0.125***	0.076***	0.07***	0.035	0.058***	0.07***	0.035***	0.085***	-0.022***	0.112***	-0.023***
High HII-Very High HII	0.096***	0.106***	0.065***	0.065***	-0.003	-0.021	0.004	0.045***	0.025	-0.012***	0.094***	-0.02**

Appendix B | Regression Results

Table B1: Probit regression results for K-8 P20W outcomes, point estimate and 95% CI in parantheses. Asterisks indicate significant p-values levels. (* p<0.05, ** p<0.01, *** p<0.001).

	KEEP Literacy	KEEP Numeracy	SAGE 3 ELA	SAGE 3 Math	SAGE 8 ELA	SAGE 8 Math	SAGE 8 Science
Intercept	-2.94*** (-3.16, -2.71)	-3.54*** (-3.78, -3.31)	-0.52*** (-0.59, -0.45)	-0.16*** (-0.23, -0.1)	-1.44*** (-1.57, -1.31)	-1.11*** (-1.23, -0.98)	-0.83*** (-0.93, -0.72)
Low HII	-0.16*** (-0.19, -0.12)	-0.15*** (-0.18, -0.11)	-0.13*** (-0.16, -0.09)	-0.1*** (-0.14, -0.07)	-0.01 (-0.06, 0.03)	-0.09*** (-0.14, -0.05)	0.01 (-0.03, 0.05)
Average HII	-0.23*** (-0.27, -0.19)	-0.19*** (-0.23, -0.15)	-0.1*** (-0.14, -0.06)	-0.13*** (-0.17, -0.09)	-0.07** (-0.12, -0.03)	-0.1*** (-0.14, -0.05)	-0.07** (-0.12, -0.02)
High HII	-0.27*** (-0.32, -0.23)	-0.22*** (-0.26, -0.17)	-0.09*** (-0.14, -0.05)	-0.12*** (-0.16, -0.07)	-0.07** (-0.12, -0.02)	-0.22*** (-0.27, -0.17)	-0.12*** (-0.17, -0.07)
Very High HII	-0.28*** (-0.33, -0.24)	-0.26*** (-0.31, -0.22)	-0.09*** (-0.13, -0.04)	-0.1*** (-0.14, -0.05)	0.1*** (0.05, 0.15)	0.02 (-0.03, 0.07)	0.04 (-0.01, 0.09)
Hispanic	-0.4*** (-0.44, -0.37)	-0.36*** (-0.4, -0.32)	-0.27*** (-0.31, -0.22)	-0.34*** (-0.39, -0.3)	-0.37*** (-0.42, -0.33)	-0.47*** (-0.52, -0.43)	-0.45*** (-0.5, -0.41)
Multiracial	-0.06 (-0.13, 0.01)	-0.05 (-0.12, 0.02)	-0.01 (-0.09, 0.06)	-0.1** (-0.18, -0.03)	-0.08 (-0.18, 0.01)	-0.14** (-0.23, -0.04)	-0.09 (-0.18, 0)
Black	-0.34*** (-0.45, -0.24)	-0.36*** (-0.46, -0.25)	-0.55*** (-0.66, -0.43)	-0.58*** (-0.69, -0.47)	-0.58*** (-0.72, -0.44)	-0.74*** (-0.89, -0.6)	-0.68*** (-0.81, -0.54)
