

National Center for Education Statistics

# NCES Teacher Compensation Survey Technical Expert Panel

---

Report by National Institute of Statistical Sciences Technical Expert Panel  
31 October 2013



The Statistics Community Serving the Nation

National Institute of Statistical Sciences

PO Box 14006

Research Triangle Park, NC 27709-4006

Tel: 919.685.9300

FAX: 919.685.9310

www.niss.org

## **Teacher Compensation Survey (TCS) Technical Expert Panel (TEP)**

### **Final Report**

October 31, 2013

## **Contents**

<b>1</b>	<b>Background</b>	<b>1</b>
1.1	The TCS . . . . .	1
1.2	Charge to the TEP . . . . .	2
1.3	TEP Membership . . . . .	3
1.4	TEP Activities . . . . .	3
1.5	NISS Data Quality Analysis . . . . .	3
<b>2</b>	<b>Findings</b>	<b>3</b>
2.1	Items of Consensus . . . . .	3
2.2	Additional Findings . . . . .	5
2.3	Minor Items . . . . .	6
<b>3</b>	<b>Recommendations</b>	<b>6</b>
3.1	Conversion of the TCS to Relational Form . . . . .	6
3.2	Items of Consensus . . . . .	8
3.3	Other Concerns . . . . .	9
3.4	Analyses Using the TCS . . . . .	10
3.5	Data Release . . . . .	10
<b>4</b>	<b>Minor Items</b>	<b>11</b>
<b>A</b>	<b>TEP Meeting: May 20–21, 2013</b>	<b>12</b>
<b>B</b>	<b>Figures</b>	<b>12</b>
<b>C</b>	<b>The TCS and School Equity</b>	<b>19</b>

# 1 Background

This is the report of the Technical Expert Panel (TEP) convened by the National Institute of Statistical Sciences (NISS) on behalf of the National Center for Education Statistics (NCES) to evaluate the quality and utility of the Teacher Compensation Survey (TCS). The most important sections are the TEP findings (§2) and associated recommendations (§3).

## 1.1 The TCS

The NCES website<sup>1</sup> states that

The Pilot Teacher Compensation Survey (TCS) is a research and development effort to see if it is possible to collect and publish teacher-level data from the administrative records residing in state departments of education. Twenty-three (23) states participated in the TCS for SY 2008-09. Participating states provided data on salaries, years of teaching experience, highest degree earned, race/ethnicity, and gender for each public school teacher. Data on government expenditures on employee benefits were also solicited, but only three states were able to provide even partial employee benefits data.

To give a sense of scale, for 2008–09, the TCS contains 1,666,721 records for 1,586,058 teachers in 47,979 schools belonging to 7,892 LEAs in 23 states.

TCS data files are flat files with one record per *teacher assignment*. An assignment is conceptualized as one teacher’s instructional activities at one school, allowing for the same teacher to have additional, non-instructional activities at that school, as well as instructional activities. Of the 1,586,058 teachers in the 2008–09 TCS, 1,542,651 have exactly one assignment; the maximum number of assignments is 33.

The TCS contains eighteen primary variables for each assignment (National Center for Education Statistics, 2011):

**BASESAL** = base salary, in dollars

**TOTPAY** = total pay, in dollars

**RETIREBEN** = retirement benefits, in dollars

**HEALTHBEN** = healthcare benefits, in dollars

**OTHERBEN** = other benefits, in dollars

**TOTALBEN** = total benefits, in dollars

**EXP** = years of experience

**DEGREE** = highest degree earned

**RACE** = race/ethnicity

**SEX** = gender

---

<sup>1</sup><http://nces.ed.gov/ccd/tcssurv.asp>

**BRTHYR** = year of birth (nnnn)

**AGE** = age, in years

**CONTRCTDAYS** = contract days

**FTE** = fraction of full-time-equivalent associated with that activity

**TCHSTIND** = teacher status indicator

**SALIND** = salary indicator

**DSTNEWTCHIND** = district new teacher indicator

**STNEWTCHIND** = state new teacher indicator.

Missing value codes are -1 for numerical variables and “M” for categorical variables. See National Center for Education Statistics (2011) for details.

## **1.2 Charge to the TEP**

The charge to the TEP was as follows.

1. Provide an assessment of the data already collected and specific guidelines for its use by researchers, addressing guidelines on current data, to include:
  - (a) State-by-state notes on data issues and possible solutions.
  - (b) Possibilities and limitations of longitudinal analysis of teachers (with special focus on consistency of identification of individuals year-to-year).
  - (c) Possibilities and limitations of year-to-year comparisons given changes in data availability, data quality and changes in definitions of data elements.
2. Make recommendations for changes to the TCS collection process that will improve data quality or increase the value of the overall collection. Examples of potential improvements are:
  - (a) Enhancements to the current collection process, for instance, identifying the optimum set of variables to collect in TCS and providing appropriate definitions of variables and instructions to states.
  - (b) Expansion of the TCS, such as recruiting more participant states, or including other school staff in addition to teachers.
  - (c) Leveraging relationships between the TCS and other surveys, which include the potential of TCS as a sampling frame for other data collections, the cost of TCS compared to other collections of teachers and the potential cost savings realized by replacing or combining TCS with other surveys, and the potential for reducing respondent burden by TCS’ replacing other collections.



### **1.3 TEP Membership**

Members of the TEP are:

- Frank Cernik, North Carolina Department of Public Instruction
- Tom Gallagher, Wyoming Department of Workforce Services
- Carrie Jones, Bureau of Labor Statistics
- Alan Karr, National Institute of Statistical Sciences
- Steven Rivkin, University of Illinois at Chicago
- Marguerite Roza, Georgetown University

Key NCES liaisons are Stephen Cornman, Mark Glander and Marie Stetser.

### **1.4 TEP Activities**

The TEP met on May 20–21, 2013 in Washington, DC. The agenda for the meeting appears in Appendix A. Communication among TEP members before and after the meeting took place by e-mail.

In preparation for the TEP, Alan Karr attended a Teacher Compensation Survey Workshop held at NCES on September 20, 2012.

### **1.5 NISS Data Quality Analysis**

Concomitantly with the TEP, NISS conducted a data quality analysis for the TEP, which supplements, and is focused differently from, those conducted previously by the American Institutes for Research (AIR), such as Cornman et al. (2010). Under license from NCES, NISS was provided copies of restricted TCS data files for 2005–06, 2006–07, 2007–08 and 2008–09, as well as some internal NCES documents, especially a draft of the unreleased analog of Cornman et al. (2010) for 2007–08. The results of the analysis appear in Karr (2013).

## **2 Findings**

This section contains the findings of the TEP.

### **2.1 Items of Consensus**

There was clear agreement among TEP members that:

1. There are significant missing data problems with the TCS. As noted in §1, the 2008–09 TCS contains 14 variables for 1,666,721 teacher records, other than the four indicator variables. Only 302,739 records contain all 14 of these variables. Other records are missing as many as 11 variables. Figure 1 in Appendix B shows the distribution of the numbers of missing variables for four years of the TCS. Many analyses of TCS data are impacted by the level of missingness, albeit not all.

2. Timeliness of TCS data should and can be improved.
3. Financial elements of the TCS are problematic with respect to the key data quality dimensions of timeliness, accuracy and interpretability (Karr et al., 2006). In particular, there are issues with the following variables:

**Base Salary (BASESAL)** creates difficulty because it is a conceptual, rather than actual, monetary value, and is therefore subject to inconsistent interpretation across, and even possibly within, states. The TEP finds that this variable is of limited usefulness because it does not represent actual financial outlays by LEAs. Properly defined, Total Pay (TOTALPAY) has fewer shortcomings, provided that it is drawn from a workforce or tax agency database.<sup>2</sup>

**Total Pay (TOTAL PAY).** The intention underlying this variable is to capture compensation for duties in addition to classroom teaching.

