

HERC LONGITUDINAL DATA CODEBOOK

Introduction

This codebook accompanies the HERC longitudinal database, which includes student-level and school-level data from the Houston Independent School District (HISD). This database is a result of a data sharing agreement between HISD and the Houston Education Research Consortium (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 five years, from the 2007-08 school year through the 2011-12 school year, and it includes the population of students enrolled in HISD schools.

Data Sources

PEIMS

The Public Education Information Management System (PEIMS) is a database of student background, demographics, and program participation, collected each academic year and reported to the state and federal government. Thus, the PEIMS is a reliable indicator of all students enrolled in HISD, as recorded on a particular day in October of each year. See PEIMS Variables: Table of Contents for more information.

Stanford 10

The Stanford 10 achievement test is used to evaluate the progress of student achievement and to make comparisons to both national and local reference groups using current and empirically derived normative information. Stanford 10 is an English language test; Spanish-speaking students in HISD took the Aprenda test, which is not included in this database. Students in grades K-11 were given the Stanford reading and math tests from 2007-08 through 2010-11. Starting in 2011-12, only students in grades K-8 were given the Stanford reading and math tests. Stanford science tests were given to students in grades 3-11 from 2007-08 through 2010-11, and grades 3-8 in 2011-12. The Stanford assesses students across five subject areas: reading, math, language, science, and social studies; this database currently only includes reading, math, and science. Stanford provides five types of scores: the raw score (number correct), scale score (criterion-referenced vertically scaled score), grade equivalent (student's grade placement at the time of testing, spread across all grades), percentile rank (compared to a norm-referenced sample, by grade and subject), and normal curve equivalent (percentile score transformed into equal-interval, continuous scoring scale). We recommend the scale score as the best scale to measure content knowledge growth across years, while the normal curve equivalent (NCE) is best for measuring changes in relative rank across years. The grade equivalent (GE) and NCE scores are scaled to the first decimal place. See the HISD 2011-12 Stanford 10 Report, Stanford-Aprenda Code & Definitions, and Jorgenson (2004) for more information.

TAKS

For accountability purposes, the Texas Assessment of Knowledge and Skills (TAKS) was the state-required achievement test from 2002-03 through 2011-12. TAKS was used to assess

campuses and districts in terms of their yearly progress, and in certain grades, students must pass the TAKS to be promoted to the next grade. TAKS was given to all students in grades 3-11 through the 2010-11 school year. In 2011-12, the TAKS was replaced with the State of Texas Assessments of Academic Readiness (STAAR) as the main state-required achievement test. In that year, only students in grades 10 and 11 were allowed to take the TAKS rather than the STAAR test. The STAAR is not yet included in the HERC database because it has not been administered long enough to be able to measure growth across years.

TAKS assesses students in five subject areas: reading, math, writing, science, and social studies. As only the reading and math tests are given to students in every grade, these are the only subjects included in this database. For each year and subject, the TAKS file includes several variables: score code (S=valid score), raw score, scale score, met standard (dichotomous variable indicating that score met minimum passing standard), met commended standard (dichotomous variable indicating that score met higher standard to qualify as commended achievement), language version (English or Spanish, which is only given to grades 3-6), and test version. The TAKS provides different test versions, including some for general education or special education students who may need testing accommodations (TAKS-A, TAKS-M, TAKS-Alt) or English language learners who need linguistic accommodations (TAKS LAT). See TAKS Documentation 2003-2011 and TAKS 2010-2011 Accommodations Manual for more information.

Vertical and horizontal scale scores. The TAKS scale variable includes two different types of scale scores. All students in grades 9-12 have a “regular” horizontal scale score, which allows direct comparisons of performance within grade and subject. Across grades 9-12, the threshold for “met standard” corresponds to a scale score of 2100, and the threshold for “commended” corresponds to a scale score of 2400. For students in grades 3-8, the TEA created a vertical scale score in reading and math tests only. The vertical scale scores range from 0-1000; a student’s vertical scale score in one grade can be compared to the student’s scale score in another grade within the same test language, test version, and subject, allowing for the measurement of knowledge growth across years. Vertical score thresholds for met standard and commended performance are specific to test subject, language, and grade (see TAKS Vertical Scale: About for thresholds). The HERC longitudinal database includes a combined scale score variable (trvrscYR and tmvrscYR); the value for students in grades 3-8 is a vertical scale score, and the value for students in grades 9-12 is a horizontal scale score. Longitudinal analyses of TAKS achievement growth can only be done for students in grades 3-8 over a four-year period, from the 2007-08 school year through the 2010-11 school year (TAKS was not given to students in grades 3-8 in 2011-12). HERC recommends using the combined scale score (trvrscYR and tmvrscYR) for longitudinal analyses examining achievement growth, only for grades 3-8.

