



December 2014

Applied Research Methods

---

# Using administrative data for research: A companion guide to *A descriptive analysis of the principal workforce in Florida schools*

---

Jessica Sidler Folsom  
La'Tara Osborne-Lampkin  
Carolyn D. Herrington  
Regional Educational Laboratory Southeast  
at Florida State University

## Summary

---

This companion guide describes the methods used to extract information from the Florida Department of Education database for data on school leaders and transform it into an understandable structure for use. The report focuses on methods of data cleaning, merging, and analysis in order to help those interested in analyzing similar databases in other states or contexts.

REL 2015–049

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

December 2014

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0011 by Regional Educational Laboratory Southeast administered by Florida Center for Reading Research, Florida State University. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Folsom, J. S., Osborne-Lampkin, L., & Herrington, C. D. (2014). *Using administrative data for research: A companion guide to A descriptive analysis of the principal workforce in Florida schools* (REL 2015–049). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the Regional Educational Laboratory website at <http://ies.ed.gov/ncee/edlabs>.

## Summary

This document is a companion guide to *A descriptive analysis of the principal workforce in Florida schools* (Folsom, Osborne-Lampkin, & Herrington, in press). It describes the methods used to extract information from the Florida Department of Education database in order to conduct a descriptive analysis of the demographic composition, certifications, and career paths of the state's school leaders in the 2011/12 school year. This companion guide aims to help those interested in using similar databases in other states or contexts for replication or further extension. This guide describes the process of data cleaning, merging, and analysis, including:

- Identifying and requesting data.
- Creating functional datasets.
- Merging datasets.
- Analyzing data.

## Contents

<b>Summary</b>	<b>i</b>
<b>About this guide</b>	<b>1</b>
<b>Steps 1–3: Refining the research questions, identifying data sources, and requesting datasets</b>	<b>1</b>
<b>Step 4: Preparing the datasets</b>	<b>2</b>
Importing data into a statistical software package	2
Assigning variable labels	3
Assigning value labels	3
Identifying data structure and unique identifiers	4
<b>Step 5: Creating a functional dataset for analysis</b>	<b>5</b>
Demographics dataset	6
Certificate dataset	6
Reported Work Experiences dataset	7
Jobs dataset	8
Merging all datasets	11
<b>Step 6: Analyzing and reporting on the data</b>	<b>11</b>
Descriptive demographic data and paired bar chart: Who are Florida’s school leaders?	12
Descriptive background data: What are the backgrounds of Florida’s school leaders?	13
Descriptive path data: Where have Florida’s school leaders served and in what capacity?	14
<b>A word of caution about using administrative data systems</b>	<b>19</b>
<b>Appendix A. Sample project timeline</b>	<b>A-1</b>
<b>Appendix B. Key terms</b>	<b>B-1</b>
<b>Appendix C. Sample data request</b>	<b>C-1</b>
<b>Appendix D. Data dictionary</b>	<b>D-1</b>
<b>Notes</b>	<b>Notes-1</b>
<b>References</b>	<b>Ref-1</b>
<b>Boxes</b>	
1 About the Education Data Warehouse of the Florida Department of Education	2
2 A note about replication	3
3 Univariate or multivariate data structures?	4
4 A note about merging datasets	10
<b>Figures</b>	
1 Example of restructured data	8
2 Example of a paneled bar graph of job titles	12

3	Example of a 100 percent stacked bar graph of demographic composition	13
4	Example of a paired bar chart of ages	13
5	Example of a 100 percent stacked bar graph of movement between school types	17
6	Example of a 100 percent stacked bar graph of movement between job categories	17

### Tables

1	Sample original Job Experiences dataset	5
2	Example of table of descriptive statistics of certifications	14
3	Example of table of descriptive statistics of experience types	15
4	Example of table of descriptive statistics of experience by school type	16
5	Example table of descriptive statistics of career paths	18

## About this guide

This document is a companion guide to the Regional Educational Laboratory Southeast report, *A descriptive analysis of the principal workforce in Florida schools* (Folsom et al., in press). The purpose of this companion guide is to clearly and succinctly describe the steps taken to extract and analyze data from the Florida Department of Education staffing databases. The companion report (Folsom et al., in press) describes the demographic composition, certifications, and career paths of Florida's school leaders in the 2011/12 school year.

While this guide is specific to the Florida Department of Education staffing database, the processes described can be modified to explore other education datasets. For example, states or districts may be interested in replicating the study on school types (such as charter schools and schools in need of improvement) or school subtypes (such as leaders of specific districts). In another application of the methods, states or districts may be interested in replicating the study on a different population—such as teachers or superintendents. Alternatively, some states or districts may be interested in extending this work to address correlational questions such as “What is the association between number of schools a principal has led and student achievement?” A sample project timeline is provided that aligns with each of the steps identified in this guide (see appendix A). However, the amount of time needed can vary substantially based on available resources and on the study team's level of expertise and background knowledge.

**One of the most critical steps in the research process after identifying and refining the research questions is identifying the data necessary to answer the questions**

The study team followed a six-step process in using the Florida Staffing Database to describe the principal workforce in Florida schools:

1. Refining the research questions.
2. Identifying data sources.
3. Requesting datasets.
4. Preparing the datasets.
5. Creating a functional dataset for analysis.
6. Analyzing and reporting on the data.

In the explanation of the six steps, datasets appear in Title Case, variables in *italics*, and syntax commands in ALL CAPS. The study team used IBM SPSS Statistics Version 21 for the analyses, and the syntax commands listed are specific to SPSS (see appendix B). However, this guide is intended to be broad enough that the methods can be applied in any basic statistical software package, including SAS and R. Readers should consult the documentation provided with their software package for the specific syntax or commands for each process described in this report.

### **Steps 1–3: Refining the research questions, identifying data sources, and requesting datasets**

One of the most critical steps in the research process after identifying and refining the research questions (step 1 in appendix A) is identifying the data necessary (step 2) to answer the questions. To conduct the descriptive analysis of the principal workforce in Florida, the study team used data that already existed for other purposes. Although using extant data eliminates the need for primary data collection, identifying appropriate data sources is imperative. Understanding the scope of the research and the available data sources must come first.

---

## **Box 1. About the Education Data Warehouse of the Florida Department of Education**

The Florida Department of Education’s data system on educators and students—the Education Data Warehouse—extends over multiple decades and is managed by the department’s Division of Accountability, Research, and Measurement. The Education Data Warehouse “integrates existing, transformed data extracted from multiple sources that are available at the state level. It provides a single repository of data concerning students served in the K–20 public education system as well as education facilities, curriculum, and staff involved in instructional activities” ([http://edwapp.doe.state.fl.us/EDW\\_Facts.htm](http://edwapp.doe.state.fl.us/EDW_Facts.htm)). The Education Data Warehouse is web based and easily accessible by the public.

To identify the specific datasets and variables necessary to describe the principal workforce in Florida, the study team reviewed the Education Data Warehouse Blueprint, which depicts the Education Data Warehouse’s structure, and the Education Data Warehouse Met@data portal, a data dictionary that provides descriptions, definitions, and codes for datasets and variables. The study team identified the Employee Demographic, Certified Staff, and Instructional Activity sections of the Blueprint as the sources with the most applicable data for the research questions. Within the Met@data portal, the study team identified Educational Staff as the primary business subject (that is, database), and Employee Demographic, Certified Staff, and Educational Staff as the primary business facets (that is, datasets) needed for the project. One further dataset was necessary: the Master School Identification Dataset. This dataset, publicly available on the department’s website, contains school-level information such as school type and grades served.

***Because clean data are crucial for ensuring reliable and valid results, data cleaning must be replicated with each variable of each dataset***

---

The study team used the Florida Department of Education’s online Education Data Warehouse to identify the applicable data for describing the principal workforce in Florida (box 1). The study team then submitted a formal data request (step 3; see appendix C for a sample data request narrative). After approving the data request, the Florida Department of Education provided three specific datasets, which are referred to throughout this report as the Demographics, Certificate, and Job Experiences datasets.

### **Step 4: Preparing the datasets**

Once steps 1–3 are complete and the data are acquired, step 4 is to prepare the datasets. A critical part of preparing the datasets is ensuring that all the data are clean. Data cleaning, an iterative or reparative process, can range from simple tasks such as renaming variables, assigning pseudonyms or numeric codes, or identifying extreme or unrealistic values to more complex tasks such as restructuring, splitting, aggregating, or filtering a dataset or creating new variables by writing complex code. Because clean data are crucial for ensuring reliable and valid results, data cleaning must be replicated with each variable of each dataset (box 2).

#### **Importing data into a statistical software package**

The Florida Department of Education provided the three datasets (Demographics, Certificate, and Job Experiences) as tab-delimited text files, which the study team imported into software designed for statistical analysis. The study team downloaded the Master School Identification Dataset Excel file from the Florida Department of Education website and imported it into the same statistical software as the other datasets (step 4.a in appendix A).

---

## Box 2. A note about replication

In the descriptive analysis of the principal workforce in Florida, replication involved the same study team conducting the cleaning and analysis twice and comparing results. Typically, the first round of cleaning or analysis (for example, cleaning a dataset to make it functional for analysis) was completed on day one, and day two was spent replicating the first day's work and comparing the results. An alternative to having a study team replicate its own work is to have a different study team conduct each step independently and compare results. Time and resources are often deciding factors in how replication occurs. What is most important is not how replication occurs, but that replication occurs.

---

Through data importing, each tab-delimited text file and Excel file became a unique dataset. See appendix D for a description of the original variables from each of the original datasets.

### Assigning variable labels

In the initial import the variable names from the text file were imported as the variable names in the statistical software. Next, the research team manually assigned variable labels (brief descriptions of each variable) and variable types (numeric or text) using the VARIABLE LABELS command (step 4.b in appendix A).