**Benefits (RETIREBEN, HEALTHBEN, OTHERBEN, TOTALBEN)** which are provided by only a few states,<sup>3</sup> contain inconsistencies that are possibly both conceptual and operational,<sup>4</sup> and are of limited value for modeling and analysis, because of the extreme level and structured nature of the missingness.

**Experience (EXP),** which seems not to be treated consistently, but which is a central determinant of salaries.<sup>5</sup> Experience is stated in TCS documentation to be “Years of teaching experience,” which connotes career experience, but in the data, appears sometimes to be state- or LEA-specific experience. Other surveys, such as the Survey of Doctorate Recipients (SDR) conducted by the National Center for Science and Engineering Statistics (NCSES), employ a “Years Since [Highest] Degree” variable instead, which has issues of its own, but is not ambiguous (Carrillo and Karr, 2012).

**FTE,** which is especially problematic under the current system, in which (conceptually) an FTE and dollar amounts, including benefits, are associated with each assignment.

The relationalization of the TCS recommended in §3 supports addressing many of these issues. In particular, it would associate FTE commitments only with assignments, and leave to analysts, if desired, how to allocate salary or other dollar amounts. Importantly, states would not be asked to perform such calculations.

4. Currently, the TCS does not contain several important variables:

**Dates of employment** In the current TCS, there may be confusion between FTE as a rate and FTE as a portion of the fiscal year. For instance, it is not clear how someone who works at 100% effort for one-half of the year would be recorded. A separation of merit pay from “ordinary” pay is also desirable.

---

<sup>2</sup>In the 2008–09 TCS, total pay is present for 1,421,158 records, and Base Salary is present for 1,535,237 records.

<sup>3</sup>In the 2008–09 TCS, only 374,546 records contain all four benefits values (retirement, health, other, and total), while 1,045,817 contain no benefits values at all.

<sup>4</sup>For instance, how should actuarially unsound payments into a retirement plan be treated?

<sup>5</sup>That experience affects salaries, as noted during the TEP meeting, is not noteworthy. That, as NISS analyses have shown, it affects salaries differently across states and even across schools within local education authorities (LEAs), may be very important, because of connections to school equity.

**Standard occupational codes (SOCs)** for assignments. The current TCS does not even contain grade level or subject matter, which prevents its being used to address important questions of resource allocation. (Any use of SOLCs would, of course, need to be preceded by investigation of precisely how they are generated, and whether, for instance, there are important state-level or LKA-level differences in the process.)

**Year-of-degree or year-of-certification**, as discussed previously, which is a characteristic of a teacher rather than a position.

5. It seems inevitable that there will be demand in the future to link the TCS to student performance data. Such linkage is conceptually possible at the assignment, teacher, school and LEA levels. It is unclear to the TEP whether NCES wishes or plans to support such linkage.

## 2.2 Additional Findings

There was general support, or at least strong interest, among TEP members regarding the following items. Not all of them received detailed consideration at the TEP meeting, however.

1. NCES should explore the feasibility of automated methods for extracting financial data from state education authority (SEA) or state workforce agency databases. There may be both political and technological barriers to doing so, especially for data in statewide longitudinal data systems (SLDS) that also contain pupil information. However, timeliness and accuracy of information might be improved significantly.
2. The TCS could be expanded to include other public school employees. At the extreme, the TCS would become, in effect, a census of public school employees. Expansion of the TCS to include other instructional staff<sup>6</sup> would allow exploration of trends of interest in public education.
3. Existing TCS data, at least other than benefits, are suitable for release.<sup>7</sup> User documentation should address in detail issues for problematic variables noted in §2.1. NISS analyses have shown that TCS data support a variety of insightful explorations and modeling tasks, even if these do not generate “new science.” For instance, there is persuasive evidence that in some LEAs, there is a tradeoff between teacher experience and pupil-to-teacher ratios, whereas in others there is not. See Appendix C.
4. Support for longitudinal modeling—for example, of salary trajectories at the individual teacher level—is explicitly not a goal of the TCS as it has existed in the past. Year-to-year comparisons are, of course, feasible. There are seeming year-to-year inconsistencies at the individual level, in the form of large salary fluctuations. NCES should review its position on longitudinal modeling with the TCS.

It is not clear that the NCESTEACHID variable, which is intended to be a teacher identifier that does not change across years, in fact has this property. In the 2007–08 TCS, there are 1,234,370 teachers—that is, values of NCESTEACHID, and in the 2008–09 TCS, there are 1,586,058 teachers. However, only 951,551 values of NCESTEACHID appear in both files. The first number *seems* smaller than it ought to be, given that NCESTEACHID does not include an LEA identifier,<sup>8</sup> but on closer inspection,

---

<sup>6</sup>There may be an issue of definition, which basing assignments on standard occupational code (SOC)s would attenuate.

<sup>7</sup>Presumably, the same is true for 2009–10 and 2010–11 data, to which NISS has not had access.

<sup>8</sup>The TCS variable TCSID is the concatenation of NCESTEACHID and NCESSCH.

as shown in Figure 2, the “problem” lies entirely in five states that have data for both 2008–08 and 2008–09, but no records in the joined data set.

## 2.3 Minor Items

These items can be addressed in the course of dealing with those in §2.1 and 2.2. They do not seem to be controversial.

1. Usability of the TCS would be improved if there were a single reserve code for each aberrant data value, indicating whether it is missing at the state level, missing at the LEA level, missing at the teacher level, missing at the assignment level, or a violation of the edit constraints.
2. Although the TCS may comply with NCES’ Statistical Standards, the current coding of sex, degree and race forces some users to create new, more accessible variables. For instance, recoding sex from 1/2 to M/F makes data summaries (for instance, maps) and model output easier to comprehend. The racial categories employed in the TCS are not consonant with current practice, because “Hispanic” is one of them.
3. As has been noted in the past by AIR, the edit constraints implemented in the TCS are somewhat less than transparent. There may also be need for further discussion of specifics. The current suspension of the TCS provides an opportunity for a thorough review of them.

The same kind of review may also be beneficial for the statistical disclosure limitation (SDL) applied to the TCS.

## 3 Recommendations

### 3.1 Conversion of the TCS to Relational Form

One TEP recommendation cuts across all of the others. *The TCS should be relationalized*—that is, converted to a relational database consisting of a teacher table, an assignment table, a payroll table and a school table. Conversion to relational form removes current ambiguities in the TCS, and supports its expansion and more efficient use. A key principle is that, consistent with Recommendation 1 in §3.2, all financial information should be obtained at the teacher level from state workforce or tax agencies.

The database schema assumes, therefore, that payroll data are available at the (person  $\times$  LEA) level. That is, employers of record for teachers are LEAs. It is not sufficient to index this file only by persons, because the 2008–09 TCS dataset contains instances of one NCESTCHID associated with multiple LEAs.<sup>9</sup> The schema is fully compatible with an operational model in which different elements of TCS data are obtained from different sources.

The four tables in the relationalized version of the TCS are explained below. For each, we list the current TCS variables in it, as well as variables that the TEP proposes as candidates (see §3.2) to be added or deleted, with the primary key denoted by \*\*.<sup>10</sup> The tables are identified by the SAS files containing them.

---

<sup>9</sup>This is one reason why having dates of employment in the TCS is important.