For 2007-08 and 2008-09, the vertical scale scores were merged in from a different file. As a result, the database also includes separate variables of regular horizontal scale scores (tmrsc08,tmrsc09, trrsc08,trrsc09), and vertical scale scores (tmvs08,tmvs09, trvs08,trvs09), in addition to the combined score variables. Several other variables may be of interest. Several thousand cases in grades 3-8 had missing vertical scale scores; these scores were directly imputed, as vertical scores are directly derived from horizontal scale scores for students within the same grade, test subject, test language, and test version. The variables trvsimp08, trvsimp09,

tmvsimp08, and tmvsimp09 are flags for imputed cases. In addition, a very small number of cases could not be imputed; these are flagged with trvsmiss08, trvsmiss09, tmvsmiss08, and tmvsmiss09. Finally, in a very small number of cases, the horizontal scale scores recorded in the original TAKS file were different from the horizontal scale scores merged in from the file that also contained the vertical scale scores; the horizontal score from the vertical score file was used as the final horizontal score in those cases. Variables trscdf08, trscdf09, tmscdf08, and tmscdf09 are flags for those cases. See TAKS Performance Report, 2010-2011, TAKS Documentation 2003-2011, TAKS Vertical Scale: About, and TAKS Vertical Scale Reporting: FAQ's for more information.

Zoned File

HISD is a district of choice, meaning that families can choose to enroll their children in their zoned schools or in non-zoned schools such as magnet, charter, or international baccalaureate schools. The zoned file contains information on the HISD campus each student is enrolled in, as well as the campus each student is zoned to, for all students enrolled in HISD in an academic year. (Note: the N for this file is considerably larger than the N for the PEIMS, since the zoned file includes all students ever enrolled throughout one year, including those who may have enrolled in HISD after the PEIMS date in October.) This file includes HERC constructed variables indicating whether a student attends a zoned school -- an indicator of school choice, as well as a flag indicating that the student resides outside of HISD district boundaries. HISD accepts a small number of out-of-district students each year; these students are missing on zoned school since they are not zoned to an HISD school.

Magnet Student File

This file includes student ID and campus ID for all students who are enrolled in an HISD magnet school program in each academic year. This file does not include all HISD students; only those enrolled in magnet programs. Researchers can combine data from the magnet student file and the zoned file to construct a measure of school choice that indicates whether students choose to attend a non-zoned school and/or whether students are enrolled in a magnet program. (Note: students can attend a non-zoned school without being enrolled in a magnet program, and students can be enrolled in a magnet program in their zoned school.)

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. This file does not include all HISD schools, only those with an HISD magnet program. It contains variables indicating magnet program type (school-within-a-school, schoolwide program, etc.), as well as program description. Program descriptions are combined into categories that follow those described in the HISD Magnet Program Catalog; some schools have programs in more than one category. See the 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 NCES has made only preliminary data available for

the 2011-12 school year; as such, this database does not include student demographics or economic status for that year. These measures may be aggregated from the student-level PEIMS data. See Documentation to the NCES Common Core of Data Public Elementary/ Secondary School Universe Survey: School Year 2010–11 for more information.

Data Structure

Each data file was cleaned separately for each year (e.g., TAKS reading 08-09, Stanford math 10-11, CCD 11-12). Since the PEIMS is the official student data reported to the NCES, we used this as the master data file and merged other data files to it accordingly. Files from a particular year (e.g., 2007-08) were then merged into a full yearly file that contained data from the following sources: PEIMS; Stanford reading, math, and science; TAKS reading and math; zoned file; magnet student file; magnet program file; CCD school universe. These five yearly files were then merged into one longitudinal file with N=383,272.

The data are organized in wide form at the student level; that is, each observation corresponds to one student who has multiple variables repeated for each year, with a suffix indicating the year (08 for 2007-08, for example). Each student has a unique student ID number (id). There are also year-specific student ID variables (id08, id09, ..., id12) derived from the PEIMS data set; students who were in PEIMS in a particular year will have a non-missing value on the yearly ID variable, while students who were not in the PEIMS (for example, students who enrolled in HISD after the October PEIMS date) will have a missing value.

Each student also has a campus ID variable (campus08, campus09, ..., campus12) for each HISD school in which the student was enrolled at the PEIMS date in each academic year. (Note: students may have changed schools after the PEIMS date but are still considered to be enrolled in the PEIMS school for the academic year. HERC is currently cleaning attendance data that would allow us to track school changes within an academic year.) Student-level data (TAKS, Stanford, zoned file, magnet student file) are merged to the PEIMS based on student ID; school-level data (magnet program file, CCD) are merged to each student-level observation based on campus ID. As such, each school-level variable is repeated for each student enrolled in that campus, within an academic year.