### Assigning value labels

For some variables, particularly in the Master School Identification Dataset file, the data were recorded as codes rather than a full string description. For example, school type was coded as 00, 01, 02, 03, 04, 05, or 07, referring to “Not yet assigned,” “Elementary,” “Middle/Jr. High,” “Senior High,” “Combination Elementary and Secondary,” “Adult,” or “Other,” respectively. For variables with codes, the study team manually assigned labels using the VALUE LABELS command for each code based on the Master School Identification Dataset data dictionary (<http://www.fldoe.org/eias/dataweb/tech/msid.pdf>; step 4.c in appendix A).

### Identifying data structure and unique identifiers

Understanding the data structure is crucial in the data preparation process (step 4.d in appendix A), particularly when multiple datasets are to be merged. To merge data across datasets, data must be laid out in a multivariate structure (box 3), and unique identifiers must be in place (step 4.e). To describe the principal workforce in Florida, each school leader needed a unique (to the individual) but common (across datasets) identifier in each dataset to enable the statistical software to accurately identify all the information associated with that individual. The variable *ID* was the only common variable between the three datasets and thus became the unique identifier and linking variable for school leaders.

Districts and schools must also have, or be assigned, a unique (to the district or school) but common (across datasets) identifier in datasets that link district or school information (step 4.e in appendix A). The Florida Department of Education datasets used unique numeric district and school codes. The Master School Identification Dataset provided both

***Understanding the data structure is crucial in the data preparation process, particularly when multiple datasets are to be merged***



---

### Box 3. Univariate or multivariate data structures?

Data can be organized in two formats: univariate and multivariate. In a univariate format, each row corresponds to a different piece of information. Typically, there is a column with the identifier for an individual, a column with the variables, and a column with the data for each variable. Thus, data for a single individual (case) can span multiple rows. It may help to think of this as “long” data for an individual. Administrative databases, such as the databases used for the analyses described in this report, often record data in a univariate format. In a multivariate format, each row represents a single individual, and there are multiple variables laid out in columns. It may help to think of this as “wide” data for an individual. Most data analysis procedures call for data to be stored in a multivariate format. Analyzing a single variable using univariate data would yield inaccurate results because the one variable actually stores multiple pieces of information. Thus, it is important to identify the data structure as soon as data are obtained and restructure if necessary.

---

**Districts and schools must also have, or be assigned, a unique (to the district or school) but common (across datasets) identifier in datasets that link district or school information**

text strings with the district and school names and unique numeric district and school codes. The numeric variables *District* and *School* were the unique identifiers linking school leaders' district and school information to the Master School Identification Dataset school information. It was necessary to use both *District* and *School* because school codes were not unique across the state, only within the district (for example, school 114 could exist in multiple districts). Thus when the study team used MERGE commands, cases were first matched by *District*, then by *School*.

The Demographics dataset was the only multivariate dataset with one row for each school leader. The Certificate dataset was a univariate dataset with a row for every certification ever received by a school leader. For example, one school leader who had been in the system for many years had 10 cases, one for each certificate received since 1976. The restructuring of the univariate Certificate dataset into a multivariate dataset is addressed later. The Job Experiences dataset, the most complex dataset, was a combination of two univariate datasets with a subdataset for each school year (table 1).

The Job Experiences dataset initially had a row for each of the eight experience types, each job classification, each year a school leader was present, and each school leader. For example, one school leader had 98 cases holding data for 11 years of service with multiple job titles for multiple years. The first subdataset, in a univariate structure, held information about each school leader's positions in each school year. The second subdataset, also in a univariate structure, held information about other self-reported work experience types. The study team split the larger Job Experiences dataset into two datasets to simplify data cleaning and analysis. The first dataset, Reported Work Experiences, retained the variables related to self-reported work experiences. The second dataset, Jobs, retained variables specific to the jobs held each year in Florida public schools. In both the Reported Work Experiences and Jobs datasets, the linking variable, *ID*, was retained so the study team could later remerge the datasets. When the study team split the larger Job Experiences dataset into two datasets, each new dataset retained its original univariate data structure.

**Table 1. Sample original Job Experiences dataset**

Year	ID	Job classification name	District	School	Work experience type	Years
2007	E	Teacher, self-contained, grade 5	Z	31	Teaching in Florida nonpublic schools	5
2008	E	Assistant principal, elementary school	Z	31	Administration in education	6
2009	E	Assistant principal, elementary school	Z	31	Administration in education	7
2010	E	Principal, elementary school	Z	50	Administration in education	0
2011	E	Principal, elementary school	Z	50	Administration in education	1
2007	F	Business director	Y	80	Administration in education	16
2008	F	Business director	Y	80	Administration in education	16
2009	F	Assistant principal, other elementary/secondary school	Y	80	Administration in education	18
2010	F	Assistant principal, other elementary/secondary school	Y	80	Administration in education	18
2011	F	Assistant principal, other elementary/secondary school	Y	80	Administration in education	18
2007	G	Principal, elementary school	X	51	Administration in education	0
2008	G	Principal, elementary school	X	51	Administration in education	1
2009	G	Principal, elementary school	X	51	Administration in education	2
2010	G	Principal, elementary school	X	51	Administration in education	3
2011	G	Principal, elementary school	X	51	Administration in education	4
2006	H	Teacher, self-contained, kindergarten	W	34	Teaching in Florida nonpublic schools	0
2007	H	Teacher, self-contained, kindergarten	W	34	Teaching in Florida nonpublic schools	1
2008	H	Reading coach, elementary school	W	60	Administration in education	0
2010	H	Teacher, self-contained, grade 3	W	21	Teaching in Florida nonpublic schools	0
2011	H	Assistant principal, other elementary secondary school	W	41	Administration in education	0
2010	J	Assistant principal, senior high school	V	33	Teaching out of state public schools	8
2010	J	Assistant principal, senior high school	V	33	Service to the district in current job code assignment	0
2011	J	Assistant principal, middle/junior high school	V	21	Service to the district in current job code assignment	1
2011	J	Assistant principal, middle/junior high school	V	21	Administration in education	1
2011	J	Assistant principal, middle/junior high school	V	21	Teaching out of state public schools	8
2006	K	Assistant principal, elementary school	U	30	Administration in education	0
2007	K	Principal, middle/junior high school	U	20	Administration in education	0
2009	K	Principal, other elementary/secondary school	U	61	Administration in education	0
2010	K	Principal, other elementary/secondary school	U	61	Administration in education	1
2011	K	Principal, elementary school	U	29	Administration in education	0

Source: Authors.

### Step 5: Creating a functional dataset for analysis

To create a functional dataset for analysis it is important to become familiar with and inspect each variable to determine whether the values or codes were intuitive and reasonable (steps 5.a.1, 5.b.1, 5.c.1, and 5.d.1 in appendix A). For example, in the Demographics dataset, it would be expected that the codes “M” and “F” in the *gender* variable would signify male and female. Similarly, a value greater than 1988 in the *birth\_year* variable would likely not be reasonable for school leaders in 2011/12. Once this inspection was complete, the study team created new variables and restructured the univariate datasets into a multivariate data structure with only one case per school leader. The process the

study team used to create a single functional dataset from the Demographics, Certificate, Reported Work Experiences, and Jobs datasets is described next.

### Demographics dataset

The Demographics dataset, the only dataset with one case per school leader, required minimal processing: the creation of a new variable, *age*, for the age of each school leader (step 5.a.2 in appendix A). The study team used a COMPUTE command where *birth\_year* was subtracted from 2011 (the focal school year). Because the Demographics dataset was already in a multivariate data structure, no additional data processing was needed.

### Certificate dataset

Florida educators hold a single Florida educator certificate. The certificate can encompass multiple coverages, each with an associated instruction level.<sup>1</sup> The Certificate dataset included coverages for school leaders dating back to 1966; therefore, many of the specific coverages and endorsement names are no longer in use.<sup>2</sup> The dataset also included expired coverages.

To clean the Certificate dataset, the study team first categorized and coded the coverages, endorsements, and instruction levels of the Florida educator certificate (step 5.b.2 in appendix A) into variables broad enough for interpretation but still meaningful to the Florida Department of Education. The study team initially created categorizations based on documentation on the department's website. The new variables, approved by the Florida Department of Education after a series of stakeholder meetings, emerged from a series of IF/THEN commands (step 5.b.3) and included:

- *Cert\_type*, which had codes identifying each coverage as an:
  - Endorsement.
  - Administrative coverage.
  - Subject area coverage.
  - Vocational coverage.
- *Admin\_type*, which had codes identifying each administrative coverage as:
  - School leader.
  - School principal.
  - Administration/supervision.
  - Local director of vocational education.
  - Administration of adult education.
- *Cert\_level*, which had codes identifying the instruction level for each coverage as:
  - All instruction levels.
  - Prekindergarten.
  - Elementary.
  - Secondary.
- *Cert\_current*, which had codes identifying if the coverage was:
  - Expired.
  - Current.

***To create a functional dataset for analysis it is important to become familiar with and inspect each variable to determine whether the values or codes were intuitive and reasonable***

Next, a series of AGGREGATE commands generated new variables with counts representing the number of cases for each code, each variable, and each school leader (step 5.b.4). The following variables were created:

- *Cert\_n\_ever*, a variable representing the number of certifications ever held.
- Two variables representing the number of:
  - Expired coverages (*cert\_n\_exp*).
  - Current coverages (*cert\_n\_cur*).

Next, a FILTER command reduced the dataset to only active (nonexpired) coverages per school leader (step 5.b.5). In step 5.b.6, another series of AGGREGATE commands created:

- Four variables representing the number of:
  - Endorsements (*cert\_n\_e*).
  - Administrative coverages (*cert\_n\_admin*).
  - Subject area coverages (*cert\_n\_subj*).
  - Vocational coverages (*cert\_n\_voc*).
- Five variables representing the specific administrative coverages:
  - School leaders (*admin\_n\_edl*).
  - School principal (*admin\_n\_sp*).
  - Administration/supervision (*admin\_n\_admin\_sup*).
  - Local director of vocational education (*admin\_n\_ldve*).
  - Administration of adult education (*admin\_n\_admin\_ae*).
- Four variables representing the number of coverages for each instruction level:
  - All levels (*cert\_n\_all*).
  - Prekindergarten (*cert\_n\_prek*).
  - Elementary (*cert\_n\_elem*).
  - Secondary (*cert\_n\_sec*).

