

Data Management Plan and Dataset Documentation

Section 1: Overview

Study Information

Report Title:	<i>Effects of reclassifying English learner students on student achievement in New Mexico</i>
REL:	<i>REL Southwest</i>
REL COR Contact Information:	<i>Chris Boccanfuso, Chris.Boccanfuso@ed.gov</i>
PI Contact Information:	<i>Rachel Garrett, rgarrett@air.org</i>
Authors:	<i>Garrett, Rachel Larsen, Eric Arellanes, Melissa</i>
Report Abstract:	<i>This study examined how attaining English proficiency and being reclassified as fluent English proficient affected achievement in English language arts and math in the first year after student reclassification in grades 3–8 in New Mexico. State policy in New Mexico bases student reclassification decisions on whether students attain a minimum overall English language proficiency level score of 5.0 on the ACCESS for ELLs (ACCESS) assessment. The study focused on achievement among English learner students in 2014/15–2018/19, a time when the ACCESS underwent a standards setting process to better align its language proficiency scoring scale with the expectations of college- and career-ready standards. After the standards setting, a smaller percentage of English learner students in New Mexico attained English proficiency and were reclassified each year. At the same time, students who scored near the English proficiency level required for reclassification performed above the statewide average in English language arts and math and were more likely to meet state content proficiency standards. However, the study found no effects of reclassification on student achievement either before or after the ACCESS standards setting. In addition, the study found no effect of reclassification on next-year English language arts and math achievement among most groups of students with different characteristics and among most districts in the study. Leaders at the New Mexico Public Education</i>

Department could use the findings of this study to consider maintaining the current reclassification threshold. In addition, the state and its districts might want to identify opportunities to strengthen the supports provided to English learner students. This could begin by collecting more systematic information on the education services and supports that English learner students receive leading up to and after they attain English proficiency.

URL for IES Publication Page: <https://ies.ed.gov/ncee/rel/Products/Publication/100814>

URL for Study Pre-Registration: *Not applicable*

If consent form(s) was/were signed by study participants or caregivers, is the form reproduced as appendix to report?
If not, include as appendix to this template

Administrative Data Provider #1

Agency and Office: *New Mexico Public Education Department*

Key Point of Contact: *Matthew Goodlaw*

Summary of agreed-upon confidentiality provisions or restrictions on data sharing: *Compliance with FERPA and all New Mexico laws and regulations; to use the data only be used for the Project; no data sharing with any other entity without prior written approval from the Department and no transference of access or storing privileges to any other individual, institutes, or entity; no merging with other data without prior written permission; no disclosures that identify individual students and no report of aggregate data based on an identifiable group of fewer than ten students; return of files to the Department and destruction/purging of data files when the project is completed or no later than November 30, 2022.*

Additional information, if applicable (e.g., provider IRB number, MOU reference number, or other project identifiers): *MOU signed on 7/31/2020 by R. Stewart, R. Garrett, E. Larsen and J. Mouer.*

Section 2. Data Sources

Source #1 State Standardized Assessment Data

Source Name:	<i>State Standardized Assessment Data</i>
Administrative Data Provider, if applicable, from Section 1	<i>NMPED (#1)</i>
Provider File Name(s):	<i>AIR_RELSW_1_2014_AVT_subset.csv AIR_RELSW_1_2015_AVT_subset.csv AIR_RELSW_1_2016_AVT_subset.csv AIR_RELSW_1_2017_AVT_subset.csv AIR_RELSW_1_2018_AVT_subset.csv AIR_RELSW_1_2019_AVT_subset.csv</i>
Brief Description of Source:	<i>Student-level scores on state standardized assessments and demographics for school years 2013/14–2018/19 using falsified, linkable student identifiers provided by NMPED. Includes all tested grades in New Mexico public schools.</i>
Unit of Analysis:	<i>Student</i>
Brief Description of Collection:	<i>State assessment tests conducted in 2013/14–2018/19</i>
If applicable, is a copy of the instrument reproduced in report? <i>If not, include as appendix to this template</i>	<i>Not applicable</i>