<sup>10</sup>The flag variables are not included in the schema; each would be in the same table as its parent.

**Teacher Table** This table contains characteristics of teachers that are not LEA-specific. In particular placing demographic information in this table prevents the rare but avoidable inconsistencies in the TCS files where a teacher has one sex for one assignment and the other sex for some other assignment.

**TCS Variables:** NCESTEACID (\*\*), EXP(erience), AGE, DEGREE, RACE, SEX, BRTHYR, TCHSTIND, STNEWTCHID

**Potential Additional Variables:** Year of degree; year of certification

**Payroll Table** This table contains (person  $\times$  LEA)-indexed (that is, person  $\times$  employer of record) financial information: actual amounts paid as salary and, to the extent discernible, benefits.

**TCS Variables:** NCESTCHID, LEAID, TOTPAY, RETIREBEN, HEALTHBEN, OTHERBEN, CONTRCT-DAYS, DSTNEWTCHIND, NCESTCHID\_LEAID (\*\*). The last variable is simply a concatenation of NCESTCHID and LEAID. In a database in normal form, NCESTCHID and LEAID would not be present, because they can be derived from NCESTCHID\_LEAID.

**Potential Additional Variables:** Employment dates, merit pay

**Potential Deleted Variables:** BASESAL, TOTALBEN, which is redundant;<sup>11</sup> RETIREBEN, for which sound information may not be obtainable; OTHERBEN, which clearly is problematic in multiple ways.

**Assignment Table** This table contains details of all assignments associated with each teacher.

**TCS Variables:** TCSID (\*\*), FTE

**Potential Additional Variables:** SOC

**Potential Deleted Variables:** None

**The School Table.** This table is essentially a crosswalk of various school characteristics.

**TCS Variables:** NCESSCH(\*\*), STID, SEASCH, SCHNAM

**Potential Additional Variables:** Other school characteristics, in order to avoid having to link to the CCD to access them.

**Potential Deleted Variables:** The variables FIPS, STABBR, SCHNO and LEAID are conceptually part of this table, but are absent because they are derivable from NCESSCH.

The school table would not require independent data collection, because information in it is collected already by NCES, except that there are schools that appear in the TCS but not in the Common Core of Data (CCD).

NCES may choose to release files that are joins of the four tables, but this is a user service decision, not a database structure decision.

NISS has provided a relationalized version of the 2008–09 TCS to NCES, in the form of four SAS files.

---

<sup>11</sup>By contrast with other datasets, there are no inconsistencies of the form  $TOTALBEN \neq RETIREBEN + HEALTH\ BEN + OTHEBEN$ .

### 3.2 Items of Consensus

TEP recommendations map directly onto findings presented in §2.

1. The TEP recommends that all financial elements in the TCS be actual expenditures taken from accounting databases. In particular,
  - The BASESAL variable should be dropped. It represents a concept (contractual salaries) rather than financial reality.
  - The TOTALPAY variable should represent actual amounts paid at the teacher level, which is consistent with the database schema in §3.1, where financial information pertains to people, not assignments.
  - Benefits information be taken from person-level financial records. The TEP is skeptical that useful information about retirement benefits is obtainable, especially for defined-benefit programs. As noted above, even when dollar amounts representing payments into a pension plan are reported, there is no way of knowing whether they are actuarially sound.<sup>12</sup> The OTHERBEN variable should be dropped. It lacks a meaningful interpretation, and is demonstrably difficult for states to provide, so that many of them do not.<sup>13</sup>

By contrast, the recent addition of employer expenditure on medical insurance to W-2 forms ensures that reliable, consistent information about this important benefit will be present.
2. In support of Recommendation 1, the TEP urges that NCES explore obtaining the financial information for the TCS from unemployment insurance (UI) or other records held by state workforce or tax agencies. Advantages include timeliness and higher quality. The TEP acknowledges that in order to do this, NCES would need to establish a new, and for it, non-traditional, set of relationships with state agencies. If NCES does move to this model for data flow, the TEP believes that it should explore possibilities for automated data feeds from the states, which are both timely and accurate. The TEP realizes that political, technological and other impediments exist, especially for data in SLDS that also contain pupil information.
3. The TEP recommends strongly that NCES investigate in detail issues associated with linking assignment, payroll and teacher data, which may require the use of Social Security numbers. The database schema identifies the primary keys for the database tables, which are not SSNs.
4. The TEP urges that NCES explore adding variables identified above to the TCS, including dates of employment, SOCs, merit pay, year-of-degree, and year-of-certification. If the latter are feasible, then the EXP variable should be dropped.
5. The TEP recommends that NCES target releasing data for the school year ending 6/30/20xx by 7/1/20(xx+1).

---

<sup>12</sup>This is not to say that the issue of whether pension programs are underfunded is unimportant, but only that the TCS is not a good means of understanding it.

<sup>13</sup>It is difficult to make sense of the values that are provided. In the 2008–09 TCS, they range from \$0 to \$116,558, with a median of \$837. The analytical value is virtually nil.

6. The TEP recommends that, as part of the relationalization of the TCS, NCES consider carefully whether it wishes the TCS at some point in the future to support linkage to (aggregated or student-level) performance data, and if so, ensure that the relationalized version contains the proper “hooks” for doing so.

### 3.3 Other Concerns

Here we present other concerns that arose during, or as a result of, the TEP meeting.

1. The TEP is concerned about coverage of the TCS in three respects.

- The most obvious coverage concern is the number of participating states. Even after multiple versions of the TCS, fewer than one-half of the states participate. As a result, the TCS is an anomaly among NCES’ data collections. As long as participation in the TCS is voluntary, full coverage seems unlikely. NCES should assess carefully whether it is comfortable in releasing data with such low coverage.<sup>14</sup>

Use of the TCS as a frame for other NCES data collections, which the TEP was requested to address, hinges on two issues. First, to be used as a frame, the TCS must contain data from all states. Second, it must also contain the design-level stratification variables.

- The TEP urges that NCES consider expanding the TCS to include public school employees other than teachers, or at least instructional staff other than teachers. At the extreme, the TCS could become in effect, a census of public school employees. Expansion of the TCS to include other instructional staff<sup>15</sup> would allow exploration of trends of interest and importance in public education.
  - If the current trend toward “privatization” of public schools continues, the TCS, which does not cover contract employees, will suffer declining coverage even in states that do participate. This issue extends beyond the TCS, of course, and the TEP anticipates that NCES is moving to address it. Use of UI financial data has the potential to capture salaries and benefits paid of contract employees, but of course, willingness of private operators of public schools to provide the information may not be high.
2. Support for longitudinal modeling—for example, of salary trajectories at the individual teacher level—is explicitly not a goal of the TCS as it has existed in the past. Year-to-year comparisons are, of course, feasible. There are some year-to-year inconsistencies at the individual level, in the form of large salary fluctuations, or different values of SEX. The TEP feels that NCES should review its position on longitudinal modeling with the TCS. Some users will wish to do longitudinal modeling, and releasing the TCS, even as restricted data, with a prohibition year-to-year record linkage seems impractical.
  3. Consistency of the TCS with the Schools and Staffing Survey (SASS), CCD and data released by the National Education Association (NEA) is reasonable but not stunning. For instance, TCS and NEA state-level teacher counts are quite close, but breakdowns by sex are not. It is, of course, not clear what degree of consistency there “should be.” To illustrate, Figures 5, 6 and 7 show that TCS and

---

<sup>14</sup>Were the issue nonresponse, NCES’ Statistical Standards seem clear: the data would not be released.

