California Department of Education

Executive Office

SBE-002 (REV. 11/2017)

memo-imb-amard-feb21item02

# MEMORANDUM

**DATE:** February 17, 2021

**TO:** MEMBERS, State Board of Education

**FROM:** Tony Thurmond, State Superintendent of Public Instruction

**SUBJECT:** Update on the Development of a Student Growth Model for the Integrated Local, State, and Federal Accountability and Continuous Improvement System

## Summary of Key Issues

Since 2017, the California Department of Education (CDE), the Educational Testing Services (ETS), and the State Board of Education (SBE) have been engaged in developing a Student Growth Model for California’s schools and local educational agencies (LEAs). This work has been shared and refined based on feedback from stakeholders, including educators, advocacy groups, and the general public. After exploring several different models over a multi-year period, the SBE directed the CDE to pursue the “residual gain” (RG) method.

For reporting purposes, RG scores must be aggregated to the LEA, school, and student group levels to summarize the progress of student achievement. A standard approach for computing aggregates of student-level growth measures (AGMs) is to report the simple average of all the individual student RG growth scores available for a school or LEA in a particular school year, for a particular subject. However, in preliminary investigations, ETS and the CDE found that the simple-average AGMs tended to have large year-to-year variation for the same school or LEA, creating concerns about the credibility of these AGMs if adopted for annual reporting and accountability use. In order to improve accuracy and year-to-year stability, ETS developed the empirical best linear prediction (EBLP) (Lockwood, Castellano, & McCaffrey, 2020).

At the request of CDE, ETS conducted a study to explore the potential for using the EBLP method to estimate AGMs for California LEAs and schools. This report, which appears in Attachment 1, provides the results of that investigation. It compares the EBLP weighted average approach to the simple average approach in terms of accuracy and stability at the school and LEA levels for all students and several student groups within schools and LEAs. Given distinct results at the LEA level in terms of improved accuracy, further explorations were conducted in support of a hybrid approach for calculating AGM for LEAs—assigning the EBLP weighted average in some cases and the simple average in others. The report also provides results of the hybrid approach and concludes the investigative study of the student-level growth model. The attached report provides the SBE with the necessary information to consider the adoption of the model at their March 2021 meeting. The March 2021 item will also contain the 2021 accountability work plan, including information on how to effectively communicate the growth model results.

## Summary of Previous State Board of Education Discussion and Action

In a June 2016 Information Memorandum, the CDE provided a progress update and clarified key issues related to the design of a school- and district-level accountability model, as opposed to reporting individual student-level growth and performance (<https://www.cde.ca.gov/be/pn/im/documents/memo-dsib-amard-jun16item01.doc>).

In February 2016, the SBE received an Information Memorandum that provided an overview of student-level growth models that can be used to communicate Smarter Balanced Summative Assessment results (<https://www.cde.ca.gov/be/pn/im/documents/memo-dsib-amard-feb16item01.doc>).

In January 2017, the SBE discussed criteria for selecting a growth model used for school and district accountability (<https://www.cde.ca.gov/be/ag/ag/yr17/documents/jan17item02.doc>).

Following the SBE discussion in January 2017, the CDE further consulted with ETS, the Technical Design Group (TDG), the CAASPP Technical Advisory Group (TAG), and the Statewide Assessment Stakeholder Group, regarding potential growth models. Three models were selected for simulation. The discussion and recommendations of the groups were summarized and presented to the SBE in a June 2017 Information Memorandum (<https://www.cde.ca.gov/be/pn/im/documents/memo-asb-adad-jun17item03.doc>).

In February 2018, the SBE received an Information Memorandum with the results of the ETS Growth Study, which provided a statistical analysis of three proposed growth models (<https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-amard-feb18item01.docx>).

In May 2018, the SBE reviewed analyses of the three student-level growth models conducted by ETS and directed the CDE to further explore the Residual Gain model for possible inclusion in the Dashboard (<https://www.cde.ca.gov/be/ag/ag/yr18/documents/may18item02.docx>).

At its July 2018 meeting, the SBE directed the CDE to conduct further analyses on the Residual Growth model, including the impact of future years of assessment data, changes in the model to reduce year-to-year volatility, consideration of additional growth models or options, and an examination of growth models implemented in other states (<https://www.cde.ca.gov/be/ag/ag/yr18/documents/jul18item01.docx>).

The CDE engaged the California Comprehensive Center to conduct this research and facilitate a stakeholder process on the future direction of this work. In February 2019, the SBE received an Information Memorandum, providing a summary of the first Student Growth Model stakeholder meeting (<https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-amard-feb19item03.docx>).

In April 2019, the SBE received an Information Memorandum, providing a summary of the second growth model stakeholder feedback group meeting (<https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-amard-apr19item02.docx>).

In November 2019, the SBE received an Information Memorandum, providing a summary of the growth model stakeholder feedback group process (<https://www.cde.ca.gov/be/pn/im/documents/nov19memoamard01.docx>).

At the March 2020 meeting, the SBE directed the CDE to provide a presentation at the May 2020 meeting regarding the work conducted to date on the development of a student-level growth model. Due to the national health crisis, this presentation was postponed until the July 2020 SBE meeting (<https://www.cde.ca.gov/be/ag/ag/yr20/documents/mar20item05.docx>).

In June 2020, the SBE received an Information Memorandum, providing the history and background on the student growth model work to date (<https://www.cde.ca.gov/be/pn/im/documents/memo-imb-amard-june20item01.docx>).

At the July 2020 SBE meeting, the CDE provided a presentation regarding the work conducted to date on the development of a student-level growth model (<https://www.cde.ca.gov/be/ag/ag/yr20/documents/jul20item02.docx>).

In September 2020, the CDE presented an update on the progress on refining the statistical methodology used to develop a Student Growth Model. In addition, the ETS presented the results of its study on the potential of the EBLP method to estimate aggregate growth measures for LEAs and schools (<https://www.cde.ca.gov/be/ag/ag/yr20/documents/sept20item01.docx>).

In November 2020, the CDE presented an item recommending that that the SBE adopt a single subject EBLP methodology to improve growth model communication (<https://www.cde.ca.gov/be/ag/ag/yr20/documents/nov20item06.docx>).

**Attachment(s)**

* Attachment 1: An Investigation of the Use of Empirical Best Linear Prediction for Aggregate Growth Measures (64 pages)

# Attachment 1

## An Investigation of the Use of Empirical Best Linear Prediction for Aggregate Growth Measures

This attachment was prepared by the Educational Testing Services (ETS) on behalf of the California Department of Education (CDE) for the State Board of Education (SBE).

**An Investigation of the Use of Empirical Best Linear Prediction for Aggregate Growth Measures**

**Contract #CN150012**

**Prepared for the California Department of Education by Educational Testing Service**

**Presented January 22, 2021**



**Table of Contents**

[**1. Executive Summary 4**](#_Toc62206836)

[**2. Motivating Problem 4**](#_Toc62206837)

[**3. EBLP as a Potential Solution 6**](#_Toc62206838)

[**4. Empirical Study Setup 6**](#_Toc62206839)

[4.A. Business Rules Adopted for AGM Reporting and EBLP Application 7](#_Toc62206840)

[4.B. Data 7](#_Toc62206841)

[**5. Results at the School Level 9**](#_Toc62206842)

[5.A. Accuracy Results 9](#_Toc62206843)

[5.B. Stability Results 1](#_Toc62206844)3

[**6. Results at the LEA Level 1**](#_Toc62206845)**6**

[6.A. Accuracy Results 1](#_Toc62206846)6

[6.B. Stability Results 2](#_Toc62206847)0

[6.C. Exploration of Reporting Options for LEA Student Groups 2](#_Toc62206848)3

[**7. Conclusions 3**](#_Toc62206849)**1**

[7.A. School-Level Data 3](#_Toc62206850)1

[7.B. LEA-Level Data 3](#_Toc62206851)1

[**8. Reference**](#_Toc62206852) **32**

[**Appendix A: Memorandum for the Technical Design Group December 2020**](#_Toc62206853) **33**

[Evaluating Aggregate Growth Reporting Options for Local Educational Agencies 33](#evaluating)

[Study Design 41](#study)

[Appendix A1: School-Level Stability and Accuracy Results 53](#Appen)

[Appendix A2: LEA-Level Stability and Accuracy Results 59](#APPN2)

**List of Tables**

[Table 1. Student Group Abbreviations 8](#_Toc62207138)

[Table 2. Improvement in Accuracy of Growth Estimates Using Two-Year EBLP Weighted Averages Versus Simple Averages at the School Level 11](#_Toc62207139)

[Table 3. Cross-Year Stability of the 2018–19 Two-Year EBLP Weighted Average and the Simple Average for Schools 15](#_Toc62207140)

[Table 4. Improvement in Accuracy of Growth Estimates Using Two-Year EBLP Weighted Averages Versus Simple Averages at the LEA Level 19](#_Toc62207141)

[Table 5. Cross-Year Stability of the 2018–19 Two-Year EBLP Weighted Average and the Simple Average for LEAs 23](#_Toc62207142)

[Table 6. Improvement in Accuracy of Growth Estimates Using the Hybrid Approach Versus Simple Averages for Student Groups Within LEAs 30](#_Toc62207143)

[Table 7. Cross-Year Stability of the 2018–19 Hybrid Approach and the Simple Average for Student Groups Within LEAs 34](#_Toc62207144)

1. **Executive Summary**

This report presents the findings of a study conducted by ETS on the use of the empirical best linear prediction (EBLP), which builds on the student growth model currently being developed by the California Department of Education (CDE).

Briefly, EBLP is a statistical approach developed by ETS to improve the accuracy and stability of student growth measures aggregated at the local educational agency (LEA), school, and student-group levels. Aggregate growth measures (AGMs) are typically calculated as a *simple average* of all individual student growth scores. However, the CDE and ETS have found that this method leads to significant variation in the results from one year to the next, especially for smaller populations. In contrast, EBLP uses a *weighted*-average approach. (A thorough explanation of these models is presented in the sections that follow.)

ETS has compared the EBLP weighted-average approach to the simple-average approach in terms of accuracy and stability at both the school and LEA levels. Given distinct results at the LEA level (given in section 6), further explorations were conducted in support of a hybrid approach for calculating AGM for LEAs—assigning the EBLP weighted average in some cases and the simple average in others. Based on ETS’ findings, which are presented in depth in this report, the Technical Design Group (TDG) recommended that a hybrid approach be applied to the student growth model as follows:

* The weighted average (i.e., the EBLP model) be applied at the school level, inclusive of all student groups at the school level; at the overall LEA level; and, at the LEA level for student groups with 500 or fewer growth scores.
* At the LEA level only, the simple average be applied for student groups when there are more than 500 growth scores in a student group.

1. **Motivating Problem**

The CDE’s student growth model is based on residual gains (RGs). The CDE defines an RG score as the difference between

1. a student’s current-year California Assessment of Student Performance and Progress (CAASPP) score in a particular content area (i.e., English language arts/literacy [ELA] or mathematics), and
2. a linear prediction of that test score, which is based on
   1. the student’s CAASPP scores in *both* content areas (ELA *and* mathematics) in the immediate prior-year CAASPP, and
   2. the same sets of scores (i.e. current and prior years) for all other students in the same grade.

Students who score higher than expected in the current year (i.e., higher than the linear prediction based on the students’ previous test scores) receive positive RG scores, and students who score lower than expected receive negative RG scores. Students who perform as expected receive an RG score of zero.

For example, a student who scored 2300 on the grade three CAASPP in 2018 may have an expected grade four score of 2450, based on the typical student growth of California’s fourth graders in 2019. If that student received a score of only 2400, that student has an individual residual gain (IRG) score of -50, which represents the difference between the student’s actual and expected score (2400 - 2450).

Once an IRG score is calculated for each student in the school or LEA, an aggregate RG score is calculated. A standard approach for calculating AGMs is a simple average of all the individual student growth scores available for a school or LEA in a particular school year for a particular subject. For example, if a school had 100 students with mathematics growth scores in spring 2019, the simplest method to compute the mathematics AGM for the school in 2019 is just the average, or mean, of the mathematics growth scores for these 100 students. While this method is straightforward, it has a downside. Simple averages that are based on few scores are less accurate than those based on many scores.

To help convey what is meant by the “accuracy” of a simple average, consider the example of two California schools, School A and School B, that received an average mathematics growth score of +22 points in 2018–19. School A is a large school, with 1,243 students, while School B is a small school, with only 15 students. Although both schools received the same average growth score, that average is based on vastly different numbers of students.

Consider next what would happen to the average if just one student growth score was removed from each school. For School A, with over 1200 growth scores, there would be little impact to the average: the possible values of the AGM for this school range from 21.9 to 22.3, both of which are very close to the original average of 22. However, for School B, which has only 15 growth scores, the removal of just one growth score could have a significant impact, dropping the average down to as low as 11.1 or raising it to as high as 28.2, depending on which specific score was removed. Thus, the average growth scores for smaller schools are much more sensitive to the specific set of scores included in the average than are the averages for larger schools.

One of the most problematic side effects of low accuracy is that it can cause excessive year-to-year fluctuations in the AGMs for the same school or LEA. For example, for a small school, AGMs might bounce around a lot from year to year. In preliminary investigations of simple averages, ETS and the CDE found that the AGMs tended to have large year-to-year variation for the same school or LEA, creating concerns about the credibility of these AGMs if adopted for annual reporting and accountability use. (Refer to the June 2018 ETS report at <https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-amard-jun18item01a1.docx>.)

A related problem is that because accuracy is driven by the number of growth scores included in the AGMs, the measures behave differently across schools or LEAs of different sizes. Consider again Schools A and B: The AGMs for School B, which has few student growth scores each year, are almost certainly going to fluctuate much more than those in School A. These kinds of disparities can erode credibility and a perception of fairness in the reporting system.

The problems caused by low accuracy can be even more pronounced for the AGMs of specific student groups within schools and LEAs, such as students with disabilities or foster youth, because there may be many schools and LEAs serving relatively small numbers of students in particular groups. Thus, for relatively rare student groups, low accuracy can be severe and widespread.