Source #2 Student Demographic Data

Source Name:	<i>Student Demographic Data</i>
Administrative Data Provider, if applicable, from Section 1	<i>NMPED (#2)</i>
Provider File Name(s):	<i>AIR 14-20 demo ALL student.csv</i>
Brief Description of Source:	<i>Student-level self-reported race, ethnicity, and English language proficiency status for school years 2013/14–2018/19 using falsified, linkable student identifiers provided by NMPED. Includes all students in grades K–12 in New Mexico public schools.</i>
Unit of Analysis:	<i>Student</i>

Brief Description of Collection:	<i>Student race/ethnicity and English language proficiency data collected by school districts and reported to NMPED during the 2013/14–2018/19 school years.</i>
If applicable, is a copy of the instrument reproduced in report? <i>If not, include as appendix to this template</i>	<i>Not applicable</i>

Source #3 ACCESS Assessment Data

Source Name:	<i>ACCESS Assessment Data</i>
Administrative Data Provider, if applicable, from Section 1	<i>NMPED (#3)</i>
Provider File Name(s):	<i>AIR_RELSW_1_2014_ACCESS_subset.csv AIR_RELSW_1_2015_ACCESS_subset.csv AIR_RELSW_1_2016_ACCESS_subset.csv AIR_RELSW_1_2017_ACCESS_subset.csv AIR_RELSW_1_2018_ACCESS_subset.csv AIR_RELSW_1_2019_ACCESS_subset.csv</i>
Brief Description of Source:	<i>Student-level scores on ACCESS for school years 2013/14–2018/19 using falsified, linkable student identifiers provided by NMPED. Includes all students in grades K–8 in New Mexico public schools.</i>
Unit of Analysis:	<i>Student</i>
Brief Description of Collection:	<i>State assessment tests conducted in 2013/14–2018/19</i>
If applicable, is a copy of the instrument reproduced in report? <i>If not, include as appendix to this template</i>	<i>Not applicable</i>

Source #4 DIBELS Assessment Data

Source Name:	<i>DIBELS Assessment Data</i>
Administrative Data Provider, if applicable, from Section 1	<i>NMPED (#4)</i>
Provider File Name(s):	<i>AIR_RELSW_1_2014_DIBELS_subset.csv</i>

Brief Description of Source:	<i>Student-level scores on DIBELS for school year 2013/14 using falsified, linkable student identifiers provided by NMPED. Includes all students in grades K–2 in New Mexico public schools.</i>
Unit of Analysis:	<i>Student</i>
Brief Description of Collection:	<i>DIBELS tests conducted in spring 2014</i>
If applicable, is a copy of the instrument reproduced in report? <i>If not, include as appendix to this template</i>	<i>Not applicable</i>

Source #5 District Identifier Data

Source Name:	<i>District Identifier Data</i>
Administrative Data Provider, if applicable, from Section 1	<i>NMPED (#5)</i>
Provider File Name(s):	<i>Districts_2014_2020.csv</i>
Brief Description of Source:	<i>Unique district codes for school years 2013/14–2018/19. Includes school districts in New Mexico.</i>
Unit of Analysis:	<i>School district</i>
Brief Description of Collection:	<i>Records of unique codes used to identify New Mexico school districts, maintained by NMPED.</i>
If applicable, is a copy of the instrument reproduced in report? <i>If not, include as appendix to this template</i>	<i>Not applicable.</i>