<sup>15</sup>There may be an issue of definition, which basing assignments on SOC’s would attenuate.

NEA data, the latter taken from National Education Association (2010), agree well on numbers of teachers in each state and average salaries in each state, but differ substantially in the percentage of male teachers in each state.<sup>16</sup> It is perplexing that agreement should be good on two measures but not the third.

4. The TEP believes that NCES should define more precisely the scope of the TCS, with the goal that the scope be sufficiently narrow to make the TCS financially and operationally viable. To illustrate, there was extended discussion in the TEP meeting of important aspects of teacher compensation, such as contractual compensation versus incentive compensation versus compensation for additional (possibly, non-teaching) duties. The TEP agrees that these are important issues, and likely to become more so. However, it is not clear that the TCS is the best mechanism for addressing them, as opposed to alternatives such as targeted surveys. Many of items discussed in this report are meant to reduce complexity of the TCS in order to improve quality. Rich detail about compensation is a step in the opposite direction.

### 3.4 Analyses Using the TCS

The extent to which the TCS can enable insightful analyses not possible by other means remains an issue. The potential of the TCS to generate novel scientific insights may inherently be limited. To illustrate, statistical analyses performed by NISS as part of its data quality assessment, show that the TCS leads to “reasonable” models of salary as a function of variables such as age, gender, experience and level of highest degree. That these factors affect salary is not novel. The values of estimated model coefficients and the extent to which they differ across LEAs or states, by contrast, may be very interesting. The extent to which models fit also varies, and is significant in itself.

To illustrate briefly, consider a linear model for  $\log(\text{TOTPAY})$  by state, with AGE, CONTRCTDAYS, DEGREE, RACE and SEX as predictors. In workforce modeling, these variables are typically predictive of salary Carrillo and Karr (2012). Figure 3 in Appendix B shows that the performance of the model varies substantially over states: the root mean squared error (on a log scale) varies from .11 to .29, and the coefficient of determination  $R^2$  varies from .23 to .63. The variability in estimated coefficients is equally striking. When the same model is fit with LEA rather than state as the unit of analysis, there is even greater variation, as shown in Figure 4, which is based on 3,740 LEAs.

NISS was requested by NCES to assess the capability of the TCS to address issues of school equity, broadly defined. Many analyses was conducted, some of which are summarized in Appendix C.

All these analyses notwithstanding, the TEP hopes that NCES will continue to delineate what is possible with the TCS, but not any other dataset.

### 3.5 Data Release

The TEP recommends that NCES release TCS data for 2008–09, other than benefits. Accompanying user documentation should address in detail issues for problematic variables noted in §2.1. Presumably, 2009–

---

<sup>16</sup>The TCS estimates were produced from a file indexed by teachers rather than assignments. For each teacher, base salary, total pay, benefits and FTE were summed over assignments. To avoid distortions from teachers who are truly part-time and problems with assignments discussed previously, only teachers with just one assignment and FTE = 1 were included: there are 1,343,649 of them.



10 and 2010–11 data<sup>17</sup> can also be released on the same basis. As articulated above and in Karr (2013), TCS data support a variety of insightful explorations and modeling tasks, even if they do not generate “new science.” Releasing at least 2008–09 TCS data relatively soon, rather than awaiting resumption of the TCS, would have the advantages of beginning to build a user community and stimulating user feedback.

## 4 Minor Items

These items can be addressed in the course of dealing with those in §3. They are not controversial.

1. Usability of the TCS would be improved if there were a single reserve code for each aberrant data value, indicating whether it is missing at the state level, missing at the LEA level, missing at the teacher level, missing at the assignment level, or a violation of the edit constraints.
2. Although the TCS may comply with NCES’ Statistical Standards, the current coding of sex, degree and race forces some users to create new, more accessible variables. For instance, recoding sex from 1/2 to M/F makes data summaries (for instance, maps) and model output much easier to comprehend. The same is true for the DEGREE and RACE variables.
3. The racial categories employed in the TCS are not consonant with current practice, because “Hispanic” is one of them. This may place the TCS in conflict NCES or Office of Management and Budget (OMB) guidelines, and also thwarts analyses involving comparison or linkage to dataset in which Hispanic is an ethnicity distinct from race.
4. As has been noted in the past by the AIR, the edit constraints implemented in the TCS are not transparent. There may also be need for further discussion of specifics. The current suspension of the TCS provides an opportunity for a thorough review of them. The same may also be true for the SDL applied to the TCS, which seems not to be mentioned in any TCS documentation.

---

<sup>17</sup>To which NISS has not had access.

## **A TEP Meeting: May 20–21, 2013**

### **Monday, May 20**

- 9:00 AM Welcome and Introductions; Discussion of TEP Charge
- 9:15 Introduction to TCS: NCES Personnel
  - Purpose, Content, History, Uses/Users
- 10:30 Break
- 10:45 Data Collection and Processing: NCES Personnel
- 12:00 N Lunch
- 1:00 PM NISS Data Quality Review of TCS: Alan Karr
- 2:30 Break
- 3:00 Initial Discussion of Use Cases
  - Examples: Analysis of Teacher Compensation, Longitudinal Analyses, Sampling Frame for other NCES data collections, School Equity [Also, linkage to other NCES datasets, such as SASS and CCD]
- 4:30 TEP Executive Session
- 5:00 Adjourn for the day

### **Tuesday, May 21**

- 9:00 AM TEP Working Session
  - Initial Formulation of Findings and Recommendation
  - Identification of Information and Research Needs
  - Plans and Schedule
- 11:30 TEP/NCES Discussion
- 12:00 N Adjourn

## **B Figures**

In order to avoid interrupting the text, all figures associated with the body of the report appear together in this appendix.

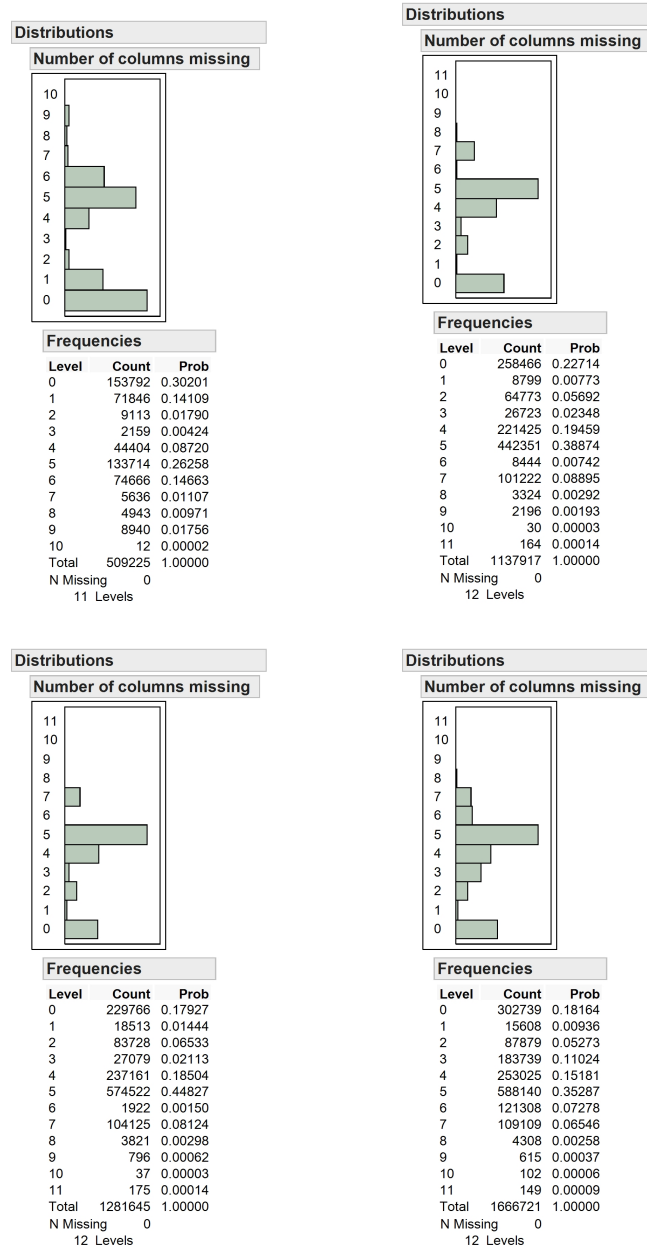


Figure 1: Numbers of variables missing in TCS data, among 14 major non-indicator variables. *Upper left:* 2005–06. *Upper right:* 2006–07. *Lower left:* 2007–08. *Lower right:* 2008–09.

