HERC LONGITUDINAL DATA CODEBOOK: CAMPUS DATA
INTRODUCTION

This codebook accompanies the HERC (Houston Education Research Consortium) longitudinal database, which includes student, school, and teacher-level data from the Houston Independent School District (HISD). This database is a result of a data sharing agreement between HISD and HERC, a partnership between HISD and Rice University. The partnership aims to produce rigorous research for the purpose of closing the socioeconomic gaps in educational achievement and attainment. The data are drawn from sources provided by HISD, the Texas Education Agency (TEA), and the Common Core of Data (CCD) from the National Center for Education Statistics (NCES). The dataset spans six years, from the 2007-08 school year through the 2012-13 school year, and it includes the population of students enrolled in HISD schools. This codebook focuses on campus-level data on HISD schools.

Data Organization

HERC has gathered data from multiple sources across diverse domains regarding the education of students in HISD. The campus-level data include information on student and staff demographics, rates of program participation, school magnet programs, college readiness, graduation and school completion, and campus accountability (based on state-required standardized tests STAAR and TAKS).

The purpose of the HERC longitudinal database is to integrate data from multiple sources into one database, with the goal of facilitating longitudinal and multilevel research that is best for producing rigorous, academic research that can also be useful to district decision-makers. To assist the merging of campus data to student or teacher measures, the campus-level data are separated by year and by source, since students and teachers can change campuses between years. HERC provides guidance on how to merge campus data to student or teacher data.

Variable Naming Strategy

HERC follows a variable naming strategy so that users can understand the source and year of the variable. The first one or two characters indicate the source of the data:

c	TEA school profiles
ca	Campus aggregated variables (Based on PEIMS)
ec	CCD school universe
mp	Magnet program
fl	Tracking or analytic sample flag

The last two digits indicate the year, based on the spring semester of the academic year: 08 for the 2007-08 academic year, 13 for the 2012-13 academic year, etc. Variable labels give full descriptions of each variable. For example:

ectype08	School Type, CCD 07-08
mpprglang09	Language Program, Magnet Schools, 08-09
cathlen13	Total Students- Home Language= English, PEIMS Aggregated Campus Vars, 2012-13

Some categorical variables also contain value labels explaining each numerical category. Value labels start with the prefix L, and then the name of the variable to which it is linked. For example, value label Lmpprgtype12 is linked to variable mpprgtype12, which is School Program Type, Magnet Schools, 2011-12. Each value label contains the numerical value of the category. For example:

```
. tab mpprgtype12, missing
```

School Program Type, Magnet Schools, 2011-12	Freq.	Percent	Cum.
1.SWAS	53	46.90	46.90
2.SWP	33	29.20	76.11
3.SWVP	15	13.27	89.38
4.SUS	10	8.85	98.23
5.SWVS	2	1.77	100.00
Total	113	100.00	

Data Sources

Magnet Program File
This is a school-level file of all schools in HISD that have magnet programs intended to attract non-zoned students with an enriched curriculum centered around a specialized theme. (Note: This file does not include all HISD schools, only those with an HISD magnet program. Each HISD school only has one magnet program.) It contains variables indicating magnet program type (school-within-a-school, schoolwide program, etc.), as well as program description, based on the content of their magnet programs.

HISD schools create their own program descriptions; when cleaning these data, HERC endeavored to collapse each unique program description into categories based on common characteristics. Program descriptions are combined into categories that follow those outlined in the 2013-14 HISD Magnet Program Catalog. Some schools had program descriptions that seemed to fit more than one category; in those cases, we counted the school in more than one category, even though it is considered to have one magnet program. The variable mpprgcountYR is a count of the number of categories in which a school's program fits; most schools fit only one, but some fit two categories. These categories are constructed by HERC, not by HISD, and

are not intended to suggest that a school has more than one program, simply that a school's program fits more than one descriptive category. See HISD School Choice Options: 2013-2014 Magnet and Specialty Schools for more information.

CCD School Universe File
The Common Core of Data (CCD) includes school-level data on school programs, demographics, and location for all public schools in the United States, collected by the National Center for Education Statistics (NCES). The CCD data include some extended missing values where possible to indicate different types of missing data: .m for missing data (meaning a value is expected but none was measured; for example, an elementary school that cannot measure the number of 3rd graders), .n for not available (meaning a value is not expected; for example no value for number of 3rd grade students in a high school), and .t for technical data issues (meaning the submitted data item does not meet NCES data quality standards; in 2010-11 and 2011-12 data only). HERC has constructed several variables to help researchers utilize the CCD data. See *Documentation to the NCES Common Core of Data Public Elementary/ Secondary School Universe Survey: School Year 2010–11* for more information.

Race/ethnicity. In 2007-08, NCES used a five-category race/ethnicity variable: American Indian, Asian/Pacific Islander, Hispanic, Black, and White. Starting in the 2008-09 year, NCES began transitioning to a seven-category variable, by separating Asians and Pacific Islanders and by adding a Multiracial category for students who select more than one race. However, districts were given until the 2010-11 school year to complete the transition process. The variable ecracecatYR indicates whether the campus uses five or seven-category race/ethnicity measures (this variable does not exist in 2007-08). In HISD, campuses used the five-category variable from 2007-08 through 2009-10, and then used the seven-category variable in 2010-11 and 2011-12. Thus, in 2008-09 and 2009-10, variables for Pacific Islander and Multiracial students are included in the data set, but are set to missing (.n); after that, valid numerical values are included for those variables for schools that have valid counts.

HERC constructed several variables to maximize comparability of data on Asians and Pacific Islanders across years. In 2007-08, variables beginning with ecap* indicate counts of Asian/Pacific Islander students, in various grade/gender combinations, and ecasianpi08 is the total number of Asian/Pacific Islander students in each school (note: the * in the middle of a variable name stands in for the middle of the variable name). Beginning in 2008-09, additional sets of variables are included that indicate the number of Asian students only (ecas*YR and ecasianYR) and Pacific Islander students only (ecpi*YR and ecpacifYR). In 2008-09 and 2009-10, values for Pacific Islander variables are set to missing (.n) and values for Asian variables are equal to the values for Asian/Pacific Islander variables. In 2010-11 and 2011-12, there are valid numerical values for the Pacific Islander variables, and the values of Asian/Pacific Islander variables are equal to the sum of Asian and Pacific Islander variables. HERC recommends that researchers use Asian/Pacific Islander variables for most longitudinal data analyses, but provides the separate measures on Asians and Pacific Islanders for users who need those detailed data.

In addition, although the CCD data include counts of specific race/ethnic groups on each campus, HERC constructed campus percentages as well (ecpaminYR, ecpasianYR, ecpasianpiYR, ecpblackYR, ecphispYR, ecpmultiYR, ecppacifYR, ecpwhiteYR), in which total

race/ethnic counts (ecaminYR, ecasianYR, ecasianpiYR, ecblackYR, echispYR, ecmultiYR, ecpacifYR, ecwhiteYR) are divided by total school membership (ectotpopYR). The variable ecminYR is the sum total of Black and Hispanic students in the campus, and ecpminYR is the percentage of students that are Black and Hispanic. Finally, ecmajminYR is a dichotomous indicator that the school has more than 50% of students who are Black or Hispanic.

Title I schools. The CCD data provide indicators of whether a campus is designated as Title I Targeted Assistance (ectitleiYR) or Title I Schoolwide Program (ecstitliYR). HERC constructed ectitswYR as a categorical variable to indicate whether a campus was either of these, neither, or missing. Variables ecnotitlYR, ectlitasYR, ectliswpYR, and ectlimisYR are dummies to indicate each of these categories, and ectitliYR indicates that a campus has any Title I designation, either Targeted Assistance of Schoolwide Program. See *Documentation to the NCES Common Core of Data Public Elementary/ Secondary School Universe Survey: School Year 2011–12* for more information on Title I designations.

TEA School Profiles Data
The Academic Excellence Indicator System (AEIS) of the Texas Education Agency (TEA) maintains a database of school-level information based on every public school student in the state of Texas. This repository includes thousands of variables indicating student and staff demographics, school financial information, standardized test achievement benchmarks, school completion and college readiness indicators, and much more. HERC has accessed and cleaned data on every school in HISD from 2007-08 through 2011-12. Due to its size and complexity, details on the TEA school profiles data are covered in a separate codebook (see *TEA HISD School Profiles Codebook*).

Campus-Level Aggregates of Student-Level Variables
HERC created several variables as campus-level counts, aggregated from the student-level PEIMS data. Variables beginning with the prefix catg*YR are counts of total students in each grade, cate*YR are counts of total students of each ethnic group, categ*YR are counts of total students of ethnic group by grade, cated*YR are counts of each economic disadvantage group, and catedg*YR are counts of economic disadvantage group by grade. Other PEIMS aggregates include campus-level counts of students by home language category (cathl*YR), LEP status, based on pdlepYR (catlepYR), special education status (catspecedYR), bilingual participation, based on pdbilingYR (catbilingYR), ESL participation (cateslYR), gifted/talented (catgtYR), at risk status (catatriskYR), and number of students with at least one dkgrdYR or chggrdYR from the cleaned, longitudinal data file (cadkgrdYR, cachggrdYR). HERC also created percentages of these variables, based on total campus counts (cactcampYR); these variables are similar to campus total variables listed above, but with the prefix capct*YR. See *HISD Longitudinal Database: Student Codebook Intro* for more information on the PEIMS and longitudinal variables on which the campus aggregates are based.

Data Deletions and Analytic Samples

HERC's goal is to make this longitudinal database as useful as possible for as many researchers as possible, so we deleted very few observations during the cleaning process, and we encourage

researchers to create analytic samples as needed for their specific research purposes. To that end, we have created several flags (where 1=valid observation) to help researchers select samples for their analyses. The tracking flags for each source indicate the total number of observations in the data file after deletions were made but with no further restrictions. In addition, for the CCD data, we created analytic sample flags with some suggested restrictions that focus on the most common school types. HERC recommends that researchers combine the relevant analytic flags across years as needed to create relevant samples, and to link to individual student and/or teacher data. Below we describe the data deletions and flags for relevant data files.

<u>**Magnet Program**</u>
Data deletions- none.
Tracking flag: This data source is already limited to schools with magnet programs.
   <u>flmagpYR</u>- no further restrictions.

<u>**CCD**</u>
Data deletions- Any non-HISD schools were deleted.
Tracking flag:
   <u>flscunYR</u>- no restrictions.
Analytic sample flag:
   <u>flanecYR</u>- not closed or inactive schools, regular school types (non-special education, alternative, or other types).

<u>**Campus Aggregates**</u>
Data deletions- none.
Tracking flag:
   <u>flcaggYR</u>- no restrictions.

Related Documents/Websites
(Note: Linked documents go to web resources. Non-linked documents are stored on the HERC online server.)

[Documentation to the NCES Common Core of Data Public Elementary/ Secondary School Universe Survey: School Year 2011–12](#)
[HISD School Choice Options: 2013-2014 Magnet and Specialty Schools](#)
TEA HISD School Profiles Codebook
HISD Longitudinal Database: Student Codebook Intro