1. **EBLP as a Potential Solution**

To improve the accuracy of the student growth model and address the year-to-year volatility in RG scores, ETS developed the ELPB statistical approach, which is applied to schools, LEAs, and student groups (Lockwood, Castellano, & McCaffrey, 2020). While EBLP is not a growth model proper, it uses the individual growth scores generated from the RG model to create a more accurate aggregate measure of those scores than one obtained from the simple average of the individual RG scores.

EBLP uses a *weighted* average of student growth measures from multiple years, giving greater weight to data from the *most recent year* and less weight to data from previous school years. The weights are specific to each school and LEA and are dependent on the numbers of growth measures for that school or LEA.

For instance, when calculating aggregate mathematics growth in the reporting year, if a school or LEA has mathematics growth measures from many students in the reporting year, the EBLP weighted average puts most of its weight on the average of those growth measures and little weight on growth measures from prior school years. And, so, a moderate-sized school with an RG of -8 in 2017–18 and an RG of +5 in 2018–19 could, depending on its number of current-year growth scores, receive an EBLP weighted average of +4 since the most recent year of student growth data would receive the vast majority of the weight.

As the number of students with mathematics growth measures in the reporting year decreases, the EBLP weighted average shifts more weight to the growth measures from prior school years. In this case, an RG of +4 in 2017–18 and an RG of +29 in 2018–19 could result in an EBLP of +24 since the prior school year’s student growth data would bear greater weight than in the previous example but still less than the weight of the current year.

As shown in the data presented in the next sections, the EBLP model is found to have the greatest impact on small groups and almost no impact on large groups. This is to be expected because simple averages are much more accurate for large groups.

1. **Empirical Study Setup**

To understand the performance of the EBLP method for California data, ETS conducted an empirical study using longitudinal test-score data from California. This report provides the results of that investigation. It compares the EBLP weighted-average approach to the simple-average approach in terms of accuracy and stability at the school and LEA levels for all students and several student groups within schools and LEAs.

The general EBLP method can combine growth data across tested subjects and across any number of years. In previous investigations of the EBLP method in California, ETS and the CDE considered applying the method using two-year or three-year data blocks (i.e., using either one or two years of data prior to the reporting year), and simultaneously modeling growth data from mathematics and ELA. However, on the basis of empirical findings, stakeholder input, and concerns about the simplicity of future communications of the method, the EBLP will use data from two-year blocks only and be applied separately for each content area. The complete set of business rules in the reporting of AGMs and application of the EBLP is presented in the next section.

**4.A. Business Rules Adopted for AGM Reporting and EBLP Application**

The following four business rules have been adopted regarding the reporting of AGMs and the application of EBLP in California:

1. AGMs will be reported only when they are based on 11 or more student growth scores in the reporting year and for which there is at least one growth score from the prior year.
2. The EBLP method will use data from two-year blocks. Thus, when computing AGMs in a given reporting year, the EBLP will use growth scores from that year as well as from the immediate prior year.
3. The EBLP method will be applied separately to mathematics and ELA growth scores, as decided by the California State Board of Education (SBE) in the November 2020 meeting. Thus, when computing mathematics AGMs in a given reporting year, the EBLP will combine mathematics growth scores from that year, and mathematics growth scores from the immediate prior year. (Note that the residual gain growth scores for individual students will still use prior scores from both subjects in the prediction model.)
4. When only a single year of growth data is available (i.e., no prior-year growth score), the simple average will be applied. For example, if a given LEA has 15 English learner students with ELA growth scores in the reporting year and no such students in the prior year, that LEA’s AGM for the ELA growth of its English learners will be the simple average of the 15 ELA growth scores in the reporting year.

These business rules were established based on empirical findings, stakeholder input, and concerns about the simplicity of future communications of the method. All reported results are based on an application of the EBLP method, which adheres to these business rules.

**4.B. Data**

Three years of data are used in this analysis. They include the mathematics and ELA growth scores from:

* the 2018–19 school year,
* the 2017–18 school year, and
* the 2016–17 school year.

The 2017–18 and 2018–19 data generate EBLPs for 2018–19, while the 2016–17 and 2017–18 data generate EBLPs for 2017–18.

These data are used to compute AGMs at aggregation levels defined by combinations of the following three factors:

1. Organizational level: school versus LEA
2. Subject: mathematics versus ELA
3. Group: 17 student groups

The 17 student groups include a group for all students as well as the 16 groups used in CDE ELA academic indicator reporting, described at <https://www.cde.ca.gov/ta/ac/cm/ela19.asp>.

Student group abbreviations are presented in [Table1](#Table1).

**Table 1. Student Group Abbreviations**

| Student Group  Abbreviation | Student Group Name |
| --- | --- |
| ALL | All Students |
| AA | Black/African American |
| AI | American Indian or Alaska Native |
| AS | Asian |
| FI | Filipino |
| HI | Hispanic |
| PI | Pacific Islander |
| WH | White |
| MR | Multiple Races/Two or More |
| EL | English Learner |
| ELO | English Learners Only |
| RFP | Reclassified fluent English proficient only |
| EO | English Only |
| SED | Socioeconomically Disadvantaged |
| SWD | Students with Disabilities |
| FOS | Foster Youth |
| HOM | Homeless Youth |

The input data used to compute 2018–19 EBLPs for a given organizational level, subject, and student group include all student growth scores in the given subject, for students who are members of the given student group and assigned to the appropriate schools or LEAs, for the 2017–18 and 2018–19 years.

1. **Results at the School Level**

The results for evaluating the accuracy and cross-year stability of the EBLP weighted averages and simple averages at the school level are presented in this section.

**5.A. Accuracy Results**

A key motivation for using the EBLP weighted average over the simple average is to improve the accuracy of reported growth measures. [Table2](#Table2) presents the improved accuracy through a ratio (i.e., mean accuracy ratio for two-year EBLP versus simple average) at the school level, for ALL students and for individual student groups:

* An accuracy ratio greater than 1 indicates that the EBLP weighted average is estimated to be more accurate than the simple average.
* An accuracy ratio less than 1 indicates that the EBLP weighted average is estimated to be less accurate than the simple average.
* An accuracy ratio equal to 1 indicates that the EBLP weighted average and the simple average are estimated to be equally accurate.

For example, an accuracy ratio of 1.50 for the EBLP weighted average means that it is estimated to be 1.5 times as accurate as the simple average. In other words, its accuracy is approximately the same as what the accuracy of the simple average would have been if it had been based on 50 percent more students.

The accuracy ratio was calculated for each school for both mathematics and ELA for the 17 student groups. The percentage of schools estimated to have EBLP weighted averages as accurate or more accurate than simple averages is also reported.

The results are summarized by content area and student group. For each student group, results are broken down by the following size intervals:

* Schools with 11 to 29 students (the range for which accountability statistics are reported but no color-coded classification is provided in the California School Dashboard)
* Schools with 30 to 149 students
* Schools with 150 or more students

In cases where there are fewer than 100 schools within one of the designated size intervals, the results will be reported using a combined interval (for example, 11–29 and ≥30).

Please note that accountability measures are not calculated for groups of 10 or fewer students. Therefore, no results will be reported when 10 or fewer students have growth scores in the reporting year.

**Table 2. Improvement in Accuracy of Growth Estimates Using Two-Year EBLP Weighted Averages Versus Simple Averages at the School Level**

| **Student Group** | **Subject** | **2018–19 School Size1** | **Number of Schools** | **Mean Accuracy Ratio** | **Percentage of Schools with Improved Accuracy2** |
| --- | --- | --- | --- | --- | --- |
| ALL | ELA | 11–29 | 186 | 1.56 | 100% |
| ALL | ELA | 30–149 | 2,172 | 1.18 | 100% |
| ALL | ELA | ≥150 | 5,019 | 1.08 | 100% |
| ALL | Mathematics | 11–29 | 179 | 1.49 | 100% |
| ALL | Mathematics | 30–149 | 2,176 | 1.14 | 100% |
| ALL | Mathematics | ≥150 | 5,016 | 1.06 | 100% |
| AA | ELA | 11–29 | 1,421 | 1.67 | 100% |
| AA | ELA | ≥30 | 967 | 1.34 | 99% |
| AA | Mathematics | 11–29 | 1,417 | 1.72 | 100% |
| AA | Mathematics | ≥30 | 967 | 1.41 | 99% |
| AI | ELA | ≥11 | 113 | 1.66 | 99% |
| AI | Mathematics | ≥11 | 113 | 1.84 | 98% |
| AS | ELA | 11–29 | 1,300 | 1.61 | 100% |
| AS | ELA | 30–149 | 1,219 | 1.28 | 100% |
| AS | ELA | ≥150 | 279 | 1.05 | 92% |
| AS | Mathematics | 11–29 | 1,296 | 1.56 | 100% |
| AS | Mathematics | 30–149 | 1,220 | 1.27 | 100% |
| AS | Mathematics | ≥150 | 278 | 1.07 | 95% |
| FI | ELA | 11–29 | 816 | 1.52 | 100% |
| FI | ELA | ≥30 | 372 | 1.28 | 100% |
| FI | Mathematics | 11–29 | 817 | 1.51 | 100% |
| FI | Mathematics | ≥30 | 372 | 1.26 | 100% |
| HI | ELA | 11–29 | 847 | 1.61 | 100% |
| HI | ELA | 30–149 | 3,478 | 1.26 | 100% |
| HI | ELA | ≥150 | 2,645 | 1.09 | 100% |
| HI | Mathematics | 11–29 | 848 | 1.53 | 100% |
| HI | Mathematics | 30–149 | 3,478 | 1.21 | 100% |
| HI | Mathematics | ≥150 | 2,644 | 1.08 | 100% |
| PI | ELA | 11–149 | 114 | 1.56 | 100% |
| PI | Mathematics | 11–149 | 113 | 1.90 | 100% |
| WH | ELA | 11–29 | 1,395 | 1.55 | 100% |
| WH | ELA | 30–149 | 2,778 | 1.24 | 100% |
| WH | ELA | ≥150 | 782 | 1.07 | 100% |
| WH | Mathematics | 11–29 | 1,398 | 1.43 | 100% |
| WH | Mathematics | 30–149 | 2,776 | 1.17 | 100% |
| WH | Mathematics | ≥150 | 781 | 1.06 | 100% |
| MR | ELA | 11–29 | 1,600 | 1.59 | 100% |
| MR | ELA | ≥30 | 617 | 1.32 | 100% |
| MR | Mathematics | 11–29 | 1,598 | 1.47 | 100% |
| MR | Mathematics | ≥30 | 615 | 1.26 | 100% |
| EL | ELA | 11–29 | 1,194 | 1.63 | 100% |
| EL | ELA | 30–149 | 3,937 | 1.29 | 100% |
| EL | ELA | ≥150 | 1,357 | 1.11 | 100% |
| EL | Mathematics | 11–29 | 1,189 | 1.57 | 100% |
| EL | Mathematics | 30–149 | 3,940 | 1.25 | 100% |
| EL | Mathematics | ≥150 | 1,355 | 1.09 | 100% |
| ELO | ELA | 11–29 | 1,867 | 1.59 | 100% |
| ELO | ELA | 30–149 | 3,322 | 1.31 | 100% |
| ELO | ELA | ≥150 | 326 | 1.12 | 100% |
| ELO | Mathematics | 11–29 | 1,869 | 1.56 | 100% |
| ELO | Mathematics | 30–149 | 3,321 | 1.29 | 100% |
| ELO | Mathematics | ≥150 | 323 | 1.12 | 100% |
| RFP | ELA | 11–29 | 2,024 | 1.57 | 100% |
| RFP | ELA | 30–149 | 3,260 | 1.30 | 100% |
| RFP | ELA | ≥150 | 447 | 1.11 | 100% |
| RFP | Mathematics | 11–29 | 2,024 | 1.50 | 100% |
| RFP | Mathematics | 30–149 | 3,262 | 1.28 | 100% |
| RFP | Mathematics | ≥150 | 445 | 1.12 | 100% |
| EO | ELA | 11–29 | 571 | 1.54 | 100% |
| EO | ELA | 30–149 | 4,182 | 1.23 | 100% |
| EO | ELA | ≥150 | 2,535 | 1.08 | 100% |
| EO | Mathematics | 11–29 | 568 | 1.45 | 100% |
| EO | Mathematics | 30–149 | 4,190 | 1.17 | 100% |
| EO | Mathematics | ≥150 | 2,526 | 1.06 | 100% |
| SED | ELA | 11–29 | 612 | 1.54 | 100% |
| SED | ELA | 30–149 | 3,455 | 1.22 | 100% |
| SED | ELA | ≥150 | 3,085 | 1.08 | 100% |
| SED | Mathematics | 11–29 | 612 | 1.46 | 100% |
| SED | Mathematics | 30–149 | 3,460 | 1.17 | 100% |
| SED | Mathematics | ≥150 | 3,078 | 1.07 | 100% |
| SWD | ELA | 11–29 | 3,436 | 1.49 | 100% |
| SWD | ELA | ≥30 | 3,041 | 1.28 | 100% |
| SWD | Mathematics | 11–29 | 3,440 | 1.41 | 100% |
| SWD | Mathematics | ≥30 | 3,034 | 1.22 | 100% |
| FOS | ELA | 11–29 | 49 | 1.46 | 100% |
| FOS | Mathematics | 11–29 | 48 | 1.48 | 100% |
| HOM | ELA | 11–29 | 1,093 | 1.47 | 100% |
| HOM | ELA | ≥30 | 632 | 1.24 | 100% |
| HOM | Mathematics | 11–29 | 1,095 | 1.48 | 100% |
| HOM | Mathematics | ≥30 | 629 | 1.24 | 100% |

1. The size intervals are based on the number of students within schools with growth scores in grade levels four through eight in 2018–19 for the student group and subject of interest (indicated in the first two columns).
2. The percentage of schools with improved accuracy for the two-year EBLP versus the simple average represents the percentage of schools whose estimated accuracy for the two-year EBLP is as good as or better than that of the simple average.

**5.A.1 Summary of Table 2 Results**

The results presented in [Table 2](#Table2) strongly support the use of EBLP at the school level. As shown in the last column, EBLP is estimated to improve accuracy for virtually all AGMs that would be reported at the school level. The gains in estimated accuracy are particularly large for small schools, as well as for student groups with few students in the schools. For example, the mean accuracy ratio for the PI student group, which has between 11 and 149 students in California schools, was 1.56 for ELA (i.e., the EBLP weighted average was 1.56 times as accurate as the simple average) and 1.9 for mathematics.