	State	Count0708	Count 0809	Count Joined	Fraction (Joined/0708)	Fraction (Joined/08009)
1	AR	31111	30918	27932	0.8978174922	0.9034219548
2	AZ	56164	56830	.	.	.
3	CO	49353	50327	.	.	.
4	FL	183259	178967	163454	0.8919289094	0.9133192153
5	IA	36051	35868	33329	0.9244958531	0.9292126687
6	ID	15521	15733	.	.	.
7	KS	35341	35837	32273	0.9131886477	0.9005497112
8	KY	43769	43670	39915	0.9119468117	0.9140141974
9	LA	49644	50418	.	.	.
10	ME	16631	16406	15412	0.9267031447	0.9394124101
11	MN	55743	55203	50471	0.9054231025	0.914280021
12	MO	70450	71767	63562	0.9022285309	0.8856716875
13	MS	34698	34897	30731	0.8856706438	0.8806201106
14	NC	.	104457	.	.	.
15	ND	.	7662	.	.	.
16	NE	21137	21226	19111	0.9041491224	0.9003580514
17	NJ	114673	116026	102800	0.896462114	0.8860083085
18	OH	.	110517	.	.	.
19	OK	43879	43760	39736	0.9055812575	0.9080438757
20	SC	49441	49662	44447	0.8989907162	0.8949901333
21	TN	.	63343	.	.	.
22	TX	327505	333439	288378	0.8805300682	0.8648598394
23	WA	.	59125	.	.	.

Figure 2: Counts, by state, of numbers of teachers in the 2007–08 TCS data file, the numbers of teachers in the 2008–09 data file, and the numbers of teachers in the file constructed by joining the two using NCESTEACHID. Five states are clearly not using this variable consistently across years.

	State	N	Mean(log salary)	Root Mean Square Error	RSquare
1	AR	25043	10.707217496	0.1445351025	0.4276706409
2	FL	165836	10.722806945	0.2960505081	0.2341493684
3	IA	34017	10.792435991	0.1392807701	0.5230319002
4	ID	14527	10.722237795	0.1773382634	0.4179163988
5	KY	24927	10.790114462	0.1111836237	0.5201789979
6	MN	36421	10.846121596	0.1694682893	0.6053702046
7	MO	59583	10.701945651	0.1986207571	0.4362057885
8	MS	27477	10.609834221	0.1650599448	0.4581563713
9	NC	80815	10.620984916	0.112118601	0.6362611927
10	ND	7272	10.661339934	0.1684927614	0.4790153755
11	NE	17795	10.730656536	0.1492220106	0.5475472917
12	OK	35628	10.57677085	0.1223257907	0.3980472737
13	TN	61701	10.709929749	0.1365770402	0.4900809218
14	WA	40592	10.984886611	0.1427455741	0.5499314548

Figure 3: Measures of regression model performance for 14 states, using the 2008–09 TCS. The response is log(TOTPAY), and predictors were teacher characteristics such as age, gender, race and highest degree. Both root mean squared error and the coefficient of determination  $R^2$  vary significantly across states.

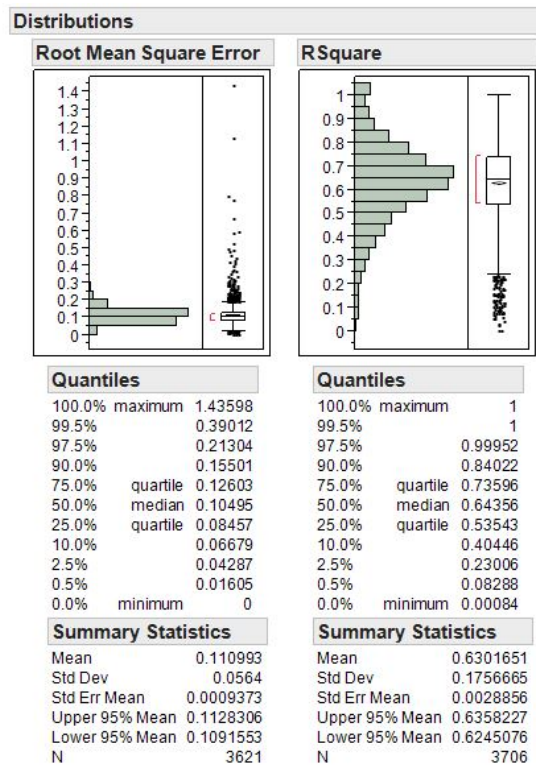


Figure 4: Measures of regression model performance for approximately 3700 LEAs, using the 2008–09 TCS. The response is  $\log(\text{TOTPAY})$ , and predictors were teacher characteristics such as age, gender, race and highest degree. Both root mean squared error and the coefficient of determination  $R^2$  vary significantly across LEAs.

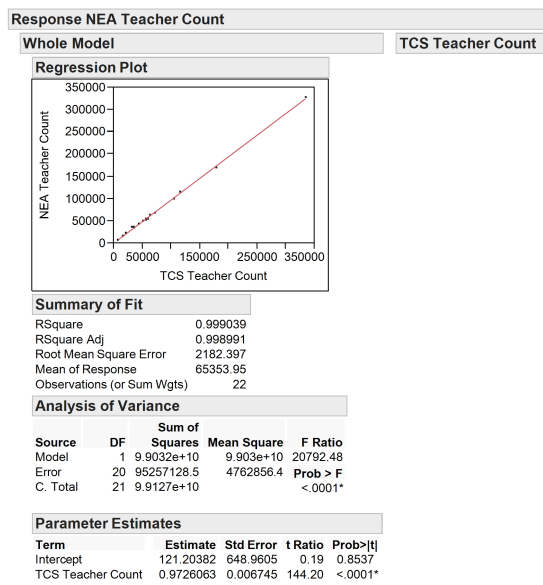


Figure 5: Regression of NEA state-level teacher counts versus TCS state-level teacher counts, for 2008–09. The fit is nearly perfect.