The study team used an AGGREGATE command to retain the new variables and create the final dataset (step 5.b.7). The study team created a series of yes/no variables using IF/THEN commands to reflect that an individual school leader could hold multiple active coverages and coverage types (step 5.b.8). For example, the variable *cert\_voc\_yn* was a dichotomous variable coded as 0 for no and 1 for yes, indicating whether a school leader held a vocational coverage. Similarly, *admin\_edl\_yn* indicated whether a school leader held the school leader coverage.

### Reported Work Experiences dataset

Some codes and associated data, provided by the Florida Department of Education, in the Reported Work Experiences dataset were not accurate. For example, the number of years teaching in the current district was consistently reported as four despite seven consecutive years of data. Similarly, there were inaccuracies with codes for service to the district in current job assignment, teaching in Florida public schools, and administration in education. In conversations with the Florida Department of Education, it was determined that these data were collected at the district level and uploaded by the district to the Education Data Warehouse. The Florida Department of Education confirmed that districts are not required to regularly update the Reported Work Experiences. Therefore, districts differ in what information is collected and when it is collected and reported to the Florida Department of Education. Because data collection and reporting of these codes were not uniform across districts, the study team, along with the department officials, decided to

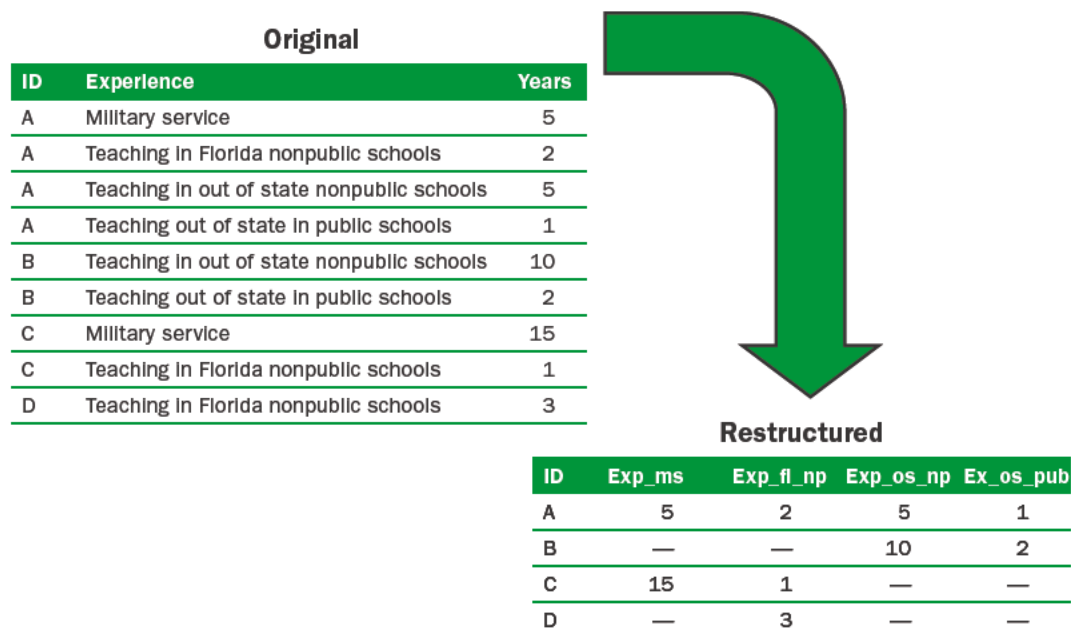
ignore codes related to Reported Work Experiences in Florida public schools (step 5.c.2 in appendix A).

Because there was no other source for capturing work experience outside of Florida public schools, the study team retained the self-reported work experience data in the final dataset. As some data were inconsistent across years, the study team decided to use the highest reported value for each experience. For example, a school leader may have started as a teacher in the Florida Department of Education, then taught in a Florida nonpublic school for two years, and then returned to the department as an assistant principal. In that case, the first recorded value for years of experience in a Florida nonpublic school would be 0, but a subsequent value would be 2. The remaining work experiences included military service (*exp\_ms*), teaching in Florida nonpublic schools (*exp\_fl\_np*), teaching in out-of-state nonpublic schools (*exp\_os\_np*), and teaching in out-of-state public schools (*exp\_os\_pub*). The AGGREGATE WITH MAX command identified the highest reported value for each work experience type for each school leader (step 5.c.3). Next, the study team used a RESTRUCTURE command to restructure the data to create a single case per school leader and a variable with the maximum reported time for each of the experiences (figure 1; step 5.c.4). The restructured dataset was saved as a new dataset, Experiences.

### Jobs dataset

For the broader objective of the study—describing the career paths of Florida’s school leaders between 2001/02 and 2011/12—preparing the Jobs dataset was critical. The most important variables were those for the specific job classification each school leader held each year. The study team used a FREQUENCY command to identify the job

**Figure 1. Example of restructured data**



**Note.** The original data structure is univariate, and the restructured data structure is multivariate.

**Source:** Authors.

classifications in the dataset; there were 366 unique job classification titles (for example, principal, records/forms analyst).

Analyzing career paths with 366 possible job classifications was not tenable. Instead, the study team sorted the job classifications into seven larger categories: classroom instruction, other instruction, support services, general administration, assistant principal, principal, and superintendent's/district office (see key terms in appendix B for definitions and examples, and see the Florida Department of Education's Automated Staff Information System Database Requirements, [http://www.fldoe.org/eias/dataweb/database\\_1112/sfappende.pdf](http://www.fldoe.org/eias/dataweb/database_1112/sfappende.pdf)). The study team consulted with the department through a series of stakeholder meetings to confirm the accuracy of the job categorizations. Each job classification was then placed into a job category (step 5.d.2 in appendix A). IF/THEN commands created two new variables for the job categories (step 5.d.3): one with a numeric code (*job\_cat*) and one with a text label (*job\_cat\_label*).

**Analyzing career paths with 366 possible job classifications was not tenable. Instead, the study team sorted the job classifications into seven larger categories**

Next, the AGGREGATE command isolated the specific job category each leader held each year (step 5.d.4). In this step the study team discovered that some school leaders held multiple jobs in multiple categories each year (step 5.d.5). For example, a school leader started an academic year as an assistant principal in one school and then transferred midyear to another school as the principal. In this instance the study team used the most recently obtained job position (based on start date) as the official position for that year. In another example a school leader held two assistant principal positions in two schools. In this instance the study team created a new variable (*multiple*), and the "official" position for the year was selected from the positions held that year, using a random number table (step 5.d.6).

When the study team had identified one primary position for each school leader for each year, an AGGREGATE command condensed the data into one row for each year and each school leader (step 5.d.7). Next, a MERGE command combined the Jobs dataset and the Master School Identification Dataset, linking cases by the *school\_ID* variable using the Master School Identification Dataset as the lookup table (step 5.d.8; box 4). The variable of interest in this step was *school\_type*, which identified the grade levels served at the school (elementary, middle, high, combination elementary/middle, adult, or other). Next, a series of AGGREGATE commands created (step 5.d.9):

- A variable indicating the number of years of data for each school leader (*n\_years\_data*).
- Three variables representing the number of different schools (*n\_schools*), districts (*n\_districts*), and school types (*n\_school\_type*).
- A variable indicating the number of job categories (*n\_job\_cat*).
- A variable representing the number of years in multiple job categories (*n\_years\_multi*).
- Six variables representing the number of years spent in each school type (*n\_years\_elem*, *n\_years\_middle*, *n\_years\_high*, *n\_years\_combo*, *n\_years\_adult*, and *n\_years\_oth*).
- Seven variables representing the number of years spent in each job category (*n\_years\_class*, *n\_years\_oth\_inst*, *n\_years\_support*, *n\_years\_admin*, *n\_years\_ap*, *n\_years\_princ*, and *n\_years\_supoff*).

The study team used the RESTRUCTURE command to rearrange the data to create a single row for each school leader, containing the new variables—this became the Jobs by Year dataset (step 5.d.10). The *year* variable served as an index for each of the variables

---

#### Box 4. A note about merging datasets

If not done with caution, data merging can lead to added or removed cases or to mismatched data. When preparing to merge data, one must first determine if the merging process will add variables or cases from other datasets.

When adding variables, there are two important details to consider. The first is identifying a linking variable unique to each case but common across datasets. The second is determining the way the datasets are to be merged. Particularly with longitudinal datasets, not all cases appear in all datasets. Therefore, it is important to identify the primary dataset, to which the remaining datasets will merge, and the datasets that will be lookup tables. The lookup table has only one case for each ID but may contain additional IDs not used in the primary dataset. The data in the lookup table can be applied to multiple cases in the primary dataset. The decision of which dataset to consider primary often depends on the research question.

In the study described in this report, the Jobs dataset was the primary dataset, and the Master School Identification Dataset was the lookup table. The Jobs dataset had multiple cases for each district/school combination, and the Master School Identification Dataset had one unique case for each district/school combination. By using the Master School Identification Dataset as the lookup table, the school-level variables for each district/school were applied to each case where the district/school appeared in the Jobs dataset. In merging the Experiences, Certificate, and Demographics datasets, the cohort of 2011/12 school leaders was the focus. This group was identified by the Experiences dataset; as described in the cleaning process, irrelevant cases were removed prior to merging the other datasets. Even though all datasets had only one case for each ID, the Certificate and Demographics datasets were considered lookup tables. That is, data from the Certificate and Demographics datasets were matched to the ID in the Experiences dataset.

---

identifying the position held by the school leader each year and where the position was held. Each row contained 11 sets of the following variables: *district*, *school*, *school\_type*, *job\_cat*, and *job\_cat\_label* (for example, *district.2001*, *school.2001*, *school\_type.2001*, *job\_cat.2001*, *job\_cat\_label.2001*, ..., *district.2011*, *school.2011*, *school\_type.2011*, *job\_cat.2011*, *job\_cat\_label.2011*). Finally, in the Jobs by Year dataset a new series of yes/no variables (created using IF/THEN commands) identified whether a school leader had any experience in each possible school type (for example, *elem\_yn*, *high\_yn*) and each possible job category (for example, *class\_yn*, *support\_yn*; step 5.d.12).