HERC has also created some year variables to help track year of observation when data are reshaped to long. The variables yr08, yr09, ..., yr12 are count variables that assign a number to each year of observation. Thus, yr08=0, yr09=1, up to yr12=4. We have also created variables indicating the academic year, based on the spring semester, so that researchers will be able to track the academic year of each person-year: schyr08="08", schyr09="09", up to schyr12="12". Also, ycampusYR and yridYR are variables that repeat campus ID and student ID, but with a prefix indicating the academic year. For example, ycampus08= "08"+campus ID for the school in which the student is enrolled in the 2007-08 academic year, and yrid11="11"+ student ID for students in the data set in the 2010-11 academic year. These variables may help researchers see, at a glance, the year of observation for which they have valid student or campus data when scanning a data table with many rows of observations.

Some measures are repeated across several data sources within a particular year; for example, PEIMS, TAKS, Stanford, zoned file, and magnet student files all have measures of student grade level. In cases where a discrepancy exists in such a variable, HERC recommends using the PEIMS as the most valid source of the variable because these are the official student data reported to the NCES.

Longitudinal Data Structure

The longitudinal database allows for analyses of cohorts of students across several years. However, the possible cohorts differ by test and subject, due to the administration schedules of the TAKS and Stanford tests. Stanford reading and math tests were given to all grades K-11 in years 2007-08 through 2010-11, and Stanford science tests were given to grades 3-11 in those same years. In 2011-12, the Stanford tests were no longer given to grades 9-11, possibly due to the introduction of the STAAR end-of-course tests in high schools. HERC recommends limiting studies of academic growth to grade 8 and lower since course sequences, and expected content knowledge, become more variable in high school. Thus, longitudinal analyses for Stanford reading and math should be limited to grades K-4 in 2007-08 through grades 4-8 in 2011-12. For Stanford science, a four-year cohort study should include grades 3-6 in 2007-08 and grades 6-8 in 2010-11, and a five-year cohort study should include grades 3-4 in 2007-08 and grades 2011-12.

The schedule for the TAKS is also complicated. As stated above, only grades 3-8 include vertically scaled scores that can be compared across grades. Also, TAKS reading and math tests were administered to grades 3-11 from 2007-08 through 2010-11. In 2011-12, TAKS was phased out to be replaced by the STAAR test; only grades 10-11 were “grandfathered” into taking the TAKS in that year. Thus, longitudinal cohort analyses should include grades 3-5 in 2007-08 through grades 6-8 in 2010-11. See Longitudinal Data Files: Intersection of Year and Cohort for more information.

Data Deletions and Analytic Samples

HERC’s goal is to make this longitudinal database as useful as possible for as many researchers as possible. As a result, 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 number of observations from the source data after deletions were made but with no further restrictions. In addition, for some data sources, we created analytic sample flags with some suggested restrictions that will allow researchers to create samples with comparable students who can be tracked over time. When each yearly file was merged into the full, longitudinal file, we did not delete any observations or create any tracking or analytic sample flags. We recommend that researchers combine the relevant yearly flags across multiple years to create longitudinal samples as needed.

Below we describe the data deletions and sample flags for relevant data files. (Note: the YR suffix is a placeholder for year of observation, based on spring semester of the academic year; for example, flanp08 is the analytic sample flag for PEIMS data from the 2007-08 school year.)

PEIMS

Data deletions- Any observations with missing student ID's were deleted.

Tracking flag:

flpeimYR- no restrictions.

Analytic sample flag:

flanpYR- full-day eligible students, based on Average Daily Attendance.

Stanford

Data deletions- Any observations with missing student ID's, missing scale scores, or duplicate student ID's were deleted. For duplicates, we kept the observation with the highest scale score, per HISD protocol.

Tracking flag:

flstanmYR- Stanford math: no restrictions.

flstanrYR- Stanford reading: no restrictions.

flstansYR- Stanford science: no restrictions.

Analytic sample flag:

flansmYR- Stanford math: non-special education students.

flansrYR- Stanford reading: non-special education students.

flanssYR- Stanford science: non-special education students.

TAKS

Data deletions- Any observations with missing student ID's, missing scale scores, a non-valid (other than S) score code, or duplicate student ID's were deleted. For duplicates, we kept the observation with the highest scale score, per HISD protocol.

Tracking flag: In 2007-08 and 2008-09, vertical scale scores were merged in from a separate file; researchers should use fltaksvmYR and fltaksvrYR as the tracking variable for those years.

fltaksmYR- TAKS math: no restrictions, (2009-10 through 2001-12).

fltaksrYR- TAKS reading: no restrictions, (2009-10 through 2001-12).

fltaksvmYR- TAKS math: no restrictions, (2007-08 & 2008-09).

fltaksvrYR- TAKS reading: no restrictions, (2007-08 & 2008-09).

Analytic sample flag:

flantmeYR- TAKS math: test version K (standard TAKS test), non-special education, English language.

flantmsYR- TAKS math: test version K (standard TAKS test), non-special education, Spanish language.

flantreYR- TAKS reading: test version K (standard TAKS test), non-special education, English language.

flantrsYR- TAKS reading: test version K (standard TAKS test), non-special education, Spanish language.

Zoned file

Data deletions- None.

Tracking flag:

flzoneYR- no restrictions.

Magnet Student

Data deletions- None.

Tracking flag: This data source is already limited to students enrolled in magnet programs.

flmagsYR- no further restrictions.

Magnet Program

Data deletions- none.

Tracking flag: This data source is already limited to schools with magnet programs.

flmagpYR- no further restrictions.

CCD

Data deletions- Any non-HISD schools were deleted.

Tracking flag:

flscunYR: no restrictions.

Analytic sample flag:

flanecYR- not closed or inactive schools, regular school types (non-special education, alternative, or other types).

Yearly Merged Files

Data deletions- none.

Tracking flag:

fmergeYR- no restrictions.

Analytic sample flag: After all data files were merged within a particular academic year, HERC created analytic sample flags for each yearly merged file that combine analytic sample flags and tracking flags from several data files: specific testing outcome (Stanford or TAKS, specific to subject and language), PEIMS, CCD, and zoned file. (Note: flags for magnet student and magnet program files are not included since these indicate only a small subset of students and/or schools. Researchers interested in those subsamples should include those flags as well.)

flmansmYR- flansmYR=1, flanpYR=1, flanecYR=1, flzoneYR=1.

flmansrYR- flansrYR=1, flanpYR=1, flanecYR=1, flzoneYR=1.

flmanssYR- flanssYR=1, flanpYR=1, flanecYR=1, flzoneYR=1.

flmantmeYR- flantmeYR=1, flanpYR=1, flanecYR=1, flzoneYR=1.

flmantmsYR- flantmsYR=1, flanpYR=1, flanecYR=1, flzoneYR=1.

flmantreYR- flantreYR=1, flanpYR=1, flanecYR=1, flzoneYR=1.

flmantrsYR- flantrsYR=1, flanpYR=1, flanecYR=1, flzoneYR=1.

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:

<u>p</u>	PEIMS
<u>sm</u>	Stanford math

<u>sr</u>	Stanford reading
<u>ss</u>	Stanford science
<u>tm</u>	TAKS math
<u>tr</u>	TAKS reading
<u>hz</u>	Zoned file
<u>ms</u>	Magnet student
<u>mp</u>	Magnet program
<u>ec</u>	CCD school universe
<u>fl</u>	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, 12 for the 2011-12 academic year, etc. Variable labels give full descriptions of each variable. For example:

<u>trtv11</u>	Test Version, TAKS Reading 10-11
<u>smsped12</u>	Special Ed, Stan Math 11-12
<u>ectype08</u>	School Type, CCD 07-08
<u>hzecampnm10</u>	Enrolled Campus Name, Zoned File, 09-10
<u>mprglang09</u>	Language Program, Magnet Schools, 08-09

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 Lphlang11 is linked to variable phlang11, which is Language Spoken at Home, PEIMS 10-11. Each value label also contains the numerical value of the category. For example:

Language Spoken at Home, PEIMS 10-11	Freq.	Percent	Cum.
0.English	105,846	27.62	27.62
1.Spanish	91,087	23.77	51.38
2.Vietnamese	1,294	0.34	51.72
3.Other	6,018	1.57	53.29
.	179,027	46.71	100.00
Total	383,272	100.00	

Related Documents

HERC Longitudinal Database: Variables Spreadsheet

[Documentation to the NCES Common Core of Data Public Elementary/ Secondary School Universe Survey: School Year 2010-11](#)

[HISD 2011-12 Stanford 10 Report](#)

[HISD School Choice Options: 2013-2014 Magnet and Specialty Schools](#)

[Jorgenson, Margaret A. 2004. *The Value of the Stanford Scale as a Common Metric: Assessment Report*. San Antonio, TX: Pearson Education Inc.](#)

Longitudinal Data Files: Intersection of Year and Cohort

PEIMS Variables: Table of Contents

Stanford-Aprenda Code & Definitions

[TAKS 2010-2011 Accommodations Manual](#)

TAKS Documentation 2003-2011

[TAKS Performance Report, 2010-2011](#)

[TAKS Vertical Scale: About](#)

[TAKS Vertical Scale Reporting: FAQ's](#)