When more students are available for calculating the aggregate growth, improvements for the EBLP weighted averages are smaller (i.e., the mean accuracy ratios near 1). This is to be expected because simple averages are much more accurate for large groups.

Across all 137,055 AGMs represented in [Table 2](#Table2), the mean estimated accuracy ratio is 1.28 and the median is 1.23. (Although not shown in the table, there are 69 cases in which the estimated accuracy ratio is less than one, and here the correlation between the EBLP and simple average AGMs is 0.96.) These findings also hold for the 2016–17 and 2017–18 data used to compute EBLPs for 2017–18, thus supporting the conclusion that using EBLP for all school-level AGMs is likely to improve accuracy for almost all cases in each reporting year.

**5.B. Stability Results**

A second motivation for using the EBLP weighted averages is to improve the cross-year stability of school or LEA growth measures. Stability can be assessed by looking at the correlation (or similarity) of AGMs between 2017–18 and 2018–19. Correlation takes on values that range from -1 to 1. Higher positive values indicate that the measures are more similar or, in this case, that growth measures are more stable across time.

Table 3 compares the cross-year stability of the two models, based on 2017–18 and 2018–19 school-level data.

**Table 3. Cross-Year Stability of the 2018–19 Two-Year EBLP Weighted Average and the Simple Average for Schools**

| **Student Group** | **Subject** | **2018–19 School Size1** | **Number of Schools** | **Correlation Between the 2017–18 and 2018–19 Simple Average** | **Correlation Between the 2017–18 and 2018–19 2-year EBLP** |
| --- | --- | --- | --- | --- | --- |
| ALL | ELA | 11–29 | 154 | 0.32 | 0.65 |
| ALL | ELA | 30–149 | 2,142 | 0.30 | 0.47 |
| ALL | ELA | ≥150 | 4,989 | 0.42 | 0.49 |
| ALL | Mathematics | 11–29 | 148 | 0.38 | 0.66 |
| ALL | Mathematics | 30–149 | 2,146 | 0.41 | 0.53 |
| ALL | Mathematics | ≥150 | 4,986 | 0.58 | 0.62 |
| AA | ELA | 11–29 | 1,245 | 0.16 | 0.58 |
| AA | ELA | ≥30 | 955 | 0.28 | 0.56 |
| AA | Mathematics | 11–29 | 1,244 | 0.20 | 0.61 |
| AA | Mathematics | ≥30 | 955 | 0.32 | 0.60 |
| AI | ELA | ≥11 | 93 | 0.12 | 0.58 |
| AI | Mathematics | ≥11 | 94 | 0.05 | 0.59 |
| AS | ELA | 11–29 | 1,124 | 0.07 | 0.51 |
| AS | ELA | 30–149 | 1,211 | 0.30 | 0.53 |
| AS | ELA | ≥150 | 279 | 0.46 | 0.55 |
| AS | Mathematics | 11–29 | 1,124 | 0.29 | 0.61 |
| AS | Mathematics | 30–149 | 1,212 | 0.52 | 0.69 |
| AS | Mathematics | ≥150 | 278 | 0.73 | 0.77 |
| FI | ELA | 11–29 | 617 | 0.03 | 0.45 |
| FI | ELA | ≥30 | 368 | 0.33 | 0.53 |
| FI | Mathematics | 11–29 | 617 | 0.30 | 0.61 |
| FI | Mathematics | ≥30 | 368 | 0.48 | 0.68 |
| HI | ELA | 11–29 | 764 | 0.17 | 0.56 |
| HI | ELA | 30–149 | 3,453 | 0.27 | 0.49 |
| HI | ELA | ≥150 | 2,631 | 0.40 | 0.48 |
| HI | Mathematics | 11–29 | 767 | 0.25 | 0.58 |
| HI | Mathematics | 30–149 | 3,453 | 0.36 | 0.52 |
| HI | Mathematics | ≥150 | 2,630 | 0.52 | 0.58 |
| PI | ELA | 11–149 | 86 | 0.08 | 0.40 |
| PI | Mathematics | 11–149 | 86 | 0.19 | 0.69 |
| WH | ELA | 11–29 | 1,216 | 0.12 | 0.51 |
| WH | ELA | 30–149 | 2,754 | 0.29 | 0.50 |
| WH | ELA | ≥150 | 775 | 0.43 | 0.49 |
| WH | Mathematics | 11–29 | 1,218 | 0.18 | 0.49 |
| WH | Mathematics | 30–149 | 2,752 | 0.41 | 0.55 |
| WH | Mathematics | ≥150 | 774 | 0.61 | 0.65 |
| MR | ELA | 11–29 | 1,323 | 0.11 | 0.53 |
| MR | ELA | ≥30 | 610 | 0.21 | 0.51 |
| MR | Mathematics | 11–29 | 1,324 | 0.30 | 0.61 |
| MR | Mathematics | ≥30 | 608 | 0.47 | 0.66 |
| EL | ELA | 11–29 | 1,091 | 0.15 | 0.57 |
| EL | ELA | 30–149 | 3,913 | 0.28 | 0.52 |
| EL | ELA | ≥150 | 1,350 | 0.43 | 0.52 |
| EL | Mathematics | 11–29 | 1,087 | 0.37 | 0.68 |
| EL | Mathematics | 30–149 | 3,916 | 0.45 | 0.62 |
| EL | Mathematics | ≥150 | 1,348 | 0.57 | 0.64 |
| ELO | ELA | 11–29 | 1,620 | 0.06 | 0.48 |
| ELO | ELA | 30–149 | 3,302 | 0.28 | 0.51 |
| ELO | ELA | ≥150 | 324 | 0.32 | 0.44 |
| ELO | Mathematics | 11–29 | 1,621 | 0.24 | 0.60 |
| ELO | Mathematics | 30–149 | 3,301 | 0.35 | 0.57 |
| ELO | Mathematics | ≥150 | 321 | 0.40 | 0.51 |
| RFP | ELA | 11–29 | 1,854 | 0.14 | 0.53 |
| RFP | ELA | 30–149 | 3,241 | 0.32 | 0.55 |
| RFP | ELA | ≥150 | 447 | 0.43 | 0.52 |
| RFP | Mathematics | 11–29 | 1,853 | 0.34 | 0.63 |
| RFP | Mathematics | 30–149 | 3,243 | 0.50 | 0.67 |
| RFP | Mathematics | ≥150 | 445 | 0.62 | 0.69 |
| EO | ELA | 11–29 | 515 | 0.24 | 0.60 |
| EO | ELA | 30–149 | 4,145 | 0.29 | 0.49 |
| EO | ELA | ≥150 | 2,519 | 0.50 | 0.56 |
| EO | Mathematics | 11–29 | 515 | 0.26 | 0.56 |
| EO | Mathematics | 30–149 | 4,153 | 0.42 | 0.56 |
| EO | Mathematics | ≥150 | 2,510 | 0.64 | 0.68 |
| SED | ELA | 11–29 | 530 | 0.13 | 0.54 |
| SED | ELA | 30–149 | 3,423 | 0.27 | 0.46 |
| SED | ELA | ≥150 | 3,067 | 0.40 | 0.48 |
| SED | Mathematics | 11–29 | 532 | 0.23 | 0.54 |
| SED | Mathematics | 30–149 | 3,428 | 0.37 | 0.51 |
| SED | Mathematics | ≥150 | 3,060 | 0.53 | 0.58 |
| SWD | ELA | 11–29 | 3,205 | 0.04 | 0.42 |
| SWD | ELA | ≥30 | 3,025 | 0.19 | 0.44 |
| SWD | Mathematics | 11–29 | 3,210 | 0.04 | 0.36 |
| SWD | Mathematics | ≥30 | 3,019 | 0.16 | 0.36 |
| FOS | ELA | 11–29 | 29 | -0.02 | 0.35 |
| FOS | Mathematics | 11–29 | 29 | -0.35 | 0.11 |
| HOM | ELA | 11–29 | 767 | 0.21 | 0.56 |
| HOM | ELA | ≥30 | 617 | 0.21 | 0.44 |
| HOM | Mathematics | 11–29 | 768 | 0.30 | 0.61 |
| HOM | Mathematics | ≥30 | 614 | 0.38 | 0.59 |

1. The size intervals are based on the number of students within schools with growth scores in grade levels four through eight in *both* 2017–18 and 2018–19 for the student group and subject of interest (indicated in the first two columns). Given that not all schools have estimates in both years, the number of schools in each interval may be smaller than that reported in [Table 2](#Table2).

**5.B.1. Summary of Table 3 Results**

As shown in Table 3, the EBLP weighted averages have higher cross-year correlations than do the simple averages. This holds true for all schools, both large and small, and for student groups. The improvements in stability are most pronounced for small schools.

1. **Results at the LEA Level**

The results for evaluating the accuracy and cross-year stability at the LEA level are presented in this section. Given some concerns with the accuracy results for student groups (excluding the ALL group) at the LEA level, additional reporting options are also discussed. The same evaluation indices used for the school-level analysis (i.e., mean accuracy ratio and percentage of LEAs with EBLP weighted averages estimated to be as accurate as or more accurate than the simple average) were used here.

**6.A. Accuracy Results**

All results are reported by student group, subject, and LEA size, with the following size intervals applied:

* LEAs with 11 to 29 students
* LEAs with 30 to 149 students
* LEAs with 150 to 1,499 students
* LEAs with 1,500 or more students

In cases where there are fewer than 100 LEAs within one of the designated size intervals, a combined interval will be used. For example, for the ALL student group for ELA, only 41 students had ELA growth scores in the 11 to 29 interval; therefore, results are reported in the combined interval of 11 to 149.

Table 4 provides the results for ALL students within the LEAs as well as for the 16 additional student groups of interest.

**Table 4. Improvement in Accuracy of Growth Estimates Using Two-Year EBLP Weighted Averages Versus Simple Averages at the LEA Level**

| **Student Group** | **Subject** | **2018–19 LEA Size1** | **Number of LEAs** | **Mean Accuracy Ratio** | **Percentage of LEAs with Improved Accuracy2** |
| --- | --- | --- | --- | --- | --- |
| ALL | ELA | 11–149 | 205 | 1.44 | 100% |
| ALL | ELA | 150–1,499 | 323 | 1.09 | 100% |
| ALL | ELA | ≥1,500 | 299 | 1.01 | 100% |
| ALL | Mathematics | 11–149 | 205 | 1.39 | 100% |
| ALL | Mathematics | 150–1,499 | 324 | 1.07 | 100% |
| ALL | Mathematics | ≥1,500 | 298 | 1.01 | 100% |
| AA | ELA | 11–149 | 240 | 1.68 | 100% |
| AA | ELA | ≥150 | 110 | 0.99 | 55% |
| AA | Mathematics | 11–149 | 240 | 2.08 | 100% |
| AA | Mathematics | ≥150 | 110 | 1.36 | 85% |
| AI | ELA | 11–1,499 | 233 | 2.17 | 100% |
| AI | Mathematics | 11–1,499 | 232 | 2.33 | 100% |
| AS | ELA | 11–149 | 218 | 1.45 | 100% |
| AS | ELA | ≥150 | 184 | 1.01 | 63% |
| AS | Mathematics | 11–149 | 218 | 1.59 | 99% |
| AS | Mathematics | ≥150 | 184 | 0.66 | 24% |
| FI | ELA | ≥11 | 321 | 1.64 | 97% |
| FI | Mathematics | ≥11 | 320 | 1.49 | 87% |
| HI | ELA | 11–149 | 264 | 1.73 | 100% |
| HI | ELA | 150–1,499 | 305 | 1.09 | 72% |
| HI | ELA | ≥1500 | 185 | 0.90 | 28% |
| HI | Mathematics | 11–149 | 265 | 1.65 | 100% |
| HI | Mathematics | 150–1,499 | 306 | 1.07 | 78% |
| HI | Mathematics | ≥1,500 | 184 | 0.94 | 34% |
| PI | ELA | 11–1,499 | 176 | 1.58 | 95% |
| PI | Mathematics | 11–1,499 | 176 | 1.98 | 99% |
| WH | ELA | 11–149 | 324 | 1.58 | 100% |
| WH | ELA | ≥150 | 422 | 1.09 | 95% |
| WH | Mathematics | 11–149 | 324 | 1.43 | 100% |
| WH | Mathematics | ≥150 | 422 | 1.06 | 100% |
| MR | ELA | 11–29 | 114 | 1.88 | 100% |
| MR | ELA | 30–149 | 197 | 1.57 | 100% |
| MR | ELA | ≥150 | 139 | 1.19 | 83% |
| MR | Mathematics | 11–29 | 113 | 1.74 | 100% |
| MR | Mathematics | 30–149 | 198 | 1.36 | 96% |
| MR | Mathematics | ≥150 | 138 | 0.98 | 48% |
| EL | ELA | 11–149 | 252 | 1.88 | 100% |
| EL | ELA | 150–1,499 | 304 | 1.06 | 59% |
| EL | ELA | ≥1,500 | 120 | 0.57 | 11% |
| EL | Mathematics | 11–149 | 251 | 2.13 | 100% |
| EL | Mathematics | 150–1,499 | 304 | 1.11 | 60% |
| EL | Mathematics | ≥1,500 | 120 | 0.35 | 3% |
| ELO | ELA | 11–149 | 289 | 1.59 | 100% |
| ELO | ELA | ≥150 | 334 | 0.91 | 43% |
| ELO | Mathematics | 11–149 | 289 | 1.87 | 99% |
| ELO | Mathematics | ≥150 | 334 | 0.86 | 39% |
| RFP | ELA | 11–149 | 273 | 1.71 | 100% |
| RFP | ELA | ≥150 | 330 | 1.01 | 54% |
| RFP | Mathematics | 11–149 | 273 | 2.04 | 100% |
| RFP | Mathematics | ≥150 | 330 | 1.12 | 60% |
| EO | ELA | 11–149 | 265 | 1.50 | 100% |
| EO | ELA | 150–1,499 | 345 | 1.08 | 97% |
| EO | ELA | ≥1,500 | 205 | 1.01 | 95% |
| EO | Mathematics | 11–149 | 265 | 1.38 | 100% |
| EO | Mathematics | 150–1,499 | 346 | 1.06 | 100% |
| EO | Mathematics | ≥1,500 | 204 | 1.01 | 100% |
| SED | ELA | 11–149 | 270 | 1.50 | 100% |
| SED | ELA | 150–1,499 | 330 | 0.98 | 60% |
| SED | ELA | ≥1,500 | 199 | 0.58 | 9% |
| SED | Mathematics | 11–149 | 270 | 1.44 | 100% |
| SED | Mathematics | 150–1,499 | 330 | 1.02 | 71% |
| SED | Mathematics | ≥1,500 | 199 | 0.87 | 18% |
| SWD | ELA | 11–29 | 121 | 1.78 | 100% |
| SWD | ELA | 30–149 | 232 | 1.35 | 100% |
| SWD | ELA | ≥150 | 322 | 1.06 | 84% |
| SWD | Mathematics | 11–29 | 121 | 1.66 | 100% |
| SWD | Mathematics | 30–149 | 232 | 1.27 | 100% |
| SWD | Mathematics | ≥150 | 322 | 1.05 | 90% |
| FOS | ELA | 11–1,499 | 195 | 1.96 | 100% |
| FOS | Mathematics | 11–1,499 | 192 | 2.13 | 100% |
| HOM | ELA | 11–29 | 130 | 1.65 | 100% |
| HOM | ELA | 30–149 | 152 | 1.32 | 98% |
| HOM | ELA | ≥150 | 101 | 0.90 | 42% |
| HOM | Mathematics | 11–29 | 129 | 1.75 | 100% |
| HOM | Mathematics | 30–149 | 152 | 1.48 | 100% |
| HOM | Mathematics | ≥150 | 101 | 1.17 | 82% |