After the study team created the Jobs by Year dataset (step 5.d.13), the original Jobs dataset became the base for a new dataset identifying the path through job categories that each school leader took, regardless of time spent in each category (step 5.d.14). The study team used an AGGREGATE WITH MIN command to generate one row for each job category and each school leader, retaining the minimum value for *year* for each job category as *year\_min* (step 5.d.15). The variable *year\_min* represented the first time (within the years captured in the dataset) a school leader held that specific job category. A SORT command sorted the cases by *ID* and *year* (step 5.d.16). Then the study team used a RESTRUCTURE command with a numeric index such that for each school leader *job\_cat.1* represented the first job category held and *job\_cat.2* represented the second job category held (step 5.d.17). In the SORT and RESTRUCTURE command sequence, a maximum of seven categories were found, creating seven variables. The restructured dataset, called the Paths dataset, resulted in the following variables: *ID*, *job\_cat.1*, *job\_cat.2*, *job\_cat.3*, *job\_cat.4*, *job\_cat.5*,

*job\_cat.6*, and *job\_cat.7*. For school leaders with fewer than seven job categories, the inapplicable variables were left empty. For example, if a school leader had only two job categories, data were present in *job\_cat.1* and *job\_cat.2*, and *job\_cat.3*, *job\_cat.4*, *job\_cat.5*, *job\_cat.6*, and *job\_cat.7* were left empty.

Next, the study team created a new variable, *path*, representing the path through job categories that each school leader took according to the information in the dataset. A CONCATENATE command strung together each of the *job\_cat* variables in order (step 5.d.18). For example, if *job\_cat.1* was “classroom teacher,” *job\_cat.2* was “assistant principal,” and *job\_cat.3* was “principal,” *path* became “classroom teacher → assistant principal → principal.”

The study team created the final Experiences dataset by merging the Jobs by Year and Paths datasets, using *ID* as the linking variable (step 5.d.20). The resulting dataset had a single row for each school leader and contained all the variables related to the school leader’s job history in the Florida Department of Education from 2001/02 to 2011/12 and all reported job history outside the department.

### Merging all datasets

Once the study team has replicated the cleaning process with matching results obtained, datasets can be merged. The merged datasets should also be inspected and cleaned because new variables may need to be created using variables from different datasets. Once all datasets are clean and merged into the final dataset, analysis can begin.

**Merged datasets should be inspected and cleaned because new variables may need to be created using variables from different datasets. Once all datasets are clean and merged into the final dataset, analysis can begin**

Because the analysis for this project was a retrospective cohort analysis investigating 2011/12 school leaders, it was imperative that the final dataset include only 2011/12 school leaders. Thus, the study team removed all irrelevant cases (step 5.d.11). The Florida Department of Education requested that the analysis focus exclusively on principals and assistant principals, so the dataset was limited to cases where *job\_cat.2011* equaled “assistant principal” or “principal.” Once the Experiences dataset was clean, the study team merged the Certificate and Demographics datasets using *ID* as the linking variable (step 5.e). By using the Experiences dataset as the primary dataset, and the Certificate and Demographics datasets as lookup tables, cases (*IDs*) found in the Certificate or Demographics datasets but not in the Experiences dataset were not merged into the final dataset. The study team named and saved this final dataset as Data for Analysis.

## **Step 6: Analyzing and reporting on the data**

The parts of this step described below are the analytical process undertaken by the study team to conduct the descriptive analysis of the principal workforce in Florida. Whenever the study team obtained an average, the standard deviation, valid *n* (the count of school leaders with data for that variable), and valid *n* percent (the percentage of school leaders with data for that variable) were also obtained with the MEANS command. All figures and tables presented in this section are examples of ways to display results associated with the described analyses; examples are taken directly from the companion report, *A descriptive analysis of the principal workforce in Florida* (Folsom et al., in press). For interpretation and discussion of the results, see the companion report.



**Descriptive demographic data and paired bar chart: Who are Florida's school leaders?**

In step 6.a the study team used FREQUENCY and CROSSTAB commands to ascertain the frequency for the number of school leaders, the number of assistant principals and principals, and the number and percentage of school leaders by job classification, school type, gender, and race/ethnicity.

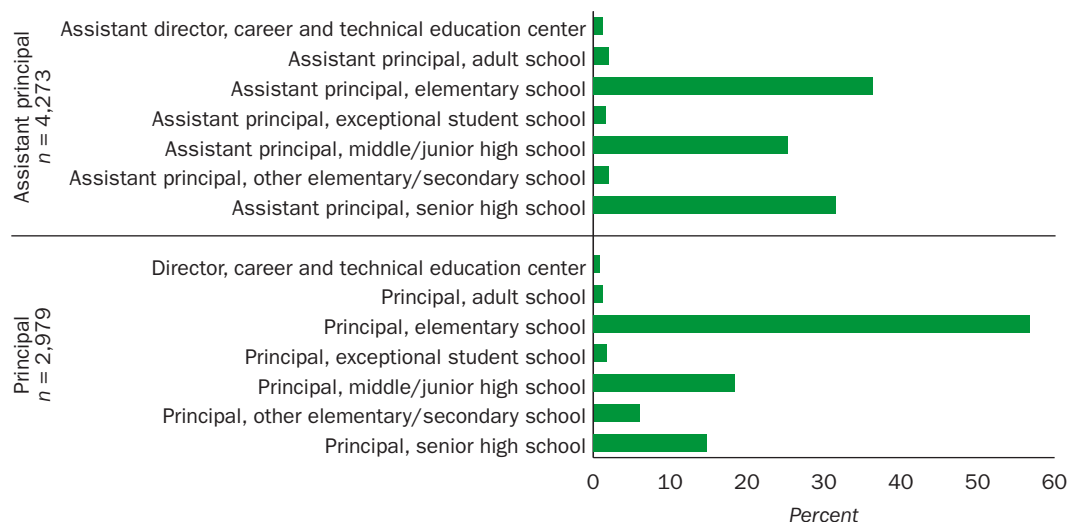
To depict the frequency distribution of school leaders by job classification, the study team created a paneled bar graph using the frequency counts obtained from the statistical analysis and copied to spreadsheet software with graphing capabilities (figure 2). The paneled bar graph was created such that each panel reflected the job category and each bar represented the percentage of school leaders by job classification.

To depict the percentage of school leaders, teachers, and students by racial/ethnic background, the study team created a 100 percent stacked bar graph using the percentage counts obtained from the statistical analysis and copied to spreadsheet software with graphing capabilities (figure 3). The study team obtained the comparison data (demographic composition of teachers and students) from the Florida Department of Education website. The bar graph was created such that each bar represented the categories of assistant principal, principal, teacher, and student and each color represented the percentage of individuals for each race/ethnicity.

Next, a series of MEANS and FREQUENCY commands provided the average age of assistant principals and principals as well as the frequency of school leaders at each age (step 6.b). To depict the distribution of school leaders by age range, the study team created a paired bar chart using the percentages obtained from the frequencies analysis and copied to spreadsheet software with graphing capabilities (figure 4). The paired bar chart was created such that each bar covered a range of five years. The height of the bar represented the percentage of school leaders in that particular age range, and each color represented a school leader type.

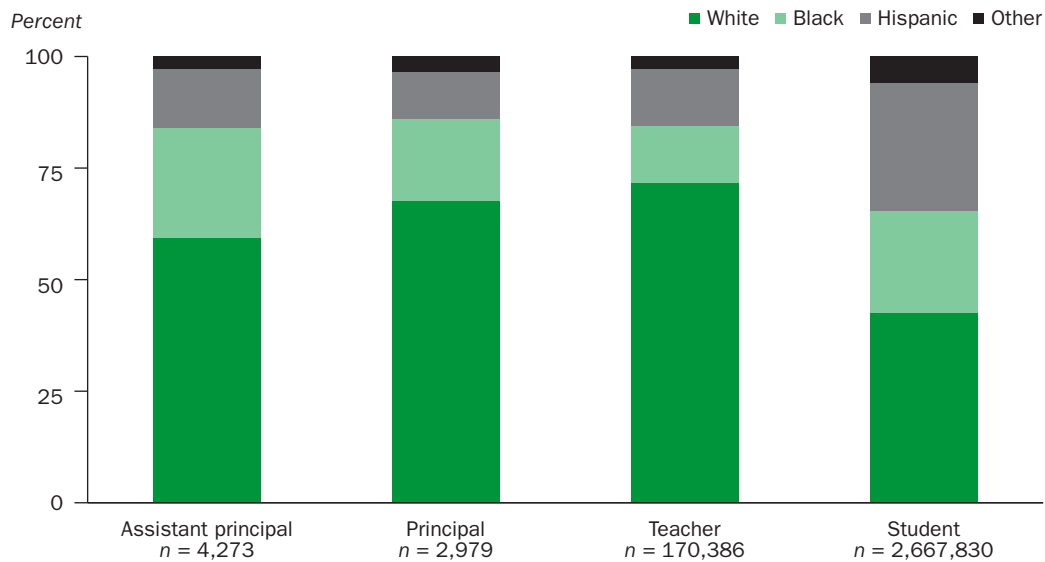
*To depict the frequency distribution of school leaders by job classification, the study team created a paneled bar graph using the frequency counts obtained from the statistical analysis and copied to spreadsheet software with graphing capabilities*

**Figure 2. Example of a paneled bar graph of job titles**



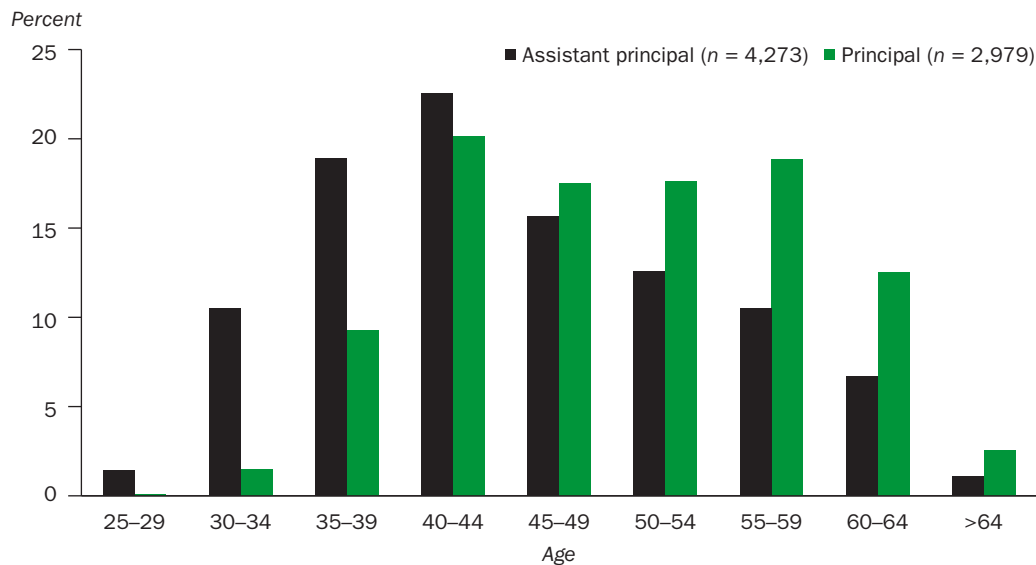
Source: Folsom et al., in press.

**Figure 3. Example of a 100 percent stacked bar graph of demographic composition**



Source: Folsom et al., in press.

**Figure 4. Example of a paired bar chart of ages**



Source: Folsom et al., in press.

**Descriptive background data: What are the backgrounds of Florida's school leaders?**

In step 6.c a series of MEANS, FREQUENCY, and CROSSTAB commands provided the following information by the school type in which the school leader served during 2011/12:

- Average number of coverages ever held and active coverages held by school leaders.
- The percentage of school leaders holding an active administrative, subject, endorsement, or vocational coverage.
- The percentage of school leaders holding an active coverage by each of the specific administrative coverage and specific instruction-level coverages.

The study team copied the analysis results to spreadsheet software (table 2).

### Descriptive path data: Where have Florida's school leaders served and in what capacity?

A series of MEANS, FREQUENCY, and CROSSTAB commands provided the following in step 6.d:

- The average number of years of experience in each job category and in each of the self-reported other job experiences, by school leader.
- The frequency counts and average number of years of data, which represents the average number of years of active employment in a Florida public school from 2001/02 to 2011/12, by school leader.
- The frequency counts and average number of districts, schools, school types, and job categories, by school leader.
- The frequency counts and average number of years spent in multiple positions across the years, by school leader.

**Table 2. Example of table of descriptive statistics of certifications**

		Elementary		Middle		High		Combination		Adult	
		Assistant principal (n = 1,463)	Principal (n = 1,674)	Assistant principal (n = 989)	Principal (n = 496)	Assistant principal (n = 1,404)	Principal (n = 541)	Assistant principal (n = 284)	Principal (n = 203)	Assistant principal (n = 92)	Principal (n = 37)
Average number of coverages ever held in the Florida Department of Education	Mean	9.4	11.8	9.2	12.2	9.1	11.6	9.0	11.6	12.1	13.8
	Standard deviation	4.3	4.5	4.3	5.0	4.6	5.4	4.5	5.1	5.3	6.8
Average number of active coverages	Mean	2.7	2.7	2.5	2.8	2.4	2.6	2.5	2.5	2.8	2.9
	Standard deviation	1.6	1.7	1.6	1.7	1.6	1.7	1.6	1.8	1.7	2.2
At least one active administrative coverage	Percent	98.8	99.3	98.0	99.0	98.3	95.7	95.8	92.5	95.7	97.4
At least one active subject coverage	Percent	98.1	98.5	97.5	96.4	96.8	92.2	95.1	92.9	96.8	89.5
At least one active endorsement	Percent	47.3	40.2	31.9	35.1	28.0	28.8	38.5	27.4	36.2	26.3
At least one active vocational coverage	Percent	0.5	0.4	0.4	1.2	1.2	2.0	0.0	0.9	3.2	13.2
<b>Specific administrative coverage</b>											
School leadership	Percent	88.1	50.8	85.6	48.3	83.0	46.1	85.1	46.2	68.1	42.1
School principal	Percent	25.7	82.5	29.3	84.4	28.5	75.2	21.5	65.6	23.4	57.9
Administration/supervision	Percent	1.6	2.9	2.1	3.8	3.0	1.8	0.7	4.7	9.6	10.5
Local director of vocational education	Percent	0.0	0.1	0.1	0.0	1.2	2.5	0.0	0.9	5.3	21.1
Administration of adult education	Percent	0.3	0.2	0.1	0.0	0.6	0.5	0.3	0.9	11.7	18.4
<b>Instruction level of subject coverage</b>											
All levels	Percent	31.8	28.9	32.6	25.9	30.9	25.7	37.5	42.0	28.7	31.6
Prekindergarten	Percent	6.2	12.2	0.7	2.8	0.7	1.3	1.7	1.4	2.1	0.0
Elementary	Percent	78.8	76.4	29.9	30.3	19.6	21.9	45.8	43.4	25.5	15.8
Secondary	Percent	19.5	20.1	66.6	70.7	71.8	72.9	43.4	42.5	69.1	68.4

Source: Folsom et al., in press.

- The average number of years spent in each school type, by the school type in which the school leader served in 2011/12.

The study team created two tables by copying the results of the analyses to spreadsheet software. The first table displayed the general experience results (table 3), and the second table displayed the results of experience by school type (table 4).

To depict the movement between school types across years for all school leaders, the study team created a paneled 100 percent stacked bar graph. A CROSSTAB command obtained the frequency counts and percentage of 2011/12 elementary, middle, senior high, combination, and adult school leaders in each school type for each school year from 2001/02 to 2010/11. The study team copied the results to spreadsheet software with graphing capabilities, where each bar represented the year and each color represented the school type the school leader served; bars were grouped by the school type in which the school leader served in 2011/12 (figure 5).

*To depict the movement between school types across years for all school leaders, the study team created a paneled 100 percent stacked bar graph*

To depict the movement between job categories across years for school leaders, the study team created a paneled 100 percent stacked bar graph (figure 6). A CROSSTAB command obtained the frequency counts and percentage of 2011/12 assistant principals and principals in each job category for each school year from 2001/02 to 2010/11. The study team copied the results to spreadsheet software with graphing capabilities, where each bar represented the year and each color represented the job category held by the school leader that year; bars were grouped by the job category held by the school leader in 2011/12.

A FREQUENCY command provided the frequency and percentages for each possible path the 2011/12 assistant principals and principals took from 2001/02 to 2011/12. Based on conversations with the Florida Department of Education, the study team determined that paths with frequency counts less than 10 were considered a “unique path,” and paths

**Table 3. Example of table of descriptive statistics of experience types**

Position	Assistant principal			Principal		
	Percent with any experience in this position	Years of experience among those with any experience		Percent with any experience in this position	Years of experience among those with any experience	
		Mean	Standard deviation		Mean	Standard deviation
Experience in Florida from 2001/02 to 2011/12						
Classroom instruction	72.1	3.9	2.0	30.4	2.5	1.3
Other instruction	16.6	2.3	1.5	6.7	2.4	1.5
Support services	6.8	3.0	2.0	3.0	2.1	1.3
General administration	4.9	2.2	1.3	2.0	1.7	1.0
Assistant principal	100.0	5.2	3.2	69.6	3.9	2.0
Principal	3.3	3.3	2.3	100.0	5.6	3.4
Superintendent's/district office	10.2	2.3	1.7	10.5	2.3	1.8
Self-reported experience outside Florida Department of Education public schools, including all years of work history						
Military service	4.1	7.0	6.3	4.0	9.0	8.2
Teaching in Florida nonpublic schools	5.3	3.8	3.4	6.2	4.1	3.6
Teaching out-of-state nonpublic schools	3.7	3.9	3.3	4.1	3.7	3.0
Teaching in out-of-state public schools	13.9	5.5	5.0	18.1	5.3	4.4

Source: Folsom et al., in press.

**Table 4. Example of table of descriptive statistics of experience by school type**

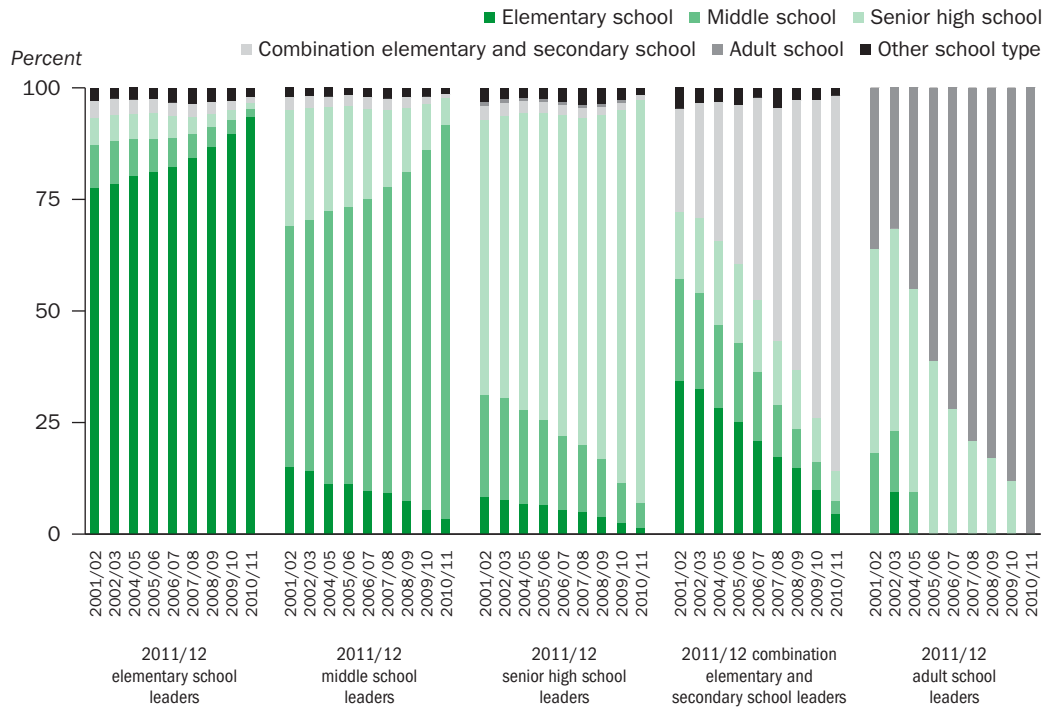
2011/12 school type	Assistant principal					Principal				
	Elementary (n = 1,467)	Middle (n = 996)	Senior high (n = 1,420)	Combination (n = 296)	Adult (n = 94)	Elementary (n = 1,670)	Middle (n = 496)	Senior high (n = 556)	Combination (n = 220)	Adult (n = 37)
Current and previous years of experience working in type of Florida public schools										
<i>Elementary</i>										
Mean	7.9	0.8	0.4	1.7	0.4	8.8	0.9	0.6	1.6	0.2
Standard deviation	2.9	1.8	1.2	2.6	1.3	2.9	2.1	1.6	2.5	1.0
<i>Middle</i>										
Mean	0.6	6.7	1.3	1.3	0.6	0.5	7.1	1.5	0.9	0.4
Standard deviation	1.6	3.2	2.2	2.4	1.4	1.3	3.4	2.3	2.0	1.3
<i>Senior high</i>										
Mean	0.4	1.5	7.3	1.2	2.4	0.3	1.7	7.1	0.9	1.3
Standard deviation	1.3	2.4	3.2	2.3	2.9	1.2	2.7	3.3	2.0	2.1
<i>Combination</i>										
Mean	0.2	.2	0.2	4.4	0.3	0.2	0.1	0.2	5.7	0.3
Standard deviation	1.0	1.0	0.8	3.1	1.2	0.9	0.6	0.9	3.3	0.9
<i>Adult</i>										
Mean	0.0	0.0	0.1	0.1	5.7	0.0	0.0	0.0	0.0	7.0
Standard deviation	0.1	0.1	0.6	0.5	3.5	0.2	0.0	0.1	0.3	3.3
<i>Other</i>										
Mean	0.3	0.2	0.2	0.3	0.1	0.2	0.2	0.3	0.3	0.3
Standard deviation	1.0	0.7	0.9	1.0	0.6	0.9	0.9	1.0	1.1	0.9
Percentage with any experience from 2001/02 to 2011/12										
Elementary	100.0	21.6	11.6	40.5	10.6	100.0	22.2	14.2	36.4	5.4
Middle	18.1	100.0	34.1	31.4	20.2	14.9	100.0	40.6	24.1	13.5
Senior high	11.5	35.9	100.0	29.4	53.2	9.8	39.9	100.0	20.9	35.1
Combination	8.2	7.6	6.8	100.0	9.6	8.1	4.2	7.0	100.0	13.5
Adult	0.3	0.5	1.5	1.4	100.0	0.4	0.0	0.5	0.5	100.0
Other	13.4	6.8	9.2	11.8	6.4	9.8	9.1	12.8	14.1	16.2

Source: Folsom et al., in press.

with 10 or more frequency counts were considered a “common path.” In instances where there was only one position (either there was only one year of data for the school leader or the school leader held the same job category across all years of available data), the school leader was considered as having had “no change.” The study team copied the frequency output to spreadsheet software and created a table with eight columns—four columns for the assistant principal paths and four for the principal paths—that held the final data (table 5). The first column listed the path description (for example, “Class → Assistant principal,” “Support → Assistant principal → Principal”), the second listed the frequency count for the path, the third listed the percentage of movers (those that had more than one position in the dataset) who took that path, and the fourth listed the percentage of all school leaders who took that path. The study team calculated the values in the third column by dividing the frequency of that path by the number of school leaders who moved. The values in the fourth column were calculated by dividing the frequency of that path by the number of all school leaders.

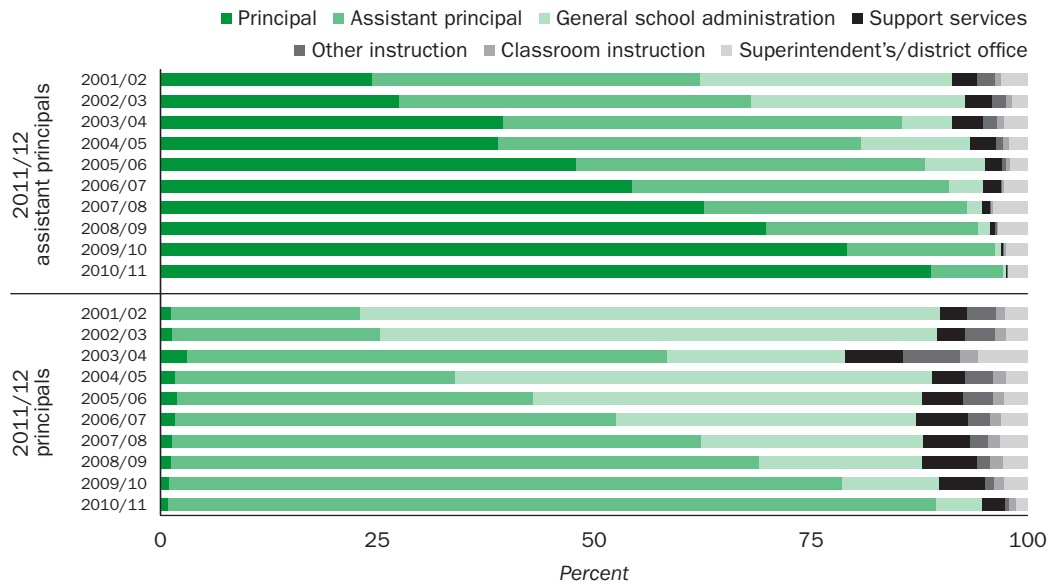
Three additional rows in the table summarized the movement among assistant principal and principal paths. The first row reflected the number of school leaders with no change,

**Figure 5. Example of a 100 percent stacked bar graph of movement between school types**



Source: Folsom et al., in press.

**Figure 6. Example of a 100 percent stacked bar graph of movement between job categories**



Source: Folsom et al., in press.

the second reflected the number of school leaders who took a common path, and the third reflected the number of school leaders who took a unique path. The study team calculated the values in the second row by summing the values for each of the common paths; values in the third row were calculated by subtracting the sum of the values of the first row (no movement) and the second row (common paths) from the total number of school leaders.

**Table 5. Example table of descriptive statistics of career paths**

Assistant principal (n = 4,273)				Principal (n = 2,979)			
Path	n	Percent who moved between 2001/02 and 2011/12	Percent of all assistant principals	Path	n	Percent who moved between 2001/02 and 2011/12	Percent of all principals
No change	741	—	17.3	No change	643	—	21.6
Common path	3,385	95.8	79.2	Common path	2,222	95.1	74.6
Unique path	147	4.2	3.4	Unique path	114	4.9	3.8
Classroom instruction→ Assistant principal	1,929	54.6	45.1	Assistant principal→ Principal	1,021	43.7	34.3
Classroom instruction→ Other instruction→Assistant principal	392	11.1	9.2	Classroom instruction→ Assistant principal→ Principal	615	26.3	20.6
Classroom instruction→ Superintendent's/district office→ Assistant principal	174	4.9	4.1	Superintendent's/district office→Principal	108	4.6	3.6
Classroom instruction→ General administration→ Assistant principal	109	3.1	2.6	Assistant principal→ Superintendent's/district office→Principal	67	2.9	2.2
Support services→ Assistant principal	105	3.0	2.5	Classroom instruction→ Principal	64	2.7	2.1
Classroom instruction→ Support services→ Assistant principal	96	2.7	2.2	Other instruction→ Assistant principal→ Principal	55	2.4	1.8
Superintendent's/district office→ Assistant principal	81	2.3	1.9	Classroom instruction→ Other instruction→ Assistant principal→Principal	52	2.2	1.7
Principal→Assistant principal	79	2.2	1.8	Superintendent's/district office→ Assistant principal→Principal	43	1.8	1.4
Other instruction→ Classroom instruction→ Assistant principal	70	2.0	1.6	Support services→ Assistant principal→ Principal	42	1.8	1.4
Other instruction→ Assistant principal	64	1.8	1.5	Classroom instruction→ Superintendent's/district office→ Assistant principal→Principal	36	1.5	1.2
Superintendent's/district office→ Classroom instruction→ Assistant principal	51	1.4	1.2	Classroom instruction→ General administration→ Assistant principal→Principal	24	1.0	0.8
Classroom instruction→ Other instruction→Superintendent's/ district office→Assistant principal	38	1.1	0.9	Other instruction→ Principal	21	0.9	0.7
Classroom instruction→ Superintendent's/district office→ Other instruction→Assistant principal	32	0.9	0.7	Classroom instruction→ Other instruction→ Principal	20	0.9	0.7

(continued)

**Table 5. Example table of descriptive statistics of career paths (continued)**

Assistant principal (n = 4,273)				Principal (n = 2,979)			
Path	n	Percent who moved between 2001/02 and 2011/12	Percent of all assistant principals	Path	n	Percent who moved between 2001/02 and 2011/12	Percent of all principals
General administration→ Assistant principal	30	0.8	0.7	General administration→ Assistant principal→Principal	19	0.8	0.6
Classroom instruction→Principal→ Assistant principal	22	0.6	0.5	Classroom instruction→ Assistant principal→ Superintendent's/district office→ Principal	17	0.7	0.6
Support services→ Classroom instruction→ Assistant principal	20	0.6	0.5	Classroom instruction→ Support services→ Assistant principal→Principal	16	0.7	0.5
Classroom instruction→ Other instruction→ General administration→ Assistant principal	16	0.5	0.4				
Principal→Superintendent's/district office→Assistant principal	14	0.4	0.3				
Other instruction→ General administration→ Assistant principal	14	0.4	0.3				
General administration→ Classroom instruction→ Assistant principal	13	0.4	0.3				
Superintendent's/district office→ Classroom instruction→ Other instruction→Assistant principal	13	0.4	0.3				
Other instruction→ Classroom instruction→ Superintendent's/district office→ Assistant principal	12	0.3	0.3				
Superintendent's/district office→ Other instruction→Assistant principal	11	0.3	0.3				

Source: Folsom et al., in press.

### **A word of caution about using administrative data systems**

Administrative data systems can differ by jurisdiction. The processes described in this report are specific to Florida Department of Education data. The procedures applied in this report may not apply to data in other states.

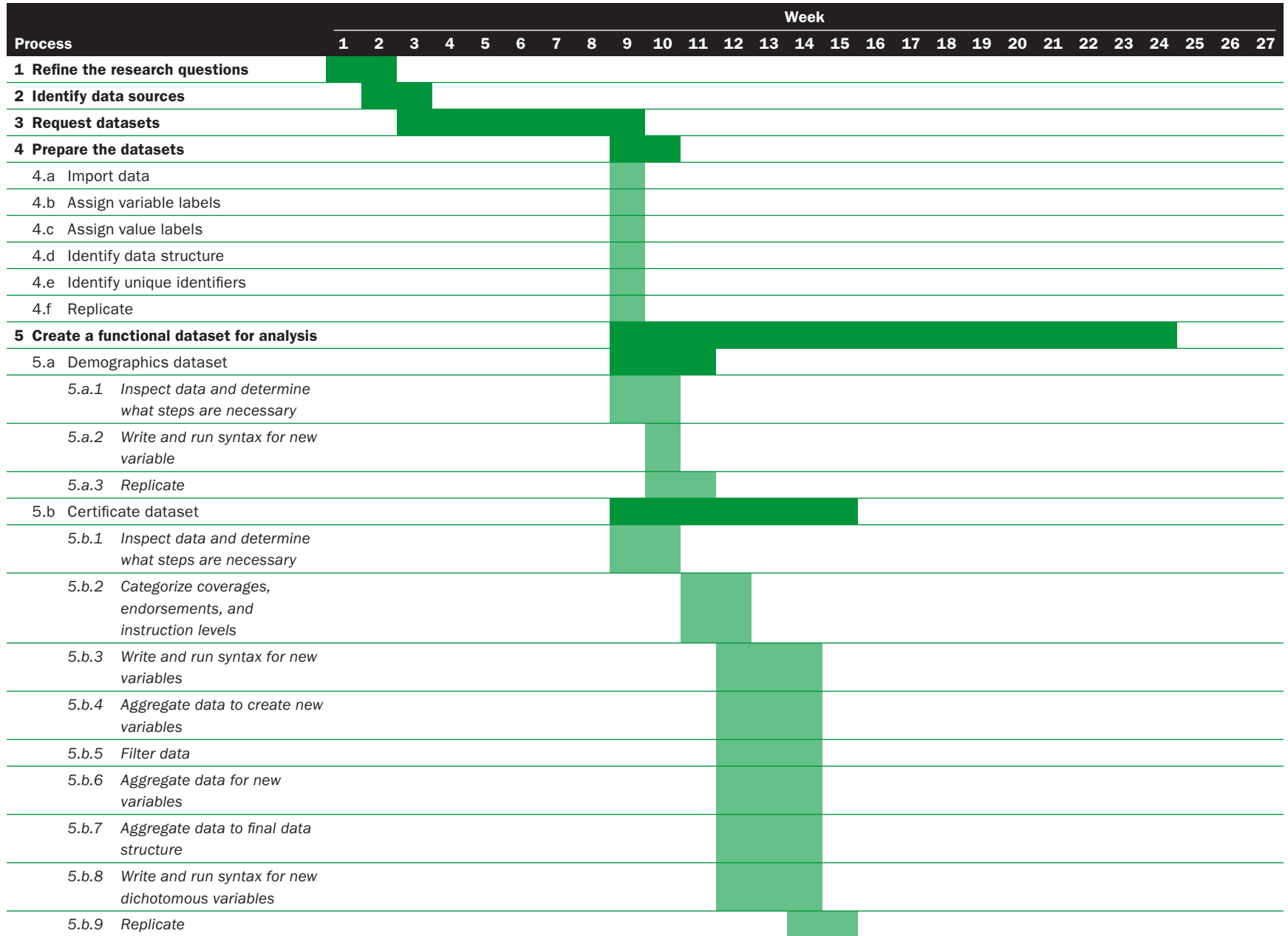
The methods described in this guide are for a retrospective cohort analysis. It does not examine workforce trends across time, but instead provides a snapshot of a selected cohort.

Though the Florida Department of Education has one of the most extensive education databases in the country, and despite efforts to carefully review and clean datasets and rerun analyses, there were instances of missing data. Challenges with data quality are not unique to Florida (see, for example, Clifford et al., 2012). Despite compliance with federal and state guidelines and efforts to ensure data security and maintenance, consistent reporting across districts is often difficult to obtain. Data issues can vary across districts within



a state. For example, as became evident in conversations with the Florida Department of Education and as noted throughout this report, districts often have flexibility in what data are collected, what is reported, and when it is reported. This was most evident in the data related to self-reported previous experiences.

Finally, administrative data systems evolve. Variables—and codes for variables—may be added or removed. The data collection and management systems may change, so issues with a univariate data structure may not apply in other data management systems that collect and store data in a multivariate structure. New issues may arise in other data systems that were not applicable in the scenarios described in this report. Therefore, study teams must carefully consider the unique features of each administrative data system before embarking on analyses.







## **Appendix B. Key terms**

**Command.** A text statement used when writing syntax instructing a software package to do something. Each software package has unique commands. The study team used SPSS for all analyses described in this report, so the commands (identified in all capital letters) are specific to SPSS. Readers will need to consult with the documentation provided with their software package for the specific commands for each process described in this report. The following commands were used:

- AGGREGATE. Collapses data and can be used to create new datasets or new variables, which can be counts of cases or summaries of other variables.
- AGGREGATE WITH MAX. Summarizes the maximum value of a specific variable.
- AGGREGATE WITH MIN. Summarizes the minimum value of a specific variable.
- COMPUTE. Creates a new variable or calculates new values for an existing variable using logical mathematical expressions.
- CONCATENATE. Combines existing string variables or other strings together to create a singular new variable.
- CROSSTAB. Creates a table to summarize at least two variables, one of which is categorical.
- FILTER. Selects in or selects out cases based on logical values.
- FREQUENCY. Creates a table summarizing the frequency of selected variables.
- GRAPH. Creates graphs whose types are specified with additional commands such as BAR or HISTOGRAM.
- IF/THEN. Executes a command (such as create a variable) only if a certain logical expression is met.
- MEANS. Creates a table summarizing continuous variables.
- MERGE. Combines multiple datasets into a single dataset using a common variable across datasets.
- RESTRUCTURE. Restructures data such that multiple cases per ID are transformed into variables.
- SORT. Sorts data in a specific order by a certain variable.
- VALUE LABELS. Assigns labels for values for categorical variables.
- VARIABLE LABELS. Creates brief descriptive labels for variables.

**Common path.** A career path taken between 2001/02 and 2011/12 by at least ten 2011/12 school leaders.

**Database.** A broader term used to describe a collection of datasets. In this report the Educational Staff database from the Education Data Warehouse was the original database; additional data were requested from the Employee Demographics, Certified Staff, and Educational Staff datasets.

**Dataset.** A specific file with a set of variables containing data. This report refers to several datasets used to create the final dataset “Dataset for Analysis.” The data obtained from the Florida Department of Education’s Education Data Warehouse were in the Employee Demographics, Certified Staff, and Educational Staff datasets.

**Data source.** A broader term referring to a specific entity that hosts data. In this report the data source is the Florida Department of Education’s Education Data Warehouse.

**Extant data.** Existing data collected by another entity. In this report the Florida Department of Education collected the staffing data.

**Florida educator certificate.** A certificate obtained from the Florida Department of Education.

- *Coverage.* Used by the department to describe the specific subjects that the Florida educator certificate covers for each school leader certificate. See <http://www.fldoe.org/edcert/subjlist.asp> for more details.
- *Instruction level.* Each coverage of the Florida educator certificate has an associated instruction level. Current<sup>3</sup> instruction level coverages include:
  - All levels. Typically K–12.
  - Elementary. Typically K–6.
  - Secondary. Typically 6–12, may include 5–9.
  - Prekindergarten. Typically birth–age 4.
  - District designation. The department does not always identify the instruction level of coverage; districts can designate the instruction level.

**Job category.** The study's broad categories related to the Florida Department of Education's job classifications. The specific job categories, with examples of department job classifications, are:

- *Assistant principal.* Limited to those identified as assistant principals or assistant directors of vocational/technical centers. According to the department's definition, assistant principals are staff members assisting the administrative head of the school.
- *Classroom instruction.* Job classifications that involve student- or classroom-level instruction. Examples include but are not limited to intermediate resource teacher, teacher of language arts, teacher of music, or teacher of varying exceptionalities.
- *General administration.* Job classifications for school-level administration other than assistant principal or principal classifications. Examples include but are not limited to administrative assistant, school clerical staff, registrar, or school secretary.
- *Principal.* Limited to those identified as principals, as defined by administrative rule 6A-4.0083,<sup>4</sup> or directors of vocational/technical centers. According to the department's definition, principals are staff members assigned as the administrative head of a school and delegated responsibility for coordinating and directing the activities of the school.
- *Other instruction.* Involves higher instruction or instruction of other professionals. Examples include but are not limited to computer systems user, educator of instructional technology, math coach, reading coach, or school librarian/media specialist.
- *Superintendent's/district office.* A job category specifically reserved for school leaders serving in the superintendent's office and indicated by the special-use school number 9001 (such that the school number rather than the job code identifies the job classification). Examples include but are not limited to district dropout prevention specialist, learning resource specialist, director of instruction/curriculum, or program specialist.
- *Support services.* Includes job classifications that provide special support services to students, teachers, or administrators but do not necessarily involve direct instruction. Examples include but are not limited to administrator on special assignment for guidance services, coordinator of pupil personnel services, counselor, diagnostic specialist, dropout prevention specialist, or parent education specialist.

**Job classification.** The specific job code assigned by the Florida Department of Education.<sup>5</sup> For this study, several job classifications were divided into the distinct job categories described.

**Retrospective cohort analysis.** Analysis that looks back in time at events specific to a group (cohort) of individuals. This study is a retrospective cohort analysis of 2011/12 school leaders.

**School leader.** All assistant principals and principals, including those with the interim/intern designation.

**School type.** Instruction level of the school designated by the Florida Department of Education.<sup>6</sup> The specific school types are:

- *Elementary schools.* Schools providing instruction at one or more grade levels from prekindergarten through 5. May include schools serving grade 6 if also serving one or more grades from prekindergarten through 5 (for example, a K–6 school).
- *Middle schools.* Schools providing instruction in middle school configurations (grades 6–8) and junior high school configurations (grades 7–9). Can also include schools serving a single grade in the 6–8 range (for example, a sixth-grade center).
- *High schools.* Schools providing instruction at one or more grade levels from 9 to 12. Includes regular high schools and ninth-grade centers.
- *Combination elementary and secondary schools.* Schools providing instruction in grade groupings that include more than one of the categories described above (for example, prekindergarten–8, K–12).
- *Adult schools.* Schools providing instruction to adult learners.
- *Other.* Schools that do not fall into one of the above categories. Typically, these schools are part of special-use school numbers such as the superintendent’s or district office.

**Tab-delimited text.** A simple text format used for storing data in a tabular structure where each value is separated by a tab. This type of file is commonly used when sharing data because it can be imported or read by all statistical software packages.

**Unique path.** A 2001/02–2011/12 path followed by fewer than ten 2011/12 school leaders.

**Work experience type.** The following codes were listed as possible work experience types that districts collected from school leaders and reported to the Florida Department of Education. As indicated in this report, several codes were determined to be inaccurate and removed.

- Administration in education.
- Military service.
- Service to the district in current job code assignment.
- Teaching in current district.
- Teaching in Florida nonpublic schools.
- Teaching in Florida public schools.
- Teaching in out-of-state nonpublic schools.
- Teaching in out-of-state public schools.

## Appendix C. Sample data request

### Project title: Characteristics of school leaders

**Abstract.** This project responds to a request from the Florida Department of Education to evaluate effective leaders—a neglected area of Florida’s Race to the Top grant. Regional Educational Laboratory Southeast worked with the Deputy Chancellor for Education Quality; bureau chiefs from Educator Recruitment, Development, and Retention; the Director of Research and Analysis in Educator Performance; and staff from Management Information Systems and from Educator Recruitment, Development, and Retention to develop research questions and identify data to describe Florida’s current pool of principals.

**Research questions.** We are requesting data to address the following research questions:

- What are the career paths of Florida’s 2011/12 school leaders?
- What are the backgrounds of Florida’s 2011/12 school leaders?

**Cohorts and characteristics.** We would like the personnel data for each 2011/12 principal and assistant principal for each year from 2001/2002 to 2011/12. This will allow us to analyze employment trends over the past 10 years.

**Methodology.** All analyses will be descriptive. Simple charts representing frequencies and distributions will be created. Where applicable, means or other measures of central tendency and standard deviations or other measures of variability will be calculated. The specific descriptive analyses we will conduct include:

- Demographic composition of principals (age, race/ethnicity, gender).
- Length of education-related service in Florida:
  - As a teacher.
  - As a district administrator.
  - As an assistant principal.
  - As a principal—total, at current school, and at previous school or schools.
- Number of Florida schools served as principal (including current school).
- Types of Florida schools served as principal (elementary, middle, high, mixed).
- Pathways to principal position (for example, teacher → assistant principal → principal; administrator → assistant principal → principal; other) and the number of principals that have taken the various tracks.

### Requested data elements

- For 2011/12 only:
  - Demographics.
  - Unique identifier to connect datasets.
  - Birthdate.
  - Race.
  - Ethnicity.
  - Gender.
  - Florida educator certificate.
  - Expiration year.
  - Number.
  - Subject coverage.
  - Type.



- For each year from 2001/02 to 2011/12:
  - School information.
  - District number.
  - School.
  - Job code.
  - Experience and employment.
  - Teaching experience.
  - Experience type.
  - Experience length.
  - Employment date, continuous employment.
  - Employment date, current position.
  - Employment date, original position.

## Appendix D. Data dictionary

This appendix describes the original variables from each of the original datasets.

### Original Demographics dataset

Variable name	Variable description	Note
ID	Unique ID for each school leader	Used to link school leaders across datasets
birth_year	Year of birth	Used to describe the age of the sample (age calculated from 2011)
racial_ethnic_cd	Race/ethnicity of the school leader	The Florida Department of Education combines race and ethnicity. Variable used to describe the sample as White, Black, Hispanic, or other
gender	Gender	Used to describe the sample

### Original Certificate dataset

Variable name	Variable description	Note
ID	Unique ID for each school leader	Used to link school leaders across datasets
expiration_date	Expiration date of the coverage	Used to determine the number of active versus expired coverages
effective_date	Date coverage became effective	
requirement_name	Code identifying if the coverage is a specific Florida Educator Certification subject area or an endorsement	Used to determine the number of subject area coverages and endorsements
certification_subject_cd	Numeric code associated with the Florida Educator Certification coverage	Used to identify the number of administrative, subject area, vocational, and endorsement coverages
certification_subject_name	Name of the specific Florida Educator Certification coverage	
instruction_level	Numeric code indicating the instruction level of the coverage	Used to determine the number of instruction level coverages
instruction_lvl_short_desc	Short description of the instruction level of the coverage	
requirement_subtype_name	Code differentiating between a Florida Educator Certification coverage and an endorsement	Used to determine the number of subject area coverages and endorsements

## Original Job Experiences dataset

Variable name	Variable description	Note
year	Year associated with the data	Used as an index variable in restructuring the dataset
ID	Unique identifier for each school leader	Used to link school leaders across datasets
source_system_class_cd	Numeric code assigned to the Florida Department of Education specific job classification	Used to categorize into job categories and identify movement between job categories
job_classification_name	Name of the Florida Department of Education specific job classification	
district	Florida district where employed	Used to identify movement between districts
school	School where employed	Used to identify movement between schools
classification_hire_date	District reported hire date of the job classification	Used to identify most current position when multiple positions held in a year
work_experience_type_cd	Numeric code assigned to the Florida Department of Education–defined, self-reported work experiences	Used to describe experiences outside of the Florida Department of Education only
work_experience_type_name	Name of the Florida Department of Education–defined, self-reported work experiences	
total_years_amt	Self-reported number of years spent in the Florida Department of Education–defined work experiences	

## Notes

1. The term “coverage” describes the specific subjects that the Florida educator certificate covers for each school leader certificate.
2. Once issued, a Florida Educator Certificate coverage or endorsement can be renewed indefinitely with the original name and instruction level, following Florida Department of Education regulations.
3. Some previous certifications and endorsements have instruction levels with different grade ranges. As part of the data cleaning, old instruction levels were matched to current Florida Department of Education instruction levels. For example, an older certification, Primary Education, had an instruction level of K–3, which now falls under the elementary instruction level.
4. Administrative rule 6A-4.0083 states, “To be eligible to receive certification as a school principal, an individual shall satisfy each of the following requirements:
  - (1) Hold a valid professional certificate covering school leadership, administration, or administration and supervision.
  - (2) Document successful performance of the duties of the school principalship. These duties shall be performed in a Department of Education approved district school principal certification program pursuant to Rule 6A-5.081, F.A.C., designed, and implemented consistent with the principal leadership standards approved by the State Board of Education. In addition, these duties shall:
    - (a) Be performed as a full-time employee in a Florida public school in a leadership position through which the candidate can fully demonstrate the competencies associated with the Florida Principal Leadership Standards.
    - (b) Be a formally planned professional development program designed and implemented to prepare the individual to effectively perform as a school principal.
    - (c) Be comprehensive of all the duties of the school principalship.
    - (d) Be performed under the direct supervision of a currently practicing school principal or district manager who has been approved by the district school board to serve as the supervising principal or manager for this program.
  - (3) Demonstrate successful performance of the competencies of the school principalship standards which shall be documented by the Florida district school superintendent based on a performance appraisal system approved by the district school board and the Department pursuant to Rule 6A-5.081, F.A.C.
  - (4) An individual who holds a valid Florida Educator’s Certificate covering administration or administration and supervision issued prior to July 1, 1986 and served as a school principal prior to July 1, 1986 for not less than one (1) school year may apply for certification as a school principal under the provisions of Rule 6A-4.0085, F.A.C.” (<http://www.fldoe.org/edcert/rules/6A-4-0083.asp>).
5. The job codes can be accessed at [http://www.fldoe.org/eias/dataweb/database\\_0809/sfappende.pdf](http://www.fldoe.org/eias/dataweb/database_0809/sfappende.pdf).
6. The designations can be accessed at <http://www.fldoe.org/eias/dataweb/tech/msid.pdf>.

## References

- Clifford, M., Condon, C., Greenberg, A., Williams, R., Gerdeman, R. D., Feters, J., et al. (2012). *A descriptive analysis of the principal workforce in Wisconsin* (Issues & Answers Report, REL 2012–135). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. <http://eric.ed.gov/?id=ED530569>
- Folsom, J. S., Osborne-Lampkin, L., & Herrington, C. D. (in press). *A descriptive analysis of the principal workforce in Florida schools* (REL 2015–068). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

## The Regional Educational Laboratory Program produces 7 types of reports



### **Making Connections**

Studies of correlational relationships



### **Making an Impact**

Studies of cause and effect



### **What's Happening**

Descriptions of policies, programs, implementation status, or data trends



### **What's Known**

Summaries of previous research



### **Stated Briefly**

Summaries of research findings for specific audiences



### **Applied Research Methods**

Research methods for educational settings



### **Tools**

Help for planning, gathering, analyzing, or reporting data or research