Pacific Islander	-0.52*** (-0.61, -0.43)	-0.44*** (-0.53, -0.34)	-0.44*** (-0.55, -0.33)	-0.48*** (-0.59, -0.38)	-0.44*** (-0.55, -0.32)	-0.4*** (-0.52, -0.28)	-0.71*** (-0.83, -0.59)
Native American	-0.61*** (-0.74, -0.47)	-0.56*** (-0.69, -0.43)	-0.66*** (-0.8, -0.52)	-0.67*** (-0.8, -0.53)	-0.48*** (-0.63, -0.34)	-0.44*** (-0.59, -0.3)	-0.53*** (-0.67, -0.39)
Asian	0.21*** (0.11, 0.31)	0.06 (-0.05, 0.16)	0.08 (-0.02, 0.18)	0.13* (0.02, 0.23)	0.24*** (0.14, 0.34)	0.33*** (0.23, 0.43)	0.16** (0.07, 0.26)
Female	0.1*** (0.08, 0.13)	-0.01 (-0.03, 0.02)	0.25*** (0.23, 0.28)	-0.1*** (-0.12, -0.07)	0.34*** (0.31, 0.37)	0.04* (0.01, 0.07)	-0.15*** (-0.18, -0.12)
Ever Homeless	-0.4*** (-0.5, -0.31)	-0.24*** (-0.33, -0.15)	-0.29*** (-0.38, -0.19)	-0.2*** (-0.29, -0.1)	-0.27*** (-0.38, -0.15)	-0.41*** (-0.53, -0.29)	-0.34*** (-0.45, -0.23)
Low Income	-0.55*** (-0.58, -0.53)	-0.47*** (-0.5, -0.44)	-0.42*** (-0.45, -0.39)	-0.39*** (-0.42, -0.36)	-0.43*** (-0.46, -0.39)	-0.43*** (-0.46, -0.39)	-0.42*** (-0.45, -0.38)
ELL N	0.49*** (0.44, 0.54)	0.56*** (0.51, 0.61)	0.7*** (0.64, 0.76)	0.59*** (0.54, 0.65)	1.28*** (1.15, 1.4)	1.15*** (1.03, 1.27)	1.1*** (1, 1.2)
ELL F	0.82*** (0.56, 1.09)	0.89*** (0.61, 1.17)	1.49*** (1.39, 1.58)	1.3*** (1.2, 1.39)	1.19*** (1, 1.39)	1.06*** (0.86, 1.25)	1.02*** (0.84, 1.2)
ELL O	0.66* (0.1, 1.22)	1** (0.37, 1.64)	0.35 (-0.18, 0.87)	-0.09 (-0.65, 0.47)	0.99* (0.11, 1.86)	0.58 (-0.46, 1.62)	0.37 (-0.66, 1.4)
Frontier	0.07* (0, 0.14)	0 (-0.08,0.07)	0.04 (-0.05, 0.14)	0.08 (-0.01, 0.18)	0.2*** (0.09, 0.31)	0.37*** (0.26, 0.48)	0.19*** (0.09, 0.3)
Rural	-0.14*** (-0.17, -0.1)	-0.09*** (-0.13, -0.06)	-0.03 (-0.06, 0.01)	0.01 (-0.02, 0.05)	0.05* (0, 0.09)	0.1*** (0.06, 0.14)	0.08*** (0.04, 0.12)
School Year 2018			-1.22** (-1.97, -0.47)	-1.41*** (-2.17, -0.65)	-4.16 (-33.79,25.48)	-0.63 (-1.39, 0.12)	-0.52 (-1.19, 0.16)