Section 3. Data File Descriptions

Files #1 through #6

File Names:	<i>AIR_RELSW_1_2014_AVT_subset.csv</i> <i>AIR_RELSW_1_2015_AVT_subset.csv</i> <i>AIR_RELSW_1_2016_AVT_subset.csv</i> <i>AIR_RELSW_1_2017_AVT_subset.csv</i> <i>AIR_RELSW_1_2018_AVT_subset.csv</i> <i>AIR_RELSW_1_2019_AVT_subset.csv</i>
Related source, from Section 2	<i>State Standardized Assessment Data</i>
Records:	<i>AIR_RELSW_1_2014_AVT_subset.csv: 296,829 records</i> <i>AIR_RELSW_1_2015_AVT_subset.csv: 330,692 records</i> <i>AIR_RELSW_1_2016_AVT_subset.csv: 361,852 records</i> <i>AIR_RELSW_1_2017_AVT_subset.csv: 372,169 records</i> <i>AIR_RELSW_1_2018_AVT_subset.csv: 370,301 records</i> <i>AIR_RELSW_1_2019_AVT_subset.csv: 331,899 records</i>
Variables:	<i>AIR_RELSW_1_2014_AVT_subset.csv: 14 variables</i> <i>AIR_RELSW_1_2015_AVT_subset.csv: 15 variables</i> <i>AIR_RELSW_1_2016_AVT_subset.csv: 15 variables</i> <i>AIR_RELSW_1_2017_AVT_subset.csv: 14 variables</i> <i>AIR_RELSW_1_2018_AVT_subset.csv: 14 variables</i> <i>AIR_RELSW_1_2019_AVT_subset.csv: 15 variables</i>
Unique identifiers:	<i>STID_PROXY, SY</i>

<i>Variable Name</i> (as it appears in the file)	<i>Variable Definition</i>	<i>Notes</i>
SY	School year	
STID_PROXY	Falsified, unique identifier for the student.	
GRADE	Grade level at the time students were assessed on the state standardized assessment.	
TESTED_GRADE	Grade level at the time students were assessed on the state standardized assessment.	This variable is available only in the following data file: <i>AIR_RELSW_1_2019_AVT_subset</i>
STARS_GRADE	Grade level the students were assessed in.	This variable is available only in the following data file: <i>AIR_RELSW_1_2019_AVT_subset</i>
SCHNUMB	Unique identifier for the school.	

<i>Variable Name</i> (as it appears in the file)	<i>Variable Definition</i>	<i>Notes</i>
TESTNAME	Name of the assessment.	
SUBTEST	The subject of the state standardized assessment.	
TESTLANG	The language of the state standardized math assessment.	
PL	Proficiency level on state summative standardized assessment.	
SS	Score on state summative standardized assessment.	
FEMALE	Whether a student is female.	This variable is available only in the following data files: AIR_RELSW_1_2014_AVT_subset AIR_RELSW_1_2015_AVT_subset AIR_RELSW_1_2016_AVT_subset
MALE	Whether a student is male.	This variable is available only in the following data files: AIR_RELSW_1_2014_AVT_subset AIR_RELSW_1_2015_AVT_subset AIR_RELSW_1_2016_AVT_subset
GENDER	Gender of a student.	This variable is available only in the following data files: AIR_RELSW_1_2017_AVT_subset AIR_RELSW_1_2018_AVT_subset AIR_RELSW_1_2019_AVT_subset
FRL	Whether a student is eligible for the National School Lunch Program.	
SWD	Whether a student is eligible for special education services.	
ELLCURRENT	Status as an English learner student.	This variable is available only in the following data files: AIR_RELSW_1_2014_AVT_subset AIR_RELSW_1_2016_AVT_subset AIR_RELSW_1_2019_AVT_subset
ELL	Status as an English learner student.	This variable is available only in the following data files: AIR_RELSW_1_2015_AVT_subset AIR_RELSW_1_2017_AVT_subset AIR_RELSW_1_2018_AVT_subset

File #7

File Name:	<i>AIR 14-20 demo ALL student.csv</i>
Related source, from Section 2	<i>Student Demographic Data</i>
Records:	<i>1,802,820 records</i>
Variables:	<i>5 variables</i>
Unique identifiers:	<i>SCHOOL_YEAR, StudentCID</i>

<i>Variable Name (as it appears in the file)</i>	<i>Variable Definition</i>
SCHOOL_YEAR	Date of the student demographics in YYYY-MM-DD format (for example, 2013/14 stored as 2014-06-30)
StudentCID	Falsified, unique identifier for the student.
ETHNIC_CODE	Student self-report of racial/ethnic background. Coded as American Indian, Asian, Black, Hawaiian Native/Pacific Islander, or White.
HISPANIC_IND	Student self-report indicating whether a student is of Hispanic ethnicity.
ENG_PROFICIENCY	Status as an English learner student.