1. The size intervals are based the number of students within LEAs with growth scores in grade levels four through eight in 2018–19 for the student group and subject of interest (indicated in the first two columns).
2. The percentage of LEAs with improved accuracy for the two-year EBLP versus the simple average represents the percentage of LEAs whose estimated accuracy for the two-year EBLP is as good as or better than that of the simple average.

**6.A.1. Summary of Table 4 Results**

The results for the ALL student group at the LEA level are similar to those attained at the school level:

* The EBLP weighted averages are estimated to be more accurate for every LEA (for the ALL group), with the largest gains in accuracy observed for LEAs serving fewer students.
* In addition, for the smaller student groups, across all LEAs, the EBLP weighted averages are estimated to be more accurate than the simple averages.

However, among the larger student groups, the percentage of LEAs with estimated gains in accuracy is noticeably lower than 100 percent (as low as 3 percent in some cases), and the mean accuracy ratio is less than 1.0.

* Seventeen percent of the 16,395 AGMs (represented by the 16 additional student groups) have EBLP weighted averages that are estimated to be less accurate than the simple average, with the median LEA size for these cases being 1,160.

The likely reason that EBLP performs differently at the school and LEA levels is that there are far fewer LEAs than there are schools. (There are approximately 10 times as many schools as LEAs in the California data analyzed.)

The EBLP method requires optimal weights to be estimated from the data, and imprecision in these optimal weights contributes to estimation errors for the EBLPs.

The optimal weights are estimated more precisely when there are a large number of aggregation units (as with schools) in the model and less precisely when there are relatively few aggregation units in the model (as with LEAs).

Given the tenfold magnitude of schools over LEAs in the data, the EBLP weights for LEAs are not as precise as they are for schools.

The ALL student group is less sensitive to these errors because it uses more LEAs than any of the other student groups. For the overall group, there were 827 LEAs with at least 11 students with growth scores in a given subject in 2018–19 and at least 1 student with growth scores in the subject of interest in 2017–18. However, for the other student groups, the number of LEAs with reportable AGMs is, on average, only 530 and dips as low as 176 (for the PI group).

Please note that for large student groups at the LEA level, such as EL and SED, the EBLP improved accuracy for fewer LEAs. For example, among LEAs with 1,500 or more growth scores, the EBLP improved the accuracy of:

* Mathematic scores for only 3 percent of the EL student groups at the LEA level
* Mathematic scores for only 18 percent of the SED student groups at the LEA level.

For these groups, the simple average would generally provide more accurate AGMs than the EBLP although the differences in the values of the EBLPs and simple averages themselves are likely small. However, for less populous student groups, such as PI and AI, which are represented in only a small number of LEAs, the EBLP resulted in more accurate AGMs.

In summary the EBLP improves the accuracy of the AGM for the ALL student group for every LEA. However, for a significant share of student groups at the LEA level—specifically, those with larger numbers of growth scores—the simple average is estimated to be more accurate. The accuracy results for the student groups at the LEA level led to further exploration of AGM reporting options for LEAs. These will be discussed in [section 6.C](#_6.C._Exploration_of).

**6.B. Stability Results**

As mentioned earlier, application of the EBLP increases stability of the growth measures across years. This stability is evident in the correlation (ranging from –1 to +1) of AGMs across school years 2017–18 and 2018–19. Higher positive values indicate that the measures are more similar or, in this case, that growth measures are more stable across time.

Table 5 compares the cross-year stability of the 2017–18 and 2018–19 two-year EBLP weighted averages to that of the simple averages at the LEA level.

**Table 5. Cross-Year Stability of the 2018–19 Two-Year EBLP Weighted Average and the Simple Average for LEAs**

| **Student Group** | **Subject** | **2018–19 LEA Size1** | **Number of LEAs** | **Correlation Between the 2017–18 and 2018–19 Simple Average** | **Correlation Between the 2017–18 and 2018–19 2-year EBLP** |
| --- | --- | --- | --- | --- | --- |
| ALL | ELA | 11–149 | 200 | 0.32 | 0.59 |
| ALL | ELA | 150–1,499 | 323 | 0.37 | 0.47 |
| ALL | ELA | ≥1,500 | 299 | 0.65 | 0.66 |
| ALL | Mathematics | 11–149 | 200 | 0.33 | 0.57 |
| ALL | Mathematics | 150–1,499 | 324 | 0.43 | 0.51 |
| ALL | Mathematics | ≥1,500 | 298 | 0.85 | 0.85 |
| AA | ELA | 11–149 | 226 | 0.13 | 0.59 |
| AA | ELA | ≥150 | 110 | 0.67 | 0.73 |
| AA | Mathematics | 11–149 | 226 | 0.08 | 0.64 |
| AA | Mathematics | ≥150 | 110 | 0.63 | 0.84 |
| AI | ELA | 11–1,499 | 217 | 0.17 | 0.60 |
| AI | Mathematics | 11–1,499 | 216 | 0.11 | 0.61 |
| AS | ELA | 11–149 | 212 | -0.03 | 0.48 |
| AS | ELA | ≥150 | 184 | 0.53 | 0.59 |
| AS | Mathematics | 11–149 | 212 | 0.33 | 0.72 |
| AS | Mathematics | ≥150 | 184 | 0.80 | 0.86 |
| FI | ELA | ≥11 | 305 | 0.06 | 0.50 |
| FI | Mathematics | ≥11 | 305 | 0.18 | 0.54 |
| HI | ELA | 11–149 | 256 | 0.20 | 0.61 |
| HI | ELA | 150–1,499 | 305 | 0.41 | 0.61 |
| HI | ELA | ≥1,500 | 185 | 0.63 | 0.68 |
| HI | Mathematics | 11–149 | 257 | 0.22 | 0.62 |
| HI | Mathematics | 150–1,499 | 306 | 0.44 | 0.57 |
| HI | Mathematics | ≥1,500 | 184 | 0.72 | 0.74 |
| PI | ELA | 11–1,499 | 169 | 0.06 | 0.43 |
| PI | Mathematics | 11–1,499 | 169 | 0.13 | 0.51 |
| WH | ELA | 11–149 | 316 | 0.21 | 0.51 |
| WH | ELA | ≥150 | 422 | 0.48 | 0.60 |
| WH | Mathematics | 11–149 | 316 | 0.28 | 0.53 |
| WH | Mathematics | ≥150 | 422 | 0.59 | 0.65 |
| MR | ELA | 11–29 | 102 | 0.18 | 0.63 |
| MR | ELA | 30–149 | 197 | 0.23 | 0.59 |
| MR | ELA | ≥150 | 139 | 0.52 | 0.69 |
| MR | Mathematics | 11–29 | 102 | 0.13 | 0.67 |
| MR | Mathematics | 30–149 | 198 | 0.47 | 0.66 |
| MR | Mathematics | ≥150 | 138 | 0.70 | 0.76 |
| EL | ELA | 11–149 | 246 | 0.06 | 0.62 |
| EL | ELA | 150–1,499 | 304 | 0.50 | 0.71 |
| EL | ELA | ≥1,500 | 120 | 0.70 | 0.80 |
| EL | Mathematics | 11–149 | 246 | 0.30 | 0.72 |
| EL | Mathematics | 150–1,499 | 304 | 0.67 | 0.86 |
| EL | Mathematics | ≥1,500 | 120 | 0.82 | 0.90 |
| ELO | ELA | 11–149 | 277 | 0.05 | 0.49 |
| ELO | ELA | ≥150 | 334 | 0.44 | 0.58 |
| ELO | Mathematics | 11–149 | 276 | 0.29 | 0.72 |
| ELO | Mathematics | ≥150 | 334 | 0.60 | 0.77 |
| RFP | ELA | 11–149 | 259 | 0.15 | 0.63 |
| RFP | ELA | ≥150 | 330 | 0.51 | 0.64 |
| RFP | Mathematics | 11–149 | 260 | 0.23 | 0.70 |
| RFP | Mathematics | ≥150 | 330 | 0.76 | 0.86 |
| EO | ELA | 11–149 | 262 | 0.31 | 0.62 |
| EO | ELA | 150–1,499 | 345 | 0.46 | 0.56 |
| EO | ELA | ≥1,500 | 205 | 0.69 | 0.70 |
| EO | Mathematics | 11–149 | 262 | 0.24 | 0.48 |
| EO | Mathematics | 150–1,499 | 346 | 0.60 | 0.64 |
| EO | Mathematics | ≥1,500 | 204 | 0.86 | 0.86 |
| SED | ELA | 11–149 | 257 | 0.15 | 0.55 |
| SED | ELA | 150–1,499 | 330 | 0.43 | 0.57 |
| SED | ELA | ≥1,500 | 199 | 0.58 | 0.62 |
| SED | Mathematics | 11–149 | 257 | 0.25 | 0.54 |
| SED | Mathematics | 150–1,499 | 330 | 0.42 | 0.52 |
| SED | Mathematics | ≥1,500 | 199 | 0.74 | 0.75 |
| SWD | ELA | 11–29 | 114 | -0.06 | 0.37 |
| SWD | ELA | 30–149 | 232 | 0.05 | 0.46 |
| SWD | ELA | ≥150 | 322 | 0.51 | 0.60 |
| SWD | Mathematics | 11–29 | 113 | -0.18 | 0.27 |
| SWD | Mathematics | 30–149 | 232 | 0.04 | 0.35 |
| SWD | Mathematics | ≥150 | 322 | 0.49 | 0.56 |
| FOS | ELA | 11–1,499 | 180 | 0.02 | 0.61 |
| FOS | Mathematics | 11–1,499 | 176 | 0.07 | 0.51 |
| HOM | ELA | 11–29 | 100 | -0.03 | 0.58 |
| HOM | ELA | 30–149 | 149 | 0.09 | 0.53 |
| HOM | ELA | ≥150 | 101 | 0.52 | 0.64 |
| HOM | Mathematics | 11–29 | 100 | 0.04 | 0.49 |
| HOM | Mathematics | 30–149 | 149 | 0.05 | 0.44 |
| HOM | Mathematics | ≥150 | 101 | 0.65 | 0.77 |

1. The size intervals refer to the number of students within LEAs with growth scores in grades four through eight in *both* 2017–18 and 2018–19 for the student group and subject of interest (indicated in the first two columns). The number of LEAs in each bin is smaller than in Table 4, given that not all schools have estimates in both years.

**6.B.1. Summary of Table 5 Results**

In terms of cross-year stability, the EBLP approach performs similarly at the LEA and school levels. When compared to the simple average for every student group, by subject and size, the EBLP weighted averages at the LEA level have as good or better AGM correlations between two consecutive years (2017–18 and 2018–19), with the largest gains posted for the smaller group sizes. For example, for LEAs with 11 to 29 students in the HOM student group, the ELA correlation is -0.03 using the simple average, compared with 0.58 using the EBLP weighted average.

In an analysis of any AGM across the 17 student groups with 11 to 149 students (with growth scores in 2018–19), the correlation is only 0.24 for simple averages, compared with 0.62 for the EBLP weighted averages (found by calculations done separately from those shown in Table 5). The EBLP weighted averages are noticeably more stable for smaller LEAs and for smaller student groups.at the LEA level.

In contrast, for the largest LEAs (i.e., those with 1,500 or more students in any of the 17 student groups), the simple average cross-year correlation is 0.81, compared with 0.83 for the EBLP weighted averages. (These correlations are not shown in the table.) This difference in stability for large student groups in LEAs is negligible.

Consequently, the EBLP weighted averages improve stability where it is most needed and reduce the disparity in cross-year stability between small and large student groups within LEAs.

**6.C. Exploration of Reporting Options for LEA Student Groups**

As described in sections 6.A and 6.B, the performance of the EBLP weighted average at the LEA level was mixed. It improved stability for every student group at the LEA level including the ALL student group, but increased accuracy for only two groups:

* The ALL student group
* Student groups at the LEA level with small numbers of growth scores

For a portion of student groups at the LEA level with large numbers of growth scores, the EBLP approach did not improve accuracy over the simple average. Therefore, ETS explored a hybrid method that combines the two approaches:

* The EBLP is applied to student groups at the LEA level that have a small number of growth scores, as determined by a specific cutoff, or *n*-size
* The simple average is applied to student groups at the LEA level that have a larger number of growth scores

ETS’ exploration of such a hybrid approach tested different options for determining when to report the EBLP and when to report the simple average. ETS presented the results of the study to the TDG. Full details of the exploration and its results are available in the December memorandum to the TDG, which is provided in Appendix A.