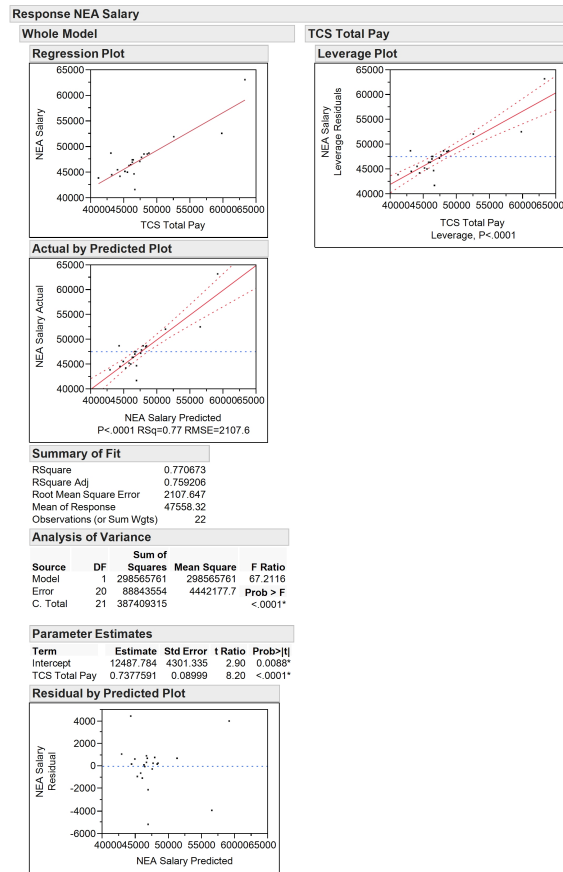


Figure 6: Regression of NEA state-level average teacher salaries versus TCS state-level average teacher salaries, for 2008–09. The fit is quite good, but not as good as for counts.

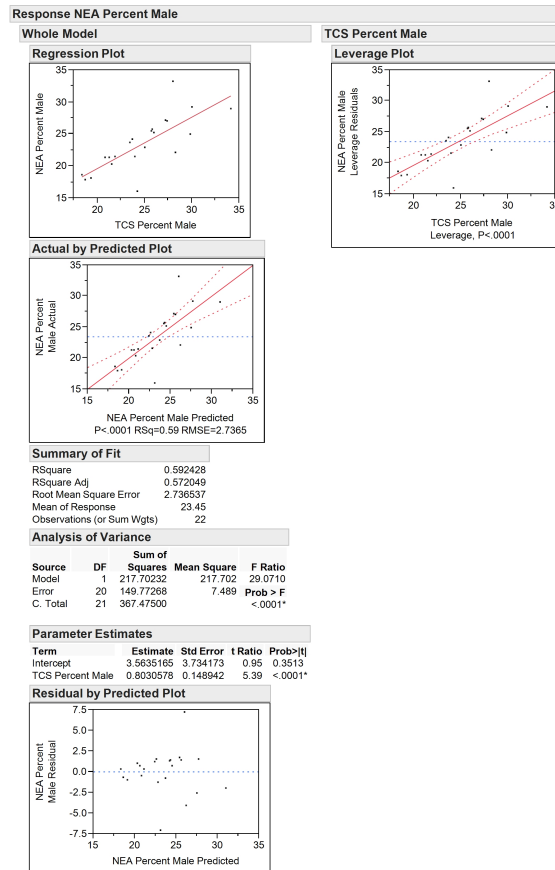


Figure 7: Regression of NEA state-level percentage of male teaches versus TCS state-level percentages of male teachers, for 2008–09. The fit is not as good as for counts or salaries.



## C The TCS and School Equity

NISS was requested by NCES to assess the capability of the TCS to address issues of school equity. Many analyses were conducted, a selected few of which are described here. Instantiation of “equity” is a complex issue with political, social and economic dimensions. Broadly speaking, the question is whether groups of students traditionally characterized as disadvantaged are receiving their fair share of educational resources. Two measures of resources relevant to the TCS were considered: teacher salaries and teacher experience.

The TCS itself contains no information regarding student characteristics, but it is straightforward to link it to the CCD to obtain such information. For the analyses described here, this was done for 2008–09. The variables employed in the analyses are listed below. In general, they are school characteristics that measure the presence of historically disadvantaged groups: Title I status, urbanicity, and the percentages of students who are black, hispanic or eligible for free or reduced-price lunches.

A major—and apparently unique—advantage of the TCS in this context is the ability to perform analyses *at the teacher level*, using actual salary, experience and demographic variables. If they were performed using school-level variables,<sup>18</sup> their insightfulness would be attenuated significantly. The analyses reported below were performed with two groupings, by LEA and by state.

*Q1: Does teacher pay depend on school characteristics in ways that suggest that teachers in schools with certain characteristics (for instance, high percentages of students eligible for free or reduced-price lunches) are paid less than their counterparts in other schools?* To address this question, a regression analysis was conducted for each of 86 large LEAs (in terms of number of schools), in which the unit of analysis is the teacher, the response is  $\log(\text{TOTPAY})$ , and the predictors are:

**Teacher-level variables from the TCS** DEGREE, EXP, RACE, SEX

**School-level variables taken directly from the CCD** LEVEL, PUPTCH, STITLI, TITLEI, ULOCAL

**School-level variables constructed from the CCD** PercentFRL = Percent of students eligible for free and reduced-price lunch, PercentBlack = percent of black students, PercentHispanic = percent of hispanic students.<sup>19</sup>

Figure 8 shows the distributions of the values of selected school-level variables. The mean of every one of the estimated coefficients is negative: on the average, teacher salaries are lower (1) the higher the percentage of free and reduced-price lunch students, the higher the percentage of black students, the higher the percentage of hispanic students, the higher the pupil-teacher ratio, if the school is Title I eligible, and if the school is School-Wide Title I.<sup>20</sup> Because of the presence of the teacher-level variables, whose effects on salary are described in §1.5, differing characteristics of the teachers are accounted for in the models.

At the state level, the message is more nuanced, but the strongly negative influence of PercentFRL is still present. See Figure 11.

*Q2: Does teacher experience depend on school characteristics in ways that suggest that teachers in schools with certain characteristics are less experienced than their counterparts in other schools?* This analysis is very similar to that for salaries. The the response is  $\log(\text{EXP})$ , and the predictors are:

<sup>18</sup>For instance, average (or median) salary, average experience, percentage of teachers with degrees at various levels, . . .

<sup>19</sup>These variables may themselves be subject to data quality problems, which are ignored in this report.

<sup>20</sup>In Figure 8, the reference category for TITLEI and STITLI is 2 = “no.”

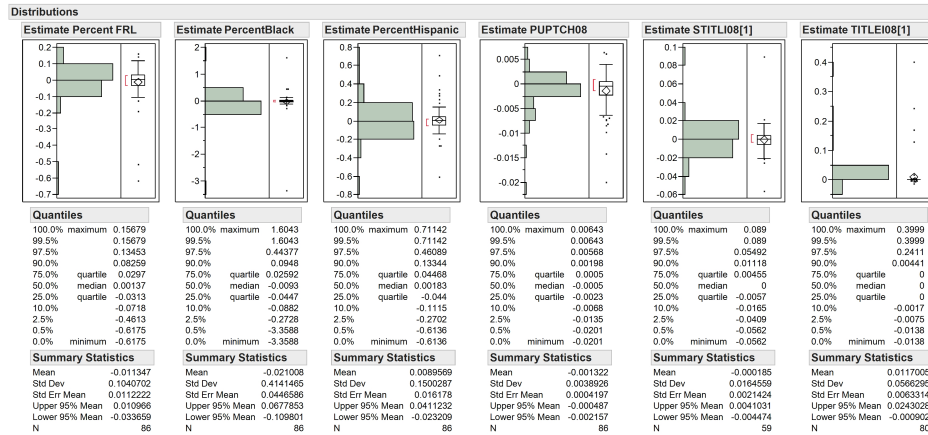


Figure 8: Distribution of coefficients of selected school-level variables in regression, by LEA, of teacher salary versus teacher and school characteristics.