Files #8 through #13

File Names:	<i>AIR_RELSW_1_2014_ACCESS_subset.csv AIR_RELSW_1_2015_ACCESS_subset.csv AIR_RELSW_1_2016_ACCESS_subset.csv AIR_RELSW_1_2017_ACCESS_subset.csv AIR_RELSW_1_2018_ACCESS_subset.csv AIR_RELSW_1_2019_ACCESS_subset.csv</i>
Related source, from Section 2	<i>ACCESS Assessment Data</i>
Records:	<i>AIR_RELSW_1_2014_ACCESS_subset.csv: 38,451 records AIR_RELSW_1_2015_ACCESS_subset.csv: 36,591 records AIR_RELSW_1_2016_ACCESS_subset: 34,967 records AIR_RELSW_1_2017_ACCESS_subset: 31,824 records AIR_RELSW_1_2018_ACCESS_subset: 35,477 records AIR_RELSW_1_2019_ACCESS_subset: 36,497 records</i>
Variables:	<i>15 variables</i>
Unique identifiers:	<i>STID_PROXY, SY</i>

<i>Variable Name</i> (as it appears in the file)	<i>Variable Definition</i>
STID_PROXY	Falsified, unique identifier for the student.
SY	School year
GRADE	Grade level at the time students were assessed on the ACCESS for ELLs (ACCESS) assessment.
TESTCODE	Code of the assessment.
TESTNAME	Name of the assessment.
LISTENINGSCALESCORE	Proficiency level score in the English language listening domain on the ACCESS.
SPEAKINGSCALESCORE	Proficiency level score in the English language speaking domain on the ACCESS.
READINGSCALESCORE	Proficiency level score in the English language reading domain on the ACCESS.
WRITINGSCALESCORE	Proficiency level score in the English language writing domain on the ACCESS.
COMPOSITEOVERALLSCALESCORE	Overall proficiency level score on the ACCESS.
LISTENINGPROFICIENCYLEVEL	Proficiency level in the English language listening domain on the ACCESS.
SPEAKINGPROFICIENCYLEVEL	Proficiency level in the English language speaking domain on the ACCESS.
READINGPROFICIENCYLEVEL	Proficiency level in the English language reading domain on the ACCESS.
WRITINGPROFICIENCYLEVEL	Proficiency level in the English language writing domain on the ACCESS.
COMPOSITEOVERALLPROFICIENCYLEVEL	Overall proficiency level on the ACCESS.
DATEOFTESTING	Date the student was assessed

File #14

File Name:	<i>AIR_RELSW_1_2014_DIBELS_subset.csv</i>
Related source, from Section 2	<i>DIBELS Assessment Data</i>
Records:	<i>30,714 records</i>
Variables:	<i>6 variables</i>
Unique identifiers:	<i>STID_PROXY, SY</i>

<i>Variable Name</i> (as it appears in the file)	<i>Variable Definition</i>
SY	School year of district information in YYYY-YYYY
STID_PROXY	Falsified, unique identifier for the student.
SCHNUMB	Unique identifier for the school.
GRADE	Grade level at the time students were assessed on the DIBELS assessment.
SS	Score on DIBELS assessment.
TESTNAME	Name of the assessment.

File #15

File Name:	<i>Districts_2014_2020.csv</i>
Related source, from Section 2	<i>District Identifier Data</i>
Records:	<i>1,064 records</i>
Variables:	<i>3 variables</i>
Unique identifiers:	<i>STID_PROXY, SY</i>

<i>Variable Name</i> (as it appears in the file)	<i>Variable Definition</i>
SY	School year of district information in YYYY-YYYY
DistCode	Unique identifier for the district.
DistName	District name

Section 4. Statistical Code or Pseudo-Code

This section provides brief descriptions of how the data obtained were transformed into the final analytic file. The study team conducted the following steps to prepare data for analysis.

I. Reading in and converting files

- Statistical code: 1_read_all_raw_files_public_data_file.do
- The study team read all 15 comma-separated values files from NMPED into Stata and converted the files to Stata_dta format.

II. Cleaning files

The study team cleaned the data files in multiple stages across the various data sources.

- **State Standardized Assessment and DIBELS Data**
 - Statistical code: 2a_process_avt_file_public_data_file.do and 3a_clean_achievement_public_data_file.do
 - Combine all files across years.
 - Recode string demographic variables to numeric. For example, the variable SWD (whether a student is eligible for special education services) was coded to Y = 1 and N = 0.
 - Consolidate the three gender variables, FEMALE and MALE in 2013/14 through 2015/16, and GENDER in 2016/17 through 2018/19.
 - Add leading zeros to school identification codes that are less than six digits.
 - Exclude records for students tested in the fall of the school year.
 - Exclude records with conflicting or duplicate testing information across multiple records.
 - Standardize scale scores within each subject, test code, and school year.
- **Student Demographic Data**
 - Statistical code: 2b_process_demo_file_public_data_file.do
 - Rename the unique identifier for the student and school years variables to match the variables names in the state standardized assessment data files.
 - Make codes for the variable eng_proficiency consistent across years.
 - Recode string race variables to numeric variables. For example, the variable ethnic_code (student self-report of racial/ethnic background) was coded to race_as = 1 if ethnic_code = A.
 - Exclude duplicate student records each year.
 - Exclude students who are in pre-Kindergarten that appear in the 2019 data file.
- **ACCESS Assessment Data**
 - Statistical code: 2c_process_access_files_public_data_file.do
 - Combine all files across years.
 - Exclude records with missing unique student identifiers.
 - Exclude the lowest ACCESS score(s) if a student has more than one ACCESS score in a year.

- **District Identifier Data**
 - Statistical code: 2d_process_districts_public_data_file.do
 - Keep one record per district because the records were consistent across years.

III. Merging and creating analysis samples

- **Merging all cleaned data files**
 - Statistical code: 4_merge_all_files_data_file.do
 - Merge files across all data sources.
 - Create school-wide-by-year demographic variables using the full sample of students in each school. The study team created the following school-wide variables: total school enrollment, the percentage of students in each racial/ethnic group, the percentage of students eligible for the National School Lunch Program, the percentage of students eligible for special education services, and the percentage of students in each English language proficiency category.
- **Creating the sample for research questions 2–4**
 - Statistical code: 5a_create_primary_sample_data_file.do
 - Filter the data file to students that met the study criteria:
 - In the baseline-year: students in grades 2–7 and identified as an English learner student.
 - In the outcome-year: students in grades 3–8, assessed in English language arts or math, identified as an English learner student and took the ACCESS, or identified as fluent English proficient and reclassified and did not take the ACCESS.
- **Creating the sample for research question 1**
 - Statistical code: 5b_create_state_sample_data_file.do
 - Filter the data file to students that met the study criteria: Students in grades 3–8 in 2014/15–2018/19 who were identified as newly reclassified and did not take the ACCESS or who were identified as English learner students and took the ACCESS and assessed in English language arts or math.

IV. Analysis – Main report and appendices

- **Preparing the data for research questions 2–4 analyses**
 - Statistical code: 6a_analysis_setup_public_data_file.do
 - Filter the data file to students that had a baseline ACCESS score that fell within the optimal bandwidth.
 - Generate running variable and interactions for baseline ACCESS scale scores.
 - Create quadratic form of the running variables.
- **Conducting supplemental analyses for research questions 2–4 analyses**
 - Statistical code: 6b_figures_public_data_file.do
 - Conduct supplemental analyses and create figures for reporting appendix C.
- **Conducting primary analyses for research questions 2–4**
 - Statistical code: 6c_analysis_public_data_file.do
 - Conduct analyses for research questions 2–4 and create figures for reporting in the main section of the report and in appendix B.

- **Create appendix A tables**
 - Statistical code: 99_create_appendix_a_tables_public_data_file.do
 - Create a table summarizing missing data and tables summarizing the characteristics of students in the sample.

- **Conducting analyses**
 - Statistical code: 99_examine_attrition_public_data_file.do
 - Create a table summarizing the overall and differential attrition.