Based on the results from this analysis, the TDG recommended that an *n*-size of 500 growth scores be used when determining whether to report the EBLP or simple average, yielding the following hybrid reporting plan at the LEA level:

1. For the AGMs for the ALL student group, report the EBLP for all LEAs with 11 or more growth scores.
2. For the AGMs for individual student groups:
3. Report the EBLP for groups in which there are 11 to 500 growth scores (Note: this rule applies only to LEAs, not schools.)
4. Report the simple average for groups in which there are more than 500 growth scores (Note: this rule applies only to LEAs, not schools.)

[Table 6](#table6) shows the improvement in the accuracy, for the 16 student groups, when the hybrid approach is applied at the LEA level. To clearly show the impact of the hybrid method, the size intervals were updated to specifically align with an *n*-size cutoff of 500:

* LEAs with 11 to 29 growth scores in the student group
* LEAs with 30 to 149 growth scores in the student group
* LEAs with 150 to 500 growth scores in the student group
* LEAs with 501 or more growth scores in the student group

As done previously, to ensure a sufficient number of LEAs in a size interval for precise estimation of the summary statistics (e.g., mean accuracy ratio), size intervals that had fewer than 100 LEAs were combined (with the exception of the size interval for groups with 501 or more students).

The simple average is used for all student groups at the LEA level with more than 500 growth scores. Consequently, the estimated accuracy for the hybrid estimator is exactly equal to that of the simple average. Accordingly, regardless of the number of LEAs in the “501 or more” interval, the mean accuracy ratio will exactly equal 1.0, and the percentage of LEAs estimated to have student-group AGMs as accurate as or more accurate than the simple average will be exactly 100 percent. For this interval, no minimum number of LEAs is needed to have precise estimation of the summary statistics. Results are shown in [Table 6](#table6).

**Table 6. Improvement in Accuracy of Growth Estimates Using the Hybrid Approach Versus Simple Averages for Student Groups Within LEAs**

| **Student Group** | **Subject** | **2018–19 LEA Size1** | **Number of LEAs** | **Mean Accuracy Ratio** | **Percentage of LEAs with Improved Accuracy2** |
| --- | --- | --- | --- | --- | --- |
| AA | ELA | 11–500 | 308 | 1.56 | 94% |
| AA | ELA | >500 | 42 | 1.00 | 100% |
| AA | Mathematics | 11–500 | 308 | 1.95 | 100% |
| AA | Mathematics | >500 | 42 | 1.00 | 100% |
| AI | ELA | 11–500 | 233 | 2.17 | 100% |
| AI | Mathematics | 11–500 | 232 | 2.33 | 100% |
| AS | ELA | 11–500 | 312 | 1.34 | 95% |
| AS | ELA | >500 | 90 | 1.00 | 100% |
| AS | Mathematics | 11–500 | 312 | 1.37 | 81% |
| AS | Mathematics | >500 | 90 | 1.00 | 100% |
| FI | ELA | 11–500 | 303 | 1.68 | 99% |
| FI | ELA | >500 | 18 | 1.00 | 100% |
| FI | Mathematics | 11–500 | 302 | 1.53 | 92% |
| FI | Mathematics | >500 | 18 | 1.00 | 100% |
| HI | ELA | 11–149 | 264 | 1.73 | 100% |
| HI | ELA | 150–500 | 152 | 1.20 | 91% |
| HI | ELA | >500 | 338 | 1.00 | 100% |
| HI | Mathematics | 11–149 | 265 | 1.65 | 100% |
| HI | Mathematics | 150–500 | 153 | 1.15 | 94% |
| HI | Mathematics | >500 | 337 | 1.00 | 100% |
| PI | ELA | 11–500 | 175 | 1.59 | 95% |
| PI | ELA | >500 | 1 | 1.00 | 100% |
| PI | Mathematics | 11–500 | 175 | 1.98 | 100% |
| PI | Mathematics | >500 | 1 | 1.00 | 100% |
| WH | ELA | 11–149 | 324 | 1.58 | 100% |
| WH | ELA | 150–500 | 199 | 1.14 | 96% |
| WH | ELA | >500 | 223 | 1.00 | 100% |
| WH | Mathematics | 11–149 | 324 | 1.43 | 100% |
| WH | Mathematics | 150–500 | 200 | 1.09 | 100% |
| WH | Mathematics | >500 | 222 | 1.00 | 100% |
| MR | ELA | 11–29 | 114 | 1.88 | 100% |
| MR | ELA | 30–149 | 197 | 1.57 | 100% |
| MR | ELA | 150–500 | 117 | 1.22 | 90% |
| MR | ELA | >500 | 22 | 1.00 | 100% |
| MR | Mathematics | 11–29 | 113 | 1.74 | 100% |
| MR | Mathematics | 30–149 | 198 | 1.36 | 96% |
| MR | Mathematics | 150–500 | 116 | 1.00 | 53% |
| MR | Mathematics | >500 | 22 | 1.00 | 100% |
| EL | ELA | 11–149 | 252 | 1.88 | 100% |
| EL | ELA | 150–500 | 160 | 1.25 | 84% |
| EL | ELA | >500 | 264 | 1.00 | 100% |
| EL | Mathematics | 11–149 | 251 | 2.13 | 100% |
| EL | Mathematics | 150–500 | 161 | 1.40 | 83% |
| EL | Mathematics | >500 | 263 | 1.00 | 100% |
| ELO | ELA | 11–149 | 289 | 1.59 | 100% |
| ELO | ELA | 150–500 | 164 | 1.05 | 75% |
| ELO | ELA | >500 | 170 | 1.00 | 100% |
| ELO | Mathematics | 11–149 | 289 | 1.87 | 99% |
| ELO | Mathematics | 150–500 | 165 | 1.09 | 65% |
| ELO | Mathematics | >500 | 169 | 1.00 | 100% |
| RFP | ELA | 11–149 | 273 | 1.71 | 100% |
| RFP | ELA | 150–500 | 164 | 1.12 | 78% |
| RFP | ELA | >500 | 166 | 1.00 | 100% |
| RFP | Mathematics | 11–149 | 273 | 2.04 | 100% |
| RFP | Mathematics | 150–500 | 165 | 1.39 | 83% |
| RFP | Mathematics | >500 | 165 | 1.00 | 100% |
| EO | ELA | 11–149 | 265 | 1.50 | 100% |
| EO | ELA | 150–500 | 158 | 1.13 | 100% |
| EO | ELA | >500 | 392 | 1.00 | 100% |
| EO | Mathematics | 11–149 | 265 | 1.38 | 100% |
| EO | Mathematics | 150–500 | 158 | 1.09 | 100% |
| EO | Mathematics | >500 | 392 | 1.00 | 100% |
| SED | ELA | 11–149 | 270 | 1.50 | 100% |
| SED | ELA | 150–500 | 165 | 1.12 | 88% |
| SED | ELA | >500 | 364 | 1.00 | 100% |
| SED | Mathematics | 11–149 | 270 | 1.44 | 100% |
| SED | Mathematics | 150–500 | 166 | 1.09 | 92% |
| SED | Mathematics | >500 | 363 | 1.00 | 100% |
| SWD | ELA | 11–29 | 121 | 1.78 | 100% |
| SWD | ELA | 30–149 | 232 | 1.35 | 100% |
| SWD | ELA | 150–500 | 189 | 1.10 | 96% |
| SWD | ELA | >500 | 133 | 1.00 | 100% |
| SWD | Mathematics | 11–29 | 121 | 1.66 | 100% |
| SWD | Mathematics | 30–149 | 232 | 1.27 | 100% |
| SWD | Mathematics | 150–500 | 189 | 1.07 | 95% |
| SWD | Mathematics | >500 | 133 | 1.00 | 100% |
| FOS | ELA | 11–500 | 194 | 1.97 | 100% |
| FOS | ELA | >500 | 1 | 1.00 | 100% |
| FOS | Mathematics | 11–500 | 191 | 2.13 | 100% |
| FOS | Mathematics | >500 | 1 | 1.00 | 100% |
| HOM | ELA | 11–29 | 130 | 1.65 | 100% |
| HOM | ELA | 30–500 | 214 | 1.24 | 87% |
| HOM | ELA | >500 | 39 | 1.00 | 100% |
| HOM | Mathematics | 11–29 | 129 | 1.75 | 100% |
| HOM | Mathematics | 30–500 | 214 | 1.41 | 99% |
| HOM | Mathematics | >500 | 39 | 1.00 | 100% |

1. The size intervals refer to the number of students within LEAs with growth scores in grade levels four through eight in 2018–19 for the student group and subject of interest (indicated in the first two columns).
2. The percentage of LEAs with improved accuracy for the hybrid approach versus the simple average represents the percentage of LEAs whose estimated accuracy for the hybrid approach is as good as or better than that of the simple average.

**6.C.1. Summary of Table 6 Results**

[Table 6](#table6) shows that:

* For all student groups with all size intervals that fall below 501 (i.e., student groups with 500 or fewer students who had growth scores in 2018-19), the mean accuracy ratio is greater than or equal to 1.0. Using the hybrid approach the majority of LEAs are estimated to have student-level AGMs as accurate as or more accurate than exclusively assigning the simple average (i.e., the percentage with improved accuracy is greater than 50 percent). For these size intervals, the hybrid approach assigns LEAs the EBLP weighted average for their AGM.
* For all student groups with 501 or more students who had growth scores in 2018–19, the mean accuracy ratio is exactly equal to 1.0, and the percentage of LEAs with equivalent or better accuracy is exactly 100 percent. This is a result of that fact that, for this size interval, the hybrid approach assigns LEAs the simple average for the AGM, resulting in the same estimated accuracy for the simple average as for the hybrid estimator.
* The hybrid approach reduces the prevalence of cases in which the EBLP weighted average is estimated to be less accurate than the simple average (from 17 percent to 3 percent).

Table 7 provides the updated stability results for the hybrid approach for the 16 student groups. Note that as with the accuracy results for the hybrid approach, the size intervals aligned to the *n*-size cutoff of 500 were used to illustrate more clearly the impact of this approach. However, unlike for the accuracy results, any size interval, including the largest (i.e., “>500”), was combined with adjacent size intervals if the interval had fewer than 100 LEAs for the group. In some cases, there are very few LEAs with 501 or more students with growth scores (e.g., for foster youth, there is only one LEA in this interval), making correlation estimates unstable or impossible to calculate. Accordingly, the size intervals may differ somewhat from those given in [Table 6](#table6).

**Table 7. Cross-Year Stability of the 2018–19 Hybrid Approach and the Simple Average for Student Groups Within LEAs**

| Student Group | Subject | 2018–19 LEA Size1 | Number of LEAs | Correlation Between the 2017–18 and 2018–19 Simple Average | Correlation Between the 2017–18 and 2018–19 Hybrid Approach |
| --- | --- | --- | --- | --- | --- |
| AA | ELA | 11–149 | 226 | 0.13 | 0.59 |
| AA | ELA | ≥150 | 110 | 0.67 | 0.73 |
| AA | Mathematics | 11–149 | 226 | 0.08 | 0.64 |
| AA | Mathematics | ≥150 | 110 | 0.63 | 0.84 |
| AI | ELA | ≥11 | 217 | 0.17 | 0.60 |
| AI | Mathematics | ≥11 | 216 | 0.11 | 0.61 |
| AS | ELA | 11–149 | 212 | -0.03 | 0.48 |
| AS | ELA | ≥150 | 184 | 0.53 | 0.59 |
| AS | Mathematics | 11–149 | 212 | 0.33 | 0.72 |
| AS | Mathematics | ≥150 | 184 | 0.80 | 0.86 |
| FI | ELA | ≥11 | 305 | 0.06 | 0.50 |
| FI | Mathematics | ≥11 | 305 | 0.18 | 0.54 |
| HI | ELA | 11–149 | 256 | 0.20 | 0.61 |
| HI | ELA | 150–500 | 152 | 0.37 | 0.63 |
| HI | ELA | >500 | 338 | 0.55 | 0.62 |
| HI | Mathematics | 11–149 | 257 | 0.22 | 0.62 |
| HI | Mathematics | 150–500 | 153 | 0.31 | 0.49 |
| HI | Mathematics | >500 | 337 | 0.68 | 0.71 |
| PI | ELA | ≥11 | 169 | 0.06 | 0.43 |
| PI | Mathematics | ≥11 | 169 | 0.13 | 0.51 |
| WH | ELA | 11–149 | 316 | 0.21 | 0.51 |
| WH | ELA | 150–500 | 199 | 0.40 | 0.55 |
| WH | ELA | >500 | 223 | 0.60 | 0.66 |
| WH | Mathematics | 11–149 | 316 | 0.28 | 0.53 |
| WH | Mathematics | 150–500 | 200 | 0.47 | 0.55 |
| WH | Mathematics | >500 | 222 | 0.74 | 0.76 |
| MR | ELA | 11–29 | 102 | 0.18 | 0.63 |
| MR | ELA | 30–149 | 197 | 0.23 | 0.59 |
| MR | ELA | ≥150 | 139 | 0.52 | 0.69 |
| MR | Mathematics | 11–29 | 102 | 0.13 | 0.67 |
| MR | Mathematics | 30–149 | 198 | 0.47 | 0.66 |
| MR | Mathematics | ≥150 | 138 | 0.70 | 0.76 |
| EL | ELA | 11–149 | 246 | 0.06 | 0.62 |
| EL | ELA | 150–500 | 160 | 0.50 | 0.74 |
| EL | ELA | >500 | 264 | 0.57 | 0.72 |
| EL | Mathematics | 11–149 | 246 | 0.30 | 0.72 |
| EL | Mathematics | 150–500 | 161 | 0.56 | 0.82 |
| EL | Mathematics | >500 | 263 | 0.81 | 0.91 |
| ELO | ELA | 11–149 | 277 | 0.05 | 0.49 |
| ELO | ELA | 150–500 | 164 | 0.37 | 0.56 |
| ELO | ELA | >500 | 170 | 0.55 | 0.62 |
| ELO | Mathematics | 11–149 | 276 | 0.29 | 0.72 |
| ELO | Mathematics | 150–500 | 165 | 0.55 | 0.77 |
| ELO | Mathematics | >500 | 169 | 0.67 | 0.76 |
| RFP | ELA | 11–149 | 259 | 0.15 | 0.63 |
| RFP | ELA | 150–500 | 164 | 0.48 | 0.63 |
| RFP | ELA | >500 | 166 | 0.56 | 0.65 |
| RFP | Mathematics | 11–149 | 260 | 0.23 | 0.70 |
| RFP | Mathematics | 150–500 | 165 | 0.72 | 0.84 |
| RFP | Mathematics | >500 | 165 | 0.82 | 0.89 |
| EO | ELA | 11–149 | 262 | 0.31 | 0.62 |
| EO | ELA | 150–500 | 158 | 0.38 | 0.50 |
| EO | ELA | >500 | 392 | 0.60 | 0.64 |
| EO | Mathematics | 11–149 | 262 | 0.24 | 0.48 |
| EO | Mathematics | 150–500 | 158 | 0.50 | 0.57 |
| EO | Mathematics | >500 | 392 | 0.76 | 0.78 |
| SED | ELA | 11–149 | 257 | 0.15 | 0.55 |
| SED | ELA | 150–500 | 165 | 0.37 | 0.55 |
| SED | ELA | >500 | 364 | 0.55 | 0.60 |
| SED | Mathematics | 11–149 | 257 | 0.25 | 0.54 |
| SED | Mathematics | 150–500 | 166 | 0.26 | 0.38 |
| SED | Mathematics | >500 | 363 | 0.69 | 0.71 |
| SWD | ELA | 11–29 | 114 | -0.06 | 0.37 |
| SWD | ELA | 30–149 | 232 | 0.05 | 0.46 |
| SWD | ELA | 150–500 | 189 | 0.48 | 0.59 |
| SWD | ELA | >500 | 133 | 0.57 | 0.63 |
| SWD | Mathematics | 11–29 | 113 | -0.18 | 0.27 |
| SWD | Mathematics | 30–149 | 232 | 0.04 | 0.35 |
| SWD | Mathematics | 150–500 | 189 | 0.47 | 0.55 |
| SWD | Mathematics | >500 | 133 | 0.55 | 0.58 |
| FOS | ELA | ≥11 | 180 | 0.02 | 0.61 |
| FOS | Mathematics | ≥11 | 176 | 0.07 | 0.51 |
| HOM | ELA | 11–29 | 100 | -0.03 | 0.58 |
| HOM | ELA | 30–149 | 149 | 0.09 | 0.53 |
| HOM | ELA | ≥150 | 101 | 0.52 | 0.64 |
| HOM | Mathematics | 11–29 | 100 | 0.04 | 0.49 |
| HOM | Mathematics | 30–149 | 149 | 0.05 | 0.44 |
| HOM | Mathematics | ≥150 | 101 | 0.65 | 0.77 |