## Teacher-level variables from the TCS DEGREE, RACE, SEX

## School-level variables taken directly from the CCD LEVEL, PUPTCH, STITLI, TITLEI, ULOCAL

**School-level variables constructed from the CCD** PercentFRL = Percent of students eligible for free and reduced-price lunch, PercentBlack = percent of black students, PercentHispanic = percent of hispanic students.

The results for LEAs appear in Figure 10, and those for states are in Figure 11. Again, the negative effect of PercentFRL is notable.

*Q3: Are LEAs “trading off” teacher experience for numbers of teachers?* While less directly a question of equity, the question of whether LEAs are trading off numbers of teachers for teacher experience, across schools, is a relevant measure of their behavior. The motivation for the question is that *employing more teachers*, which lowers PUPTCH, and *employing teachers with more experience* are different responses to perceived student needs. This question was addressed, for the 50 largest LEAs, by means of a regression analysis in which schools (within each LEA) are the unit of analysis, PUTTCH is the predictor and median teacher experience is the response. Because of extreme values of PUPTCH, its logarithm was used as the predictor. The results appear in Figures 12 and 13, each of which shows the results of the analysis for 25 LEAs. When the slope of the least squares regression is positive, experience is higher in schools with larger values of PUPTCH, which suggests that a tradeoff is being made: schools with more pupils per teacher have teachers with more experience. By contrast, when the slope of the line is negative, as it is for many LEAs, schools with more pupils per teacher also have teachers with less experience. A lack of equity is one interpretation of the latter phenomenon.

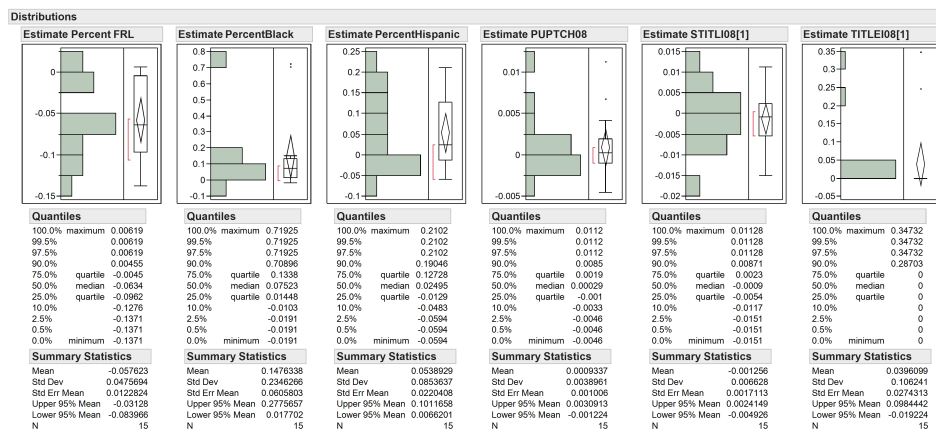


Figure 9: Distribution of coefficients of selected school-level variables in regression, by state, of teacher salary versus teacher and school characteristics.

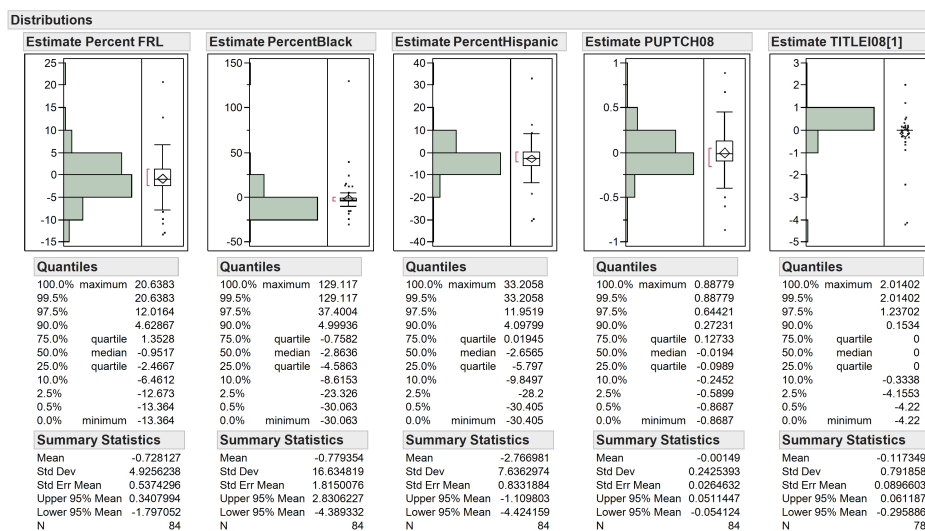


Figure 10: Distribution of coefficients of selected school-level variables in regression, by LEA, of teacher experience versus teacher and school characteristics.

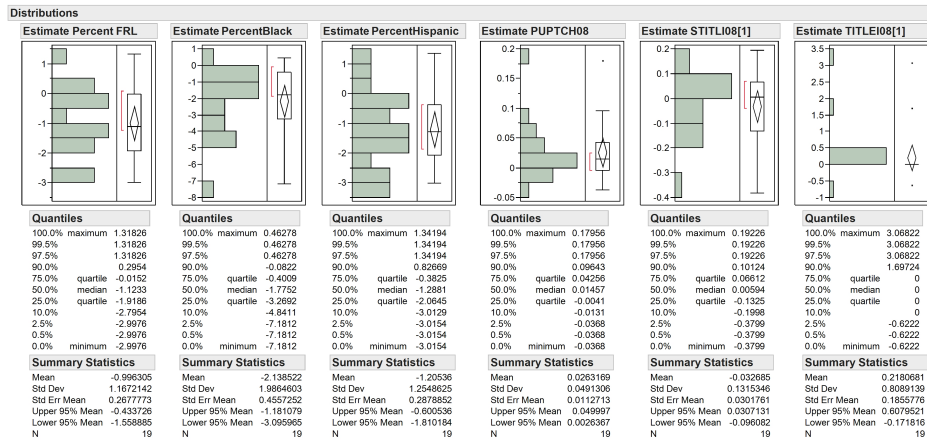


Figure 11: Distribution of coefficients of selected school-level variables in regression, by state, of teacher experience versus teacher and school characteristics.