Table B2: Probit regression results for Post-Secondary P20W outcomes, point estimate and 95% CI in parantheses. Asterisks indicate significant p-values levels. (* p<0.05, ** p<0.01, *** p<0.001).

	Attempt Prep Courses	ACT College Ready	Post-Secondary Enrollment	STEM Major
Intercept	0.05 (0, 0.1)	0.28** (0.1,0.46)	0.33*** (0.25,0.41)	-1.01*** (-1.22,-0.81)
Low HII	0.12*** (0.1, 0.13)	-0.1*** (-0.14,-0.06)	-0.06*** (-0.09,-0.04)	0.01 (-0.04,0.06)
Average HII	-0.16*** (-0.18, -0.15)	-0.26*** (-0.31,-0.22)	-0.08*** (-0.1,-0.06)	-0.01 (-0.06,0.04)
High HII	-0.1*** (-0.11, -0.08)	-0.31*** (-0.36,-0.27)	-0.07*** (-0.09,-0.05)	-0.01 (-0.06,0.05)
Very High HII	-0.16*** (-0.18, -0.14)	-0.25*** (-0.3,-0.19)	-0.28*** (-0.31,-0.25)	0.11** (0.04,0.18)
Hispanic	-0.13*** (-0.14, -0.11)	-0.61*** (-0.66,-0.57)	-0.29*** (-0.31,-0.27)	0.05 (-0.01,0.11)
Multiracial	0 (-0.03, 0.03)	-0.04 (-0.14,0.05)	0 (-0.04,0.04)	0 (-0.11,0.11)
Native American	-0.49*** (-0.55, -0.44)	-0.74*** (-0.89,-0.59)	-0.39*** (-0.46,-0.32)	0.17 (-0.02,0.36)
Pacific Islander	-0.31*** (-0.35, -0.27)	-0.54*** (-0.65,-0.43)	-0.56*** (-0.62,-0.5)	-0.22* (-0.42,-0.01)
Asian	0.33*** (0.3, 0.36)	-0.11* (-0.21,0)	0.28*** (0.23,0.33)	0.46*** (0.37,0.55)
Black	-0.38*** (-0.42, -0.34)	-0.79*** (-0.92,-0.67)	-0.35*** (-0.42,-0.29)	0.08 (-0.08,0.25)
Female	0.12*** (0.11, 0.13)	0.1*** (0.07,0.13)	0.25*** (0.23,0.26)	-0.38*** (-0.41,-0.35)
Gender U	-3.43 (-15.38, 8.51)			
Ever Homeless	-0.34*** (-0.37, -0.3)	-0.3*** (-0.41,-0.2)	-0.27*** (-0.32,-0.22)	0.05 (-0.08,0.17)
Low Income	-0.4*** (-0.41, -0.38)	-0.49*** (-0.52,-0.45)	-0.38***(-0.4,-0.36)	0.01 (-0.03,0.05)
ELL N	-0.05 (-0.1, 0.01)	0.6*** (0.42,0.77)	0.14*** (0.06,0.22)	-0.18 (-0.38,0.03)
ELL O	-0.32*** (-0.5,-0.14)	-0.39 (-1.27,0.48)	-0.22 (-0.48,0.05)	-0.07 (-0.85,0.72)
ELL Y	-0.65*** (-0.71,-0.59)	-1.04*** (-1.28,-0.8)	-0.53*** (-0.62,-0.44)	-0.14 (-0.39,0.12)
Frontier	0.18*** (0.14,0.22)	0 (-0.12,0.13)	0.5*** (0.45,0.56)	-0.08 (-0.2,0.05)
Rural	-0.14*** (-0.16,-0.13)	-0.23*** (-0.27,-0.19)	0.08*** (0.06,0.1)	-0.06* (-0.12,-0.01)
School Year 2018	0.04*** (0.03,0.05)	-0.88*** (-1.08,-0.68)	-0.32*** (-0.34,-0.31)	-0.25*** (-0.29,-0.22)
9th Grade	-1.14*** (-1.15,-1.12)			
10th Grade	-0.74*** (-0.76,-0.73)	3.15 (-42.31,48.62)		
11th Grade	-0.21*** (-0.22,-0.2)	-0.67** (-1.07,-0.27)		

Table B3: Pairwise comparision of HII group regression coefficients. Asterisks indicate significant p-values levels. (* p<0.05, ** p<0.01, *** p<0.001).