1. The size intervals refer to the number of students within LEAs with growth scores in grade levels four through eight in *both* 2017–18 and 2018–19 for the student group and subject of interest (indicated in the first two columns). The number of LEAs in each interval may be smaller than in
2. [Table 6](#table6), given that not all LEAs have estimates in both years.

**6.C.2. Summary of Table 7 Results**

In terms of stability, the hybrid approach improves cross-year stability where it is most needed—for small group sizes—and thus reduces the discrepancies between small and large groups. In fact, for small interval sizes that fall completely below 501, the cross-year stability results for the hybrid approach are identical to those seen for the EBLP approach (presented in Table 5, though note that not all the same size intervals are used in Table 5 and [Table 6](#table6)). Moreover, for intervals above 501, the hybrid approach yields results identical to those seen for the simple average (presented in column 5 of [Table 6](#table6)). Accordingly, cross-year stability results are not substantially impacted by using the hybrid approach in place of the EBLP approach.

**7. Conclusions**

ETS conducted a series of evaluations of AGMs for the CDE. These studies found that reporting the simple averages of individual students within schools, LEAs, and student groups yielded unstable AGMs at both the school and LEA levels, particularly, whenever small samples of students were used to calculate the AGMs (e.g., for small schools or LEAs or for rare student groups). The primary factor contributing to this cross-year instability was the low accuracy of the AGMs under these conditions. In response, ETS developed the EBLP method, which improves the accuracy of AGMs for small schools, LEAs, and rare student groups and, in turn, the stability of the AGMs across years.

Using 2017–18 and 2018–19 data, ETS found that the EBLP improved the stability of AGMs for all students and all student groups at both the school and LEA levels.

**7.A. School-Level Data**

For 2018–19 data, EBLP improved the accuracy of AGMs nearly universally for schools and student groups in schools.

* There are only 69 cases among the 137,055 school-level AGMs (for the All student group as well as the 16 student groups) for which the estimated accuracy ratio comparing the accuracy of EBLP to the simple average is less than 1. That is, the EBLP is estimated to be at least as accurate as the simple average for 99.95 percent of school-level AGMs.
* The results hold for 2018 data as well.

**7.B. LEA-Level Data**

For all LEAs, the EBLPs for the AGMs for the ALL student group were estimated to be more accurate than the simple averages. However, when examining individual student groups, application of the EBLP did not consistently result in more accurate results than the simple average.

Among student groups with larger numbers of students with growth scores, the percentage of LEAs for which the EBLP improved the accuracy of the AGM over the simple average was as low as 3 percent, and the mean accuracy ratio was less than 1.0, indicating that, on average, the simple average is in fact estimated to be more accurate than the EBLP weighted average.

Specifically, across all 16,395 AGMs represented by the 16 student groups other than ALL students, EBLP weighted averages are estimated to be less accurate than the simple average in 17 percent of the cases.

With this in mind, ETS explored the use of a hybrid method in which the EBLP would be used to report AGMs for some LEA student groups and the simple average would be used for others. Based on this exploration, ETS: (1) found evidence supporting the use of a sample-size rule (i.e., the EBLP would be reported when the AGM was calculated with fewer students in a student group and the simple average would be reported otherwise), and (2) developed methods for selecting a single cutoff for the sample-size rule.

Presented with the results of this analysis, the TDG recommended that a cutoff of 500 student growth scores be used when deciding whether to report the EBLP or simple average. Accordingly, the following hybrid reporting plan is recommended for LEAs:

1. For the ALL student group for LEAs with 11 or more students with growth scores, report the EBLP.
2. For individual student-group-level results at the LEA level, report as follows:
3. If the student group has 500 or fewer students with growth scores, report the EBLP for that student group.
4. If the student group has more than 500 students with growth scores, report the simple average for that student group.

Using this reporting plan, the reported AGM was found to be as or more accurate than the simple average for 97 percent of AGMs at the LEAs level. Moreover, AGMs for LEAs remain substantially more stable than those reported using the simple average, and stability is more uniform across LEAs of all sizes.

In sum, application of the EBLP for all schools and a hybrid reporting plan for LEAs provides a means for reporting aggregate student growth that greatly mitigates the concerns raised with reporting simple averages of student growth.

**8. Reference**

Lockwood, J. R., Castellano, K. E., & McCaffrey, D. M. (2020). Improving accuracy and stability of aggregate student growth measures using best linear prediction. (Unpublished manuscript).

**Appendix A: Memorandum for the Technical Design Group****December 2020**

**Evaluating Aggregate Growth Reporting Options for Local Educational Agencies**

J.R. Lockwood, Katherine E. Castellano, & Daniel F. McCaffrey

**Summary**

Educational Testing Service (ETS) and the California Department of Education (CDE) have been evaluating the “Empirical Best Linear Prediction (EBLP)'' method for improving accuracy and stability of aggregate growth measures (AGMs), relative to simple averages of growth scores, for schools and local educational agencies (LEAs). Preliminary findings on the performance of the EBLP method were presented to the Technical Design Group (TDG) in August 2020 (in a briefing and in the August 4 memorandum by Castellano et al., 2020), and in a briefing to the California State Board of Education (SBE) in September 2020.

As reported in the August 4 memorandum to the TDG, application of the EBLP method to growth data from California indicated that the EBLP method improves accuracy and stability for nearly all AGMs that would be reported for schools. Alternatively, we found that while EBLP improves accuracy and stability of AGMs reported at the LEA level when those AGMs are based on small-to-moderate numbers of students, EBLP could reduce accuracy of AGMs relative to simple averages at the LEA level when based on large numbers of students, particularly among AGMs for specific student groups (e.g., students with disabilities). ETS and the CDE recommended, and the TDG and SBE concurred, that an evaluation of options for reporting AGMs for LEAs that mitigated this loss of accuracy was warranted.

This memorandum presents results of this evaluation. Specifically, we evaluated two options for AGM reporting at the LEA level:

1. (the “sample-size rule”) Establishing a sample size threshold such that if an AGM is based on that many, or fewer, growth scores in the reporting year, EBLP would be reported, and otherwise the simple average would be reported; and
2. (the “estimated-accuracy rule”) When the estimated accuracy of the EBLP is greater than the estimated accuracy of the simple average, EBLP would be reported, and otherwise the simple average would be reported.

The key findings are:

* The two decision rules both appear to result in extremely small overall accuracy loss relative to an infeasible “ideal'' decision rule in which the most accurate AGM is always reported. Also, using either rule appears to yield AGMs that are on average more accurate than using EBLP for all AGMs, and substantially more accurate than using the simple average for all AGMs. In this sense, either decision rule would be reasonable.
* For the sample-size rule, values of the threshold in the range of approximately 450 to 700 have extremely similar performance to one another.
* The sample-size rule, with any threshold between 450 and 700, is estimated to have lower overall classification error (i.e., EBLP gets reported when simple average would have been more accurate, or vice versa) than the estimated-accuracy rule.
* The sample-size rule, with any threshold between 450 and 700, is estimated to result in higher overall accuracy than the estimated-accuracy rule.
* The sample-size rule, with any threshold between 450 and 700, is not estimated to be equally effective across all student groups but is substantially more effective than either using simple averages for all student groups, or EBLP weighted averages for all student groups.
* These findings, along with simplicity of description to nontechnical audiences, suggest that the sample-size rule may be the more attractive option for AGM reporting at the LEA level. Given insensitivities of the performance of the sample-size rule to the choice of threshold between 450 and 700, it would be reasonable to use non-statistical criteria such as simplicity of communication to select a specific value.

The remainder of this document provides additional background, technical details on the evaluation methods, and results.

**Background**

Through previous decision processes, the CDE has elected to compute growth scores for individual students using the “residual gain” method, in which a student's growth score in a given year and subject (mathematics or English Language Arts (ELA)) is defined as the difference between their test score in that year and subject, and a linear prediction of that test score based on their mathematics and ELA test scores from the previous school year. Students who score higher than predicted based on their previous-year scores receive positive residual gain scores, and students who score lower than predicted receive negative residual gain scores.[[1]](#footnote-2)

Simple averages of these student growth scores to the school and LEA levels, and for groups of students within these entities, are intended to provide diagnostic information about aggregate achievement progress of policy-relevant student groups. However, in preliminary investigations of these simple averages, ETS and the CDE found that the AGMs tended to have large year-to-year variation for the same school or LEA, creating concerns about the credibility of these AGMs if adopted for annual reporting and accountability use.

One source of annual fluctuations in AGMs for the same school or LEA is inaccuracy due to the fact that the simple averages, in some cases, are based on only a modest number of student growth scores. When the number of student growth scores contributing to the simple average is small, the average can be sensitive to the idiosyncrasies of the specific growth scores included in the average. This can exacerbate year-to-year fluctuations in the AGMs, particularly for schools and LEAs serving smaller numbers of students, and for relatively less populous student groups within schools and LEAs.

**EBLP as a Potential Solution**

ETS developed the empirical best linear prediction (EBLP) method for AGMs to reduce the inaccuracy in the group-level growth measures and potentially reduce fluctuations in the measures for the same school or LEA over time (Lockwood et al., 2020). The EBLP method is not a growth model for generating individual student growth measures. It is a statistical procedure that uses multiple school years of individual student growth data to calculate more accurate group-level growth measures for a given school year. The EBLP for a particular reporting year and subject improves the aggregate growth measures by creating a (approximate) weighted average of student growth measures from two or more school years rather than the simple average for the most recent year (the “reporting year”). Accordingly, we sometimes refer to the EBLP as the “EBLP weighted average” in contrast to the “simple average” of residual gains. The weights for the EBLP weighted average are calculated by a statistical procedure using individual student growth data from the entire state, from multiple school years. The procedure generates a set of weights specific to each AGM - that is, specific to each combination of school or LEA, subject, and student group of interest. In calculating AGMs, the EBLP weighted average tends to give greater weight to data from the “reporting year” and less weight to data from previous school years. How much weight is given to the growth scores from the reporting year for a particular AGM depends on how many growth scores there are in the reporting year for that AGM. The more growth scores there are, the more weight the EBLP places on those scores. Lockwood et al., (2020) provides technical details on the EBLP estimation method. ETS has also developed the “schoolgrowth” package (Lockwood, 2020) for the R statistical computing environment (R Core Team, 2020) to produce EBLPs.

**Business Rules Adopted for AGM Reporting and EBLP Application**

The general EBLP method can combine growth data across tested subjects and across any number of years. In previous investigations of the EBLP method in California, ETS and the CDE considered applying the method using two-year or three-year data blocks (i.e., using either one or two years of data prior to the reporting year), and simultaneously modeling growth data from mathematics and ELA.

On the basis of empirical findings, stakeholder input, and concerns about simplicity of future communications of the method, four business rules have been adopted regarding the reporting of AGMs and application of EBLP in California:

1. AGMs will be reported only when they are based on 11+ student growth scores in the reporting year.
2. The EBLP method will use data from two-year blocks. That is, when computing AGMs for a given reporting year, the EBLP method will use growth scores from that year, and the immediate prior year.
3. The EBLP method will be applied separately to mathematics and ELA growth scores as decided by the California State Board of Education in the November 2020 meeting.
4. EBLP will not be reported for any AGMs that have data from only the reporting year; the simple average will be reported for such cases. For example, if a given LEA has 15 English language learner students with ELA growth scores in the reporting year, and no such students in the prior year, the AGM for that LEA for the ELA growth of its English language learners will be the simple average of the 15 ELA growth scores in the reporting year.

Thus, when computing mathematics AGMs in a given reporting year, the EBLP will combine mathematics growth scores from that year, and mathematics growth scores from the immediate prior year. Analogously, when computing ELA AGMs in a given reporting year, the EBLP will combine ELA growth scores from that year, and ELA growth scores from the immediate prior year. As a result, for example, the mathematics EBLP AGM for a school in a reporting year will be approximately a weighted average of the school's mathematics growth scores in the reporting year and the school's mathematics growth scores from the prior year.

Note that while mathematics and ELA growth scores will be treated separately for the purposes of EBLP computation, the residual gain growth scores for individual students will still use scores from both subjects in the prediction model. For example, the ELA residual gain scores for individual students will be based on a regression model that includes both mathematics and ELA scores for those students from the prior school year, but for the purposes of aggregating ELA growth scores to higher levels, student ELA growth scores will not be combined with student mathematics growth scores.

All results reported in this memorandum are based on an application of the EBLP method that adheres to these business rules. All counts of AGMs reported here include only cases in which there are 11+ student growth scores in the reporting year, and for which there is at least one growth score from the prior year.

**Empirical Results Motivating the Current Study**

This section summarizes the data we analyzed, and the results obtained, that motivated the current study of reporting options for LEA AGMs.

**Data**

Data used in this analysis are mathematics and ELA growth scores from the 2018-2019 school year (“2019”), the 2017-2018 school year (“2018”), and the 2016-2017 school year (“2017”). The data from 2018 and 2019 support the application of the EBLP method for the 2019 reporting year using growth scores from a two-year block. The data from 2017 and 2018 analogously can be used for the 2018 reporting year.

These data are used to compute AGMs at aggregation levels defined by combinations of the following three factors:

1. (organizational level) School vs. LEA
2. (subject) Mathematics vs. ELA
3. (group) Seventeen “student groups”

The 17 student groups include “all” students as well as the 16 groups used in CDE reporting as shown in **Table 8**.

**Table 8. Student Group Abbreviations**

| **Student Group Abbreviation** | **Student Group Name** |
| --- | --- |
| ALL | All |
| AA | Black/African American |
| AI | American Indian or Alaska Native |
| AS | Asian |
| FI | Filipino |
| HI | Hispanic |
| PI | Pacific Islander |
| WH | White |
| MR | Multiple Races/Two or More |
| EL | English Learner |
| ELO | English Learners Only |
| RFP | RFEPs Only |
| EO | English Only |
| SED | Socioeconomically Disadvantaged |
| SWD | Students with Disabilities |
| FOS | Foster Youth |
| HOM | Homeless Youth |

To compute 2019 EBLPs for a given organizational level, subject, and student group, the input data are all student growth scores in the given subject, for students who are members of the given student group, and the links of these students to their appropriate schools or LEAs, for the 2018 and 2019 years. While all these growth scores contribute to the application of the EBLP method, per the business rules previously noted, EBLP AGMs are computed only for cases where at least 11 student growth scores are available in the reporting year. For instance, consider the computation of EBLPs at the school level, for mathematics growth, for the students with disabilities group. Suppose that school A has 10 students with disabilities in 2019 who have math growth scores, and school B has 15 students with disabilities in 2019 who have mathematics growth scores. All 25 of these students are used in the modeling to estimate statistical model parameters required for EBLP, but an EBLP AGM would be computed for only school B because school A has fewer than 11 eligible students in the reporting year.

**Results at the School Level**

Results at the school level strongly support the use of EBLP. EBLP is estimated to improve accuracy for virtually all AGMs that would be reported at the school level. Such a finding was presented in the August memorandum (Castellano et al., 2020), and the results hold after changing the EBLP application to model mathematics and ELA growth separately.

Here we briefly summarize accuracy results at the school level. Stability results were reported in Castellano et al. (2020) and do not change consequentially as a result of modeling mathematics and ELA separately. Stability results for the student groups not presented in Castellano et al. (2020) are all favorable for EBLP. See Appendix Table A1 for the stability results for all 17 student groups using the updated business rules for computing the EBLP (i.e., only using the matched subject scores in each subject’s EBLP weighted average).

For 2019, there are a total of 137,055 school-level AGMs with 11+ students contributing, and for which EBLP is an eligible choice because there are relevant data from 2018. Each one of these 137,055 AGMs is for a combination of a particular school, subject, and student group. For example, a large school may have as many as 34 AGMs based on the two subjects and 17 student groups.

For each AGM, we define the “accuracy ratio” as the mean squared error (MSE) of the simple average divided by the MSE of the EBLP. Values greater than one indicate that the EBLP is more accurate than the simple average, and values less than one indicate that the EBLP is less accurate than the simple average. For each AGM, we estimate the accuracy ratio by plugging estimates of the respective MSEs into this expression. Across all 137,055 AGMs, the mean estimated accuracy ratio is 1.28 and the median is 1.23. Among the 137,055 AGMs, there are only 69 cases for which the estimated accuracy ratio is less than one. That is, the EBLP is estimated to be at least as accurate as the simple average for 99.95% of school-level AGMs. Furthermore, among these 69 cases, the correlation of the EBLP and simple average AGMs is 0.96. Thus, EBLP is estimated to be less accurate than the simple average for only a relative handful of AGMs, and in those cases, the two sets of AGMs are extremely similar.

These findings are replicated using 2017 and 2018 data to compute EBLPs for 2018. Among the 137,484 eligible school-level AGMs for 2018, the mean estimated accuracy ratio is 1.27 and the median is 1.22. Among the 137,484 AGMs, there are only 88 cases for which the estimated accuracy ratio is less than one, so that the EBLP is estimated to be at least as accurate as the simple average for 99.94% of school-level AGMs. Among the 88 cases where the simple average would be preferred, the correlation of the EBLP and simple average AGMs is again 0.96.[[2]](#footnote-3)

These analyses support the conclusion that using EBLP for all school-level AGMs is likely to improve accuracy and stability for almost all cases in each reporting year.

**Results at the LEA Level**

In the August memorandum (Castellano et al., 2020), we reported that at the LEA level, the conclusions about EBLP were less favorable. EBLP improved stability for all LEA AGMs, and improved accuracy for LEA AGMs based on small-to-moderate numbers of students, but EBLP had somewhat less accuracy than simple averages for LEA AGMs based on larger numbers of students. As discussed in more detail in the August memorandum, the likely source of this different behavior of EBLP at the school and LEA levels is the fact that there are so many fewer LEAs than there are schools. The EBLP method requires optimal weights to be estimated from the data, and imprecision in these optimal weights contributes to estimation error for the EBLPs. The optimal weights are estimated more precisely when there are a large number of aggregation units in the model (as with schools) and are estimated less precisely when there are relatively few aggregation units in the model (as with LEAs). There are approximately 10 times as many schools as LEAs in the California data we analyzed. As a result, optimal EBLP weights are estimated less precisely for LEAs than they are for schools. The consequence of this extra imprecision is that for larger LEAs, the EBLP can be less accurate than the simple average, as the former has error due to uncertainty about optimal weights, and the latter already tends to be extremely precise because so many students contribute to it.

These general findings were replicated using two-year EBLPs with separate modeling of the subjects, with one important exception: with the revised procedure, we found that the EBLP AGMs for the “all” student groups at the LEA level were all more accurate than the simple averages. This was true for both mathematics and ELA, and for both the 2018 and 2019 reporting years. This finding is more favorable than what was reported in the August memorandum, where jointly modeling mathematics and ELA in the EBLP procedure resulted in some LEA-level AGMs for the “all” student groups for which the EBLP weighted average was less accurate than the simple average. The likely cause of the current results being more favorable is that when mathematics and ELA are modeled separately, there are many fewer unknown parameters determining the optimal weights that must be estimated from the data. As a result, there is reduced opportunity for uncertainty about the optimal weights to negatively affect the accuracy of the EBLP weighted averages, particularly for AGMs based on many students, as often occurs for AGMs for “all” students.

However, consistent with the findings reported in the August memorandum, the EBLP weighted averages for specific student groups are sometimes estimated to be less accurate than the corresponding simple averages. Specifically, restricting attention to the 16 student groups in Table 1 other than the “all” group, for 2019, there are a total of 16,395 LEA-level AGMs with 11+ students contributing, and for which EBLP is an eligible choice because there are relevant data from 2018. Each one of these 16,395 AGMs is for a combination of a particular LEA, a particular subject, and one of 16 possible student groups. A large LEA may have as many as 32 AGMs based on the two subjects and 16 student groups.

Across all 16,395 AGMs, the mean estimated accuracy ratio is 1.35 and the median is 1.27. However, in contrast to the school-level results, there are 2,787 (17%) cases for which the estimated accuracy ratio is less than one. That is, EBLP is estimated to be at least as accurate as the simple average for only 83% of LEA-level AGMs. Consistent with results reported in Castellano et al. (2020), the cases for which EBLP is estimated to be less accurate than simple average tend to be those where the simple average is based on a relatively large number of students. Specifically, the median number of growth scores contributing to the simple average among cases where EBLP is estimated to be more accurate is 90, whereas the corresponding median among cases where EBLP is estimated to be less accurate is 1,160. The correlation between the EBLP and simple average among the cases where EBLP is estimated to be less accurate is 0.98, so that the estimates are very similar even when EBLP is estimated to be less accurate.

These findings are replicated using 2017 and 2018 data to compute EBLPs for 2018. Among the 16,603 eligible LEA-level AGMs for 2018, there are 2,719 (16.4%) cases for which the estimated accuracy ratio is less than one. As with 2019, these cases tend to be those for which there are a relatively large number of student growth scores contributing to the AGM. Also, like 2019, the correlation between the EBLP and simple average among the cases where EBLP is estimated to be less accurate is high (0.99), so that the estimates are very similar even when EBLP is estimated to be less accurate. Despite this similarity, these analyses support the conclusion that using EBLP for all LEA-level AGMs appears to risk reporting AGMs that are less accurate than possible in each reporting year.

**Evaluating Reporting Options for LEAs**

In light of these findings, the goal of this document is to explore options for reporting LEA-level AGMs that results in the reporting the most accurate AGMs possible for the largest number of cases. As noted in the Summary section, we evaluate the following two decision rules:

1. (the “sample-size rule”) Establishing a sample size threshold such that if an AGM is based on that many, or fewer, growth scores in the reporting year, EBLP would be reported, and otherwise the simple average would be reported; and
2. (the “estimated-accuracy rule”) When the estimated accuracy of the EBLP is greater than the estimated accuracy of the simple average, EBLP would be reported, and otherwise the simple average would be reported.

The goal of the evaluation is to decide which of these two decision rules leads to more efficient reporting of LEA AGMs. By “more efficient” we mean that the more accurate of the two AGMs (EBLP vs simple average) is reported for as many cases as possible, and when the wrong choice is made, the loss of accuracy is as small as possible. The remainder of this section describes challenges to making this evaluation, our evaluation methods, and results.

**Challenges**

Given a real dataset, the EBLP estimation procedure provides, for each AGM, an estimate of the MSE of the simple average, an estimate of the MSE of the EBLP weighted average, and an estimate of the accuracy ratio computed by dividing the estimated MSE of the simple average by the estimated MSE of the EBLP weighted average. The true values of the underlying parameters that these quantities estimate are unknown. As an analogy, suppose there is an infinite population where some quantity has mean M and variance V, and both M and V are unknown. Further suppose that a simple random sample of size n is drawn from this population. Statistical theory yields that the sample mean computed from these n sampled values has mean M, variance V/n, and MSE of V/n. Because V is unknown, the MSE of the sample mean is also unknown. The observed data can be used to compute an unbiased estimator of the MSE using standard procedures, but the fact remains that the true MSE cannot be determined from the observed data, it can only be estimated. The specific value of the estimated MSE generally will vary across hypothetical samples of size n.

The situation with the application of the EBLP method is similar – the observed data (e.g., the growth scores from two consecutive years) can be used only to estimate the MSE of the simple average and the MSE of the EBLP weighted average for each AGM. Furthermore, with application of the EBLP method at the LEA level, the estimated MSEs can be less accurate than they are at the school level, and may have some bias as well, because having fewer aggregation units in the model makes accurate MSE estimation more challenging.

These facts imply that given an observed dataset, it is not clear how to decide which of the sample-size rule or the estimated-accuracy rule is more accurate to use for deciding how to report AGMs at the LEA level. For the sample-size rule, it is possible to pick a sample-size threshold that, for example, minimizes the average of the estimated MSEs, but this may not be the same threshold that would minimize the average of the unknown true MSEs. The estimated-accuracy rule is even more difficult to evaluate – by definition of this rule, for each AGM, EBLP is reported when it is estimated to be more accurate, and the simple average is reported when it is estimated to be more accurate. Thus, internal to the observed data, the estimated-accuracy rule appears to make perfect decisions. However, these decisions may not be perfect because the decisions are based on estimated rather than true MSEs.

**Study Design**

Given these challenges, we use a simulation study to evaluate the performance of the two decision rules. The simulation study can be used to compute estimates of the true MSEs for the simple average and EBLP weighted average for each AGM, which permit evaluation of the performance of the different decision rules when applied to specific realizations of the data and the associated estimated MSEs that can be computed from a given dataset. The model used to simulate data in the simulation study was based on real data to increase the likelihood that the findings from the simulations reflect how the decision rules perform with real data.

Prior to conducting the simulation study, for each of the 16 student groups and each of the 2 subjects (for a total of 32 conditions), we used the appropriate LEA-matched growth scores from 2018 and 2019 to estimate the model parameters required to implement EBLP. Conditional on the number of LEAs and the numbers of students linked to each LEA for a given student group *g* and subject *s*, the key model parameters that affect the MSEs of the EBLP and simple averages are variance-covariance matrices denoted by and in Lockwood et al. (2020), where the subscript “*gs*” emphasizes that these parameters are allowed to vary across student groups and subjects. Roughly speaking, in this context where subjects are modeled separately (e.g., only ELA growth scores are used to implement EBLP for ELA), is the variance-covariance matrix of random effects defined at the level of (LEA grade year) intended to capture deviations in average growth at that level from a given LEA's overall average growth (in subject *s* for students in group *g*) that does not decay to zero as the number of students goes to infinity. Alternatively, is the variance-covariance matrix of student-level residual errors that capture variations among student growth in the same LEA, grade, and year as well as covariances of growth scores from the same student across years, again in subject *s* for students in group *g*. For each student group *g* and subject *s,* denote the values of and estimated from the 2018 and 2019 California growth data by and .

For each student group *g* and subject *s*, the values of and were treated as the true values of these unknown parameters for the simulation study. In addition, for each *g* and *s*, we took the exact set of LEAs, and the exact allocations of students to LEAs, that are observed in the 2018 and 2019 California data, and these features of the data were held constant across simulation replications for group *g* and subject *s*. A single replication of the simulation for group *g* and subject *s* then proceeded as follows. For each LEA, we generated a mean-zero vector of random effects with variance-covariance matrix . The target AGM for each LEA is a weighted average of the elements of this random vector, where the weights depend on how many students are in each grade in 2019 (Lockwood et al., 2020). We then generated mean-zero student-level residual errors with variance-covariance matrix **.** The simulated values of the random effects and student-level residual errors were combined to compute simulated growth scores for each student. The resulting dataset was then used to compute the simple average AGM for each LEA, and also analyzed using the EBLP estimation procedure, which consists of using the observed data to estimate the unknown model parameters, and then using the resulting parameter estimates to compute the EBLP AGM for each LEA. This procedure also provides, for a given simulated dataset, the estimated MSE of the simple average, and estimated MSE of the EBLP weighted average, for each AGM.

For each student group *g* and subject *s*, we conducted 500 independent replications of the simulation. We used these simulation replications to evaluate the performance of a variety of methods for deciding between reporting either the simple average or EBLP weighted average, for each AGM. The following subsection provides details about the evaluation procedures.

**Evaluation Procedures**

The key output of one of the 500 replications of the simulation for group *g* and subject *s*, for each LEA, is

1. The target AGM (as defined above)
2. An estimate of the target AGM using the simple average growth score
3. An estimate of the target AGM using the EBLP weighted average
4. The estimated MSE of the simple average growth score
5. The estimated MSE of the EBLP weighted average
6. The estimated accuracy ratio, defined as the ratio of item (4) to item (5).
7. The decision about whether to report the simple average or EBLP weighted average for each of a variety of decision rules, described below.

Before describing the decision rules, it is important to note that the **true** MSE of the simple average for each AGM is known for the purposes of the simulation because it depends on only , and the counts of students and their allocations to grade levels in a given LEA, both of which are treated as known in the simulation and do not vary across simulation replications. In addition, the **true** MSE of the EBLP weighted average for each AGM can be estimated from the simulation replications by computing, for each AGM, the average across the 500 simulation replications of the squared difference between the EBLP weighted average (#3 above) and the target AGM (#1 above). We define the **true** accuracy ratio as the true MSE of the simple average divided by the true MSE of the EBLP weighted average. For each AGM, these values do not vary across simulation replications, and can be used to evaluate how often a particular decision rule makes the correct decision, with respect to these working true states, in any given simulation replication.

The decision rules that we applied to each simulated dataset are as follows:

1. (“simple average”) Report the simple average for all AGMs
2. (“EBLP”) Report the EBLP weighted average for all AGMs
3. (“sample size rule with ”) Report the EBLP weighted average if the number of students in the reporting year (here, 2019) for a particular AGM is less than or equal to a threshold , and otherwise report the simple average. The number is defined as the sample size that minimizes the average **estimated** MSE of all of the assigned AGMs across all LEAs, all 16 student groups, and both subjects. Specifically, in a given simulation replication, each of the 16,395 AGMs (aggregating across student groups and subjects) has an estimated MSE of the simple average and an estimated MSE of the EBLP weighted average based on calculations applied to the simulated data from that simulation replication. Each AGM also has a fixed number of students with growth scores in the reporting year. If the rule was to report the EBLP weighted average when the number of students was less than or equal to, for example, 100, and otherwise report the simple average, the estimated overall error of this decision would be the average of 16,395 numbers. The number for a given AGM would be the estimated MSE of the EBLP weighted average when its reporting-year sample size was less than or equal to 100, and would be the estimated MSE of the simple average when its reporting-year sample size was greater than 100. The specific value for this average estimated MSE of the AGMs that get reported will vary across different values of the sample-size threshold (100 in the example). The number is defined as the threshold that minimizes the average estimated MSE.
4. (“sample size rule with plus bias correction ”) Report the EBLP weighted average if the number of students in the reporting year (here, 2019) for a particular AGM is less than or equal to a threshold , and otherwise report the simple average. The number is as defined above, and the number is an adjustment that might improve by correcting for potential bias in . Because the estimated MSEs can be noisy and potentially biased, we thought could be error-prone and possibly biased so that it might be systematically too small or too large. By adding a constant to we might be able improve the accuracy of the resulting decision rule. Based on preliminary simulation analyses we expected might be too small, but we considered a range of positive and negative values for the bias correction. Specifically, we considered -200, -100, 100, 150, 200, and 250 as possible values of , for a total of 6 decision rules of this kind.
5. (“estimated-accuracy rule”) Report the EBLP is the **estimated** MSE for the EBLP is smaller than the **estimated** MSE for the simple average; otherwise report the simple average.

Thus, a total of 10 decision rules were applied to each simulated dataset. The performance of each these rules was evaluated against what we call the “optimal rule”: report the EBLP if the **true** MSE for the EBLP is smaller than the **true** MSE for the simple average; otherwise report the simple average. This rule yields the AGMs with the smallest MSEs, but it is not feasible with real data because it relies on the true values that are known only in the simulation. We include this rule as a benchmark to assess the loss of accuracy from making decisions using the 10 feasible decision rules.

For each of the 10 decision rules, we consider the following evaluation criteria.

1. “Probability of misclassification”: For a given simulated dataset, a given decision rule reports either the EBLP weighted average or the simple average for each AGM. Some of these decisions are correct in the sense that the decision corresponds to the one that would be made by the optimal rule using true MSEs. Others are incorrect either because they selected the simple average when the EBLP weighted average has smaller true MSE, or because they selected the EBLP weighted average when the simple average has the smaller true MSE. The misclassification rate in a given simulated dataset, for a given decision rule, is the fraction of the total AGMs for which the wrong estimate is selected by the decision rule. The average of the misclassification rate across the 500 simulation replications provides the estimated probability of misclassification for the decision rule. For example, if the probability of misclassification for a given decision rule is 0.10, it means that on average, across hypothetical realizations of the data, we expect the decision rule to make the wrong decision for 10% of AGMs.
2. “Percentage excess MSE”: We define the “average optimal MSE” as the average, across all AGMs, of the true MSE for the optimal decision rule. This is a measure of overall error for the optimal rule. For a given simulated dataset, when a decision rule selects the wrong estimate, it incurs excess MSE relative to the optimal decision in the sense that for a particular AGM, it assigns an estimate whose true MSE is larger than the MSE for the optimal decision. Alternatively, when a decision rule selects the optimal estimate, it incurs no excess MSE. We define the “average assigned MSE” for a given simulated dataset and decision rule as the average, across all AGMs, of the true MSE corresponding to the estimator that was assigned by the decision rule. By definition, the average assigned MSE is greater than or equal to the average optimal MSE. For a given simulated dataset and decision rule, we compute 100\*(average assigned MSE - average optimal MSE)/(average optimal MSE), which expresses the average excess MSE as a percentage of the average optimal MSE. We define the “percentage excess MSE” for a given decision rule as the average of this quantity across the 500 simulation replications. For example, if the percentage excess MSE for a given decision rule is 5%, it means that on average, across hypothetical realizations of the data, we expect the decision rule to make assignments whose true average MSE across all AGMs is 5% larger than the true average MSE of the optimal decision rule.
3. “Average correlation for misclassified AGMs”: For a given simulated dataset and decision rule, the AGMs for which the decision rule makes the wrong decision relative to the optimal decision rule can be identified. For this subset of AGMs, we compute the correlation between the EBLP weighted averages and the simple averages. This can be considered an indicator of the impact of a decision rule making a wrong decision. When the correlation is high, it suggests that the impact of making the wrong decision is not large in the sense that the two sets of estimates are highly correlated. For each decision rule, we compute the average of this correlation across the 500 simulation replications.

**Overall Results**

**Table 9** provides the overall results of the evaluation of the decision rules using the simulation study pooled over all student groups. That is, over all 16,395 AGMs from each of the 500 replications of the simulation. The decision rules are rank ordered in the table by their average rank for the first two criteria—probability misclassified and percentage excess MSE—using unrounded values. The table shows that *N\** plus 150, 100, 200, 250, or 0 have almost identical results when rounded to two decimal places across the three criteria, indicating that any choice of bias correction *B* in the range of 0 to 250 would be defensible and could be chosen more based on secondary policy and practical consideration than statistical ones. For all five of these decision rules, based on the simulation, there is only a 10% probability of misclassification of any of the LEAs’ AGMs, the assigned MSE is only 0.10-0.11% larger on average than the optimal MSE, and the simple average and EBLP estimates among misclassified cases are strongly related with correlations of 0.98. The decision rules of *N*\*-100, *N*\*-200, the estimated accuracy ratio, and using EBLP only result in a somewhat larger probability of misclassification, larger percentage excess MSE, and lower correlation between AGM estimates. The simple average rule in which every AGM is assigned the simple average performs demonstrably worse than any of the other decision rules with, for example, a 77% chance of misclassification. Accordingly, the simulation indicates that using a sample size threshold rule of *N*\* plus 0 to 250 would be better than using the estimated accuracy ratio rule or always assigning the EBLP or simple average.

**Table 9. Overall Results**

| **Decision Rule** | **Probability Misclassified** | **Percentage Excess MSE** | **Correlation between Simple Average and EBLP estimates among Misclassified** |
| --- | --- | --- | --- |
| N\* + 150 | 0.10 | 0.10% | 0.98 |
| N\* + 200 | 0.10 | 0.10% | 0.98 |
| N\* + 100 | 0.10 | 0.10% | 0.98 |
| N\* + 250 | 0.10 | 0.10% | 0.98 |
| N\* + 0 | 0.10 | 0.11% | 0.98 |
| N\* + -100 | 0.12 | 0.17% | 0.97 |
| Estimated Accuracy Ratio | 0.12 | 0.20% | 0.96 |
| N\* + -200 | 0.15 | 0.32% | 0.97 |
| EBLP | 0.23 | 0.33% | 0.98 |
| Simple Average | 0.77 | 73.66% | 0.84 |

We considered sample-size rules that shifted the value of up or down by B due to the concern that because is selected on the basis of estimated MSEs, random or systematic errors in these estimated MSEs could lead to bias in . The simulation suggests that this concern was reasonable because, as shown in Table 9, shifting upward leads to a small but consistent improvement in both probability of misclassification and percentage excess MSE relative to no shift. Investigations into the simulation results indicate that the likely cause of this result is that estimated MSEs for the EBLP weighted averages have a small positive bias for true MSEs for some AGMs. This bias is most pronounced for AGMs based on many students, and AGMs for the least populous student groups. While the bias is small in absolute terms and is small relative to the sampling variability of the estimated MSEs, it is large enough to cause to be negatively biased. Specifically, if estimated EBLPs MSEs for AGMs based on many students are positively biased, it has the effect of pushing downward because the apparent excess MSE of assigning the EBLP to AGMs based on many students is larger than the true excess MSE. As a result, adding a positive constant to results in a decision rule with modestly improved performance. This also explains why shifting downward (using a negative value of B) further degrades performance of the sample-size rule – it takes an already-negatively-biased estimate and makes it even more negatively biased.

**Results by Student Group**

The overall results in **Table 9** pool over all student groups. To explore implications for individual student groups, we also consider results separately by student group. **Table 10** to **Table 12** provide the results for each of the three evaluation criteria for four of the decision rules: N\*+ bias correction of 150, N\* + 0 (no bias correction), EBLP only, and simple average only. We include the N\*+150 rule as it was ranked highest overall in **Table 9**, albeit with very little distinction in performance than adding 100, 200, or 250 to N\* instead. We compare the results of this rule against the more straightforward approaches of using the sample size threshold rule without a bias correction, always assigning EBLP, or always assigning the simple average.

As shown in **Table 10**, the probability of misclassification (averaged over the 500 replications) for the two sample size rules for the student groups are similar: N\*+150 ranges from 0.00 to .25 with an average of .08 and N\*+0 ranges from 0.00 to 0.28 with an average of 0.08. All chances of misclassification are within five percentage points of each other. For all but two groups, the chance of misclassification is smaller using the sample size rules than assigning the EBLP to all AGMs. The EO student group for mathematics has the largest chance of misclassification at 25% for the N\*+150 rule and 28% for the N\*+0 rule, but these rates are only slightly larger than the 23% chance of misclassification if EBLP is always assigned. For the RFP student group for mathematics, only the N\*+0 rule results in a slightly higher misclassification rate than EBLP only rule: 14% vs 13%. The EBLP rule results in 0 to 40% chance of misclassification (average of 18%), whereas the simple average rule results in a 60 to 100% chance of misclassification (average of 82%). Consistent with the overall results, the sample size rule substantially reduces the probability of misclassification over using the simple average rule and is generally a moderate improvement over using the EBLP rule.

The percentage excess MSE results, as shown in **Table 11**, demonstrate a similar story. For the N\*+150 rule, the average assigned MSE is 0% to 0.35% the size of the optimal MSE with an average of 0.11%. For the N\*+0 rule, the average assigned MSE falls over a larger range of 0% to 0.66%, but it has a similar average of 0.12%. On average, the absolute difference in this criterion for the two sample size rules is 0.06% (with the largest difference being 0.36%). As with the misclassification probability, the EO student group is the only one for which always assigning the EBLP is slightly better than using the N\*+150 rule. But the N\*+0 rule results in four cases for which assigning EBLP exclusively has smaller percentage excess MSE by 0.03% to 0.17% for the White (WH), English Learner (EL), RFP, and EO groups all for mathematics. The EBLP only rule results in 0% to 1.27% average excess MSE (average = 0.35%) and the simple average only