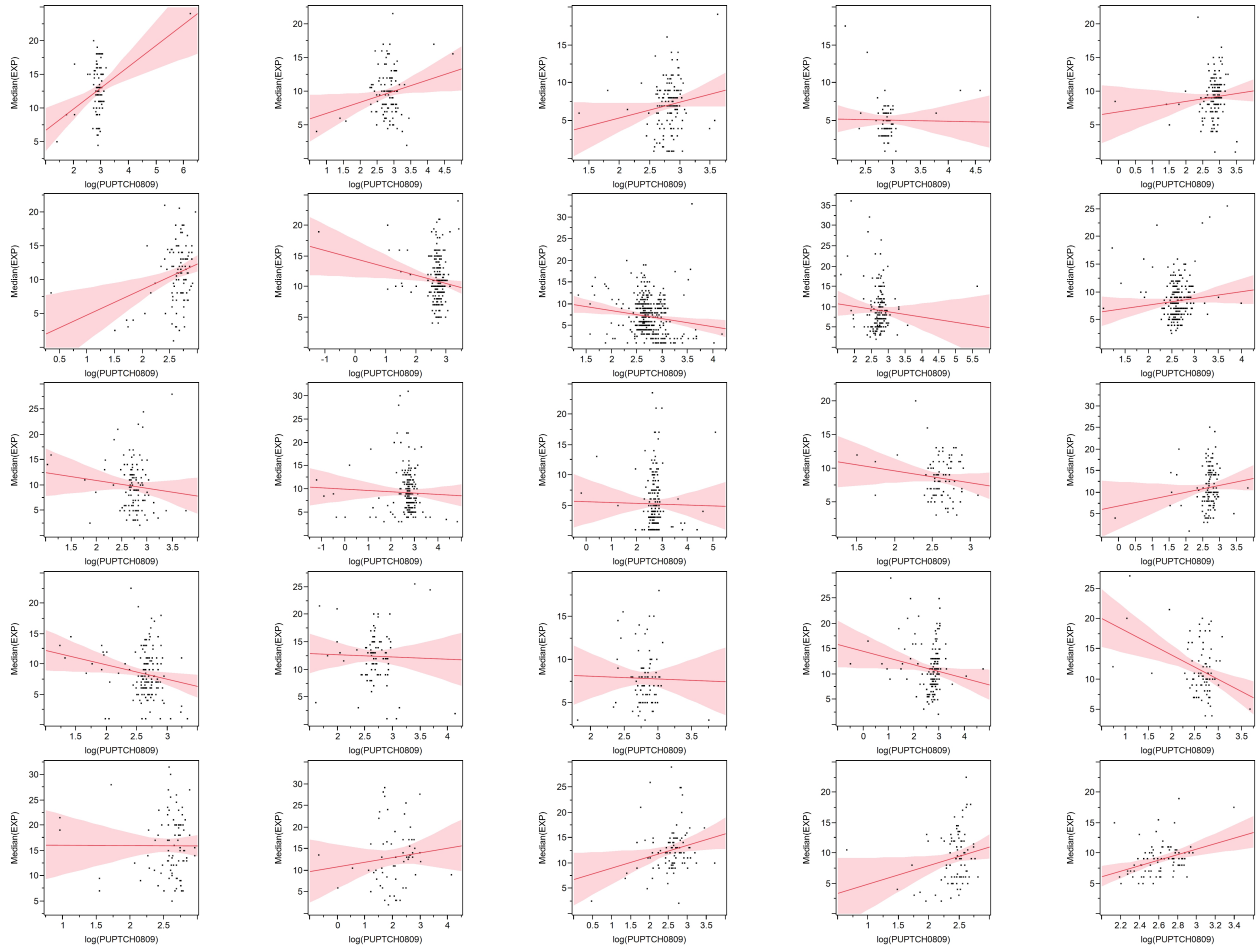


Figure 12: Regression, by school, of median teacher experience versus pupil-teacher ratio, for 25 of the 50 largest LEAs.

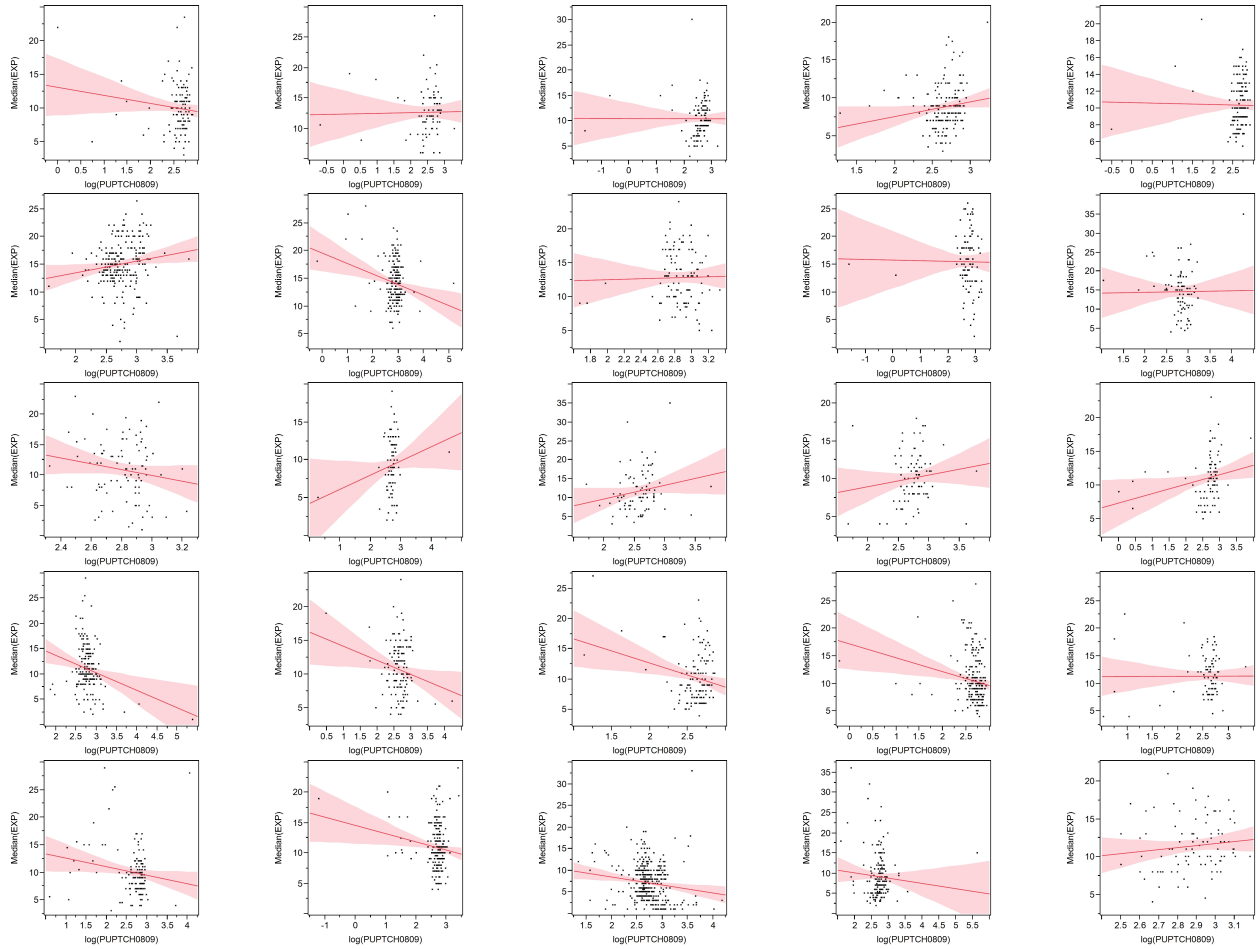


Figure 13: Regression, by school, of median teacher experience versus pupil-teacher ratio, for 25 more of the 50 largest LEAs.

## References

- Carrillo, I. and Karr, A. F. (2012). Combining cohorts in longitudinal surveys. *Survey Methodology*. To appear; available online as NISS Technical Report 180, at <http://www.niss.org/sites/default/files/tr180.pdf>.
- Cornman, S. Q., Johnson, F., Zhou, L., Honegger, S., and Noel, A. M. (2010). An Evaluation of the Data From the Teacher Compensation Survey: School Year 2006-07. NCES Research and Development Report 2010-329, available on-line at <http://www.eric.ed.gov/PDFS/ED511938.pdf>.
- Karr, A. F. (2013). Evaluation of data quality for the Teacher Compensation Survey. In preparation.
- Karr, A. F., Sanil, A. P., and Banks, D. L. (2006). Data quality: A statistical perspective. *Statistical Methodology*, 3(2):137–173.
- National Center for Education Statistics (2011). Teacher Compensation Survey School Year 2009–10 Instruction Manual. OMB Control number 1850-0874.
- National Education Association (2010). Rankings & Estimates: Rankings of the States 2009 and Estimates of School Statistics 2010. Available on-line at <http://www.nea.org/assets/docs/010rankings.pdf>.