	KEEP Literacy	KEEP Numeracy	SAGE 3 ELA	SAGE 3 Math	SAGE 8 ELA	SAGE 8 Math	SAGE 8 Science	Attempt Prep Course	ACT College Ready	Post-Secondary Enrollment	STEM Major
Low HII - Very Low HII	-0.158***	-0.146***	-0.126***	-0.104***	-0.015	-0.094***	0.007	0.115***	-0.102***	-0.065***	0.011
Average HII - Very Low HII	-0.232***	-0.193***	-0.096***	-0.133***	-0.075*	-0.095***	-0.07*	-0.163***	-0.262***	-0.08***	-0.009
High HII - Very Low HII	-0.274***	-0.215***	-0.091***	-0.118***	-0.073*	-0.219***	-0.124***	-0.096***	-0.312***	-0.072***	-0.005
Very High HII - Very Low HII	-0.283***	-0.264***	-0.088**	-0.098***	0.102***	0.024	0.042	-0.158***	-0.249***	-0.278***	0.107*
Average HII - Low HII	-0.074***	-0.047	0.03	-0.029	-0.06	-0.002	-0.077**	-0.279***	-0.16***	-0.016	-0.02
High HII - Low HII	-0.117***	-0.07**	0.035	-0.014	-0.058	-0.125***	-0.131***	-0.211***	-0.209***	-0.007	-0.017
Very High HII - Low HII	-0.125***	-0.118***	0.038	0.006	0.116***	0.117***	0.035	-0.274***	-0.147***	-0.214***	0.096*
High HII - Average HII	-0.042	-0.023	0.005	0.015	0.002	-0.124***	-0.054	0.067***	-0.049	0.009	0.004
Very High HII - Average HII	-0.051	-0.071**	0.009	0.034	0.176***	0.119***	0.111***	0.005	0.013	-0.198***	0.116**
Very High HII - High HII	-0.008	-0.049	0.003	0.02	0.174***	0.243***	0.166***	-0.062***	0.062	-0.207***	0.113*

Table B4: Probit regression results for chronic absenteeism, point estimate and 95% CI in parantheses. Colons represent interaction terms. Asterisks indicate significant p-values levels. (* p<0.05, ** p<0.01, *** p<0.001).

Intercept	-1.25*** (-1.27,-1.22)
Low HII	-0.09*** (-0.12,-0.07)
Average HII	0 (-0.03,0.03)
High HII	0.04** (0.02,0.07)
Very High HII	-0.1*** (-0.13,-0.07)
Kindergarten	-0.03* (-0.06,0)
1st Grade	-0.14*** (-0.17,-0.11)
2nd Grade	-0.18*** (-0.21,-0.15)
3rd Grade	-0.19*** (-0.22,-0.16)
4th Grade	-0.16*** (-0.19,-0.13)
5th Grade	-0.14*** (-0.17,-0.11)
6th Grade	-0.2*** (-0.23,-0.17)
7th Grade	-0.42*** (-0.45,-0.39)
8th Grade	-0.28*** (-0.31,-0.24)
9th Grade	-0.29*** (-0.32,-0.26)
10th Grade	-0.28*** (-0.31,-0.25)
11th Grade	-0.17*** (-0.2,-0.14)
Hispanic	0.15*** (0.14,0.16)
Multiracial	0.12*** (0.11,0.14)
Black	0.12*** (0.1,0.14)
Pacific Islander	0.33*** (0.31,0.35)
Native American	0.4*** (0.38,0.42)
Asian	-0.19*** (-0.21,-0.17)

Female	0 (0,0.01)
Gender U	0.38 (-0.08,0.83)
Gender O	-1.75 (-10.3,6.81)
Ever Homeless	0.48*** (0.47,0.5)
Low Income	0.42*** (0.41,0.42)
ELL N	0.07*** (0.05,0.08)
ELL F	-0.24*** (-0.27,-0.22)
ELL O	0.1* (0.01,0.2)
Frontier	-0.07*** (-0.09,-0.05)
Rural	0.07*** (0.06,0.08)
Low HII:Kindergarten	0.17*** (0.13,0.21)
Average HII:Kindergarten	0.11*** (0.07,0.15)
High HII:Kindergarten	0.13*** (0.09,0.17)
Very High HII:Kindergar-ten	0.25*** (0.2,0.3)
Low HII:1st Grade	0.12*** (0.08,0.16)
Average HII:1st Grade	0.06** (0.02,0.1)
High HII:1st Grade	0.05* (0.01,0.1)
Very High HII:1st Grade	0.22*** (0.17,0.26)
Low HII:2nd Grade	0.1*** (0.06,0.14)
Average HII:2nd Grade	0.05* (0.01,0.09)
High HII:2nd Grade	0.02 (-0.02,0.06)
Very High HII:2nd Grade	0.16*** (0.12,0.21)
Low HII:3rd Grade	0.11*** (0.07,0.15)

Table B4 Continued: Probit regression results for chronic absenteeism, point estimate and 95% CI in parantheses. Colons represent interaction terms. Asterisks indicate significant p-values levels. (* p<0.05, ** p<0.01, *** p<0.001).

Average HII:3rd Grade	0.03 (-0.01,0.07)
High HII:3rd Grade	-0.03 (-0.07,0.01)
Very High HII:3rd Grade	0.12*** (0.08,0.17)
Low HII:4th Grade	0.07*** (0.03,0.11)
Average HII:4th Grade	0.01 (-0.03,0.05)
High HII:4th Grade	-0.07*** (-0.12,-0.03)
Very High HII:4th Grade	0.07** (0.03,0.12)
Low HII:5th Grade	0.1*** (0.06,0.14)
Average HII:5th Grade	0.01 (-0.03,0.05)
High HII:5th Grade	-0.02 (-0.06,0.02)
Very High HII:5th Grade	0.06* (0.01,0.1)
Low HII:6th Grade	0.14*** (0.1,0.18)
Average HII:6th Grade	0.04 (0,0.08)
High HII:6th Grade	-0.06** (-0.1,-0.01)
Very High HII:6th Grade	0.01 (-0.04,0.06)
Low HII:7th Grade	0.22*** (0.18,0.26)
Average HII:7th Grade	0.14*** (0.1,0.19)
High HII:7th Grade	0.02 (-0.03,0.07)
Very High HII:7th Grade	0.11*** (0.06,0.16)
Low HII:8th Grade	0.17*** (0.13,0.21)
Average HII:8th Grade	0.06* (0.01,0.1)
High HII:8th Grade	-0.02 (-0.06,0.03)
Very High HII:8th Grade	0.09*** (0.05,0.14)
Low HII:9th Grade	0.24*** (0.2,0.28)
Average HII:9th Grade	0.13*** (0.09,0.17)
High HII:9th Grade	0.08*** (0.03,0.12)
Very High HII:9th Grade	0.23*** (0.19,0.28)
Low HII:10th Grade	0.13*** (0.09,0.17)
Average HII:10th Grade	0.12*** (0.08,0.16)
High HII:10th Grade	0.13*** (0.09,0.17)
Very High HII:10th Grade	0.29*** (0.24,0.34)
Low HII:11th Grade	0.1*** (0.06,0.14)
Average HII:11th Grade	0.06** (0.02,0.1)
High HII:11th Grade	0.08*** (0.04,0.12)
Very High HII:11th Grade	0.23*** (0.18,0.27)

Table B5: OLS Regression for HII. Asterisks indicate significant p-values levels. (* p<0.05, ** p<0.01, *** p<0.001).

Intercept	77.69*** (62.78, 92.6)
Rural	0.24 (-18.61, 19.08)
Urban	-29.45*** (-40.31, -18.59)
%Minority	0.69*** (0.4, 0.98)
%Collage Age	1.42*** (1.09, 1.74)
%Over 65	1.07*** (0.48, 1.66)
HHI	-11.03 (-27.97, 5.91)
Rural*%Minority	-0.84 (-2.07, 0.4)
Urban*%Minority	0.5** (0.16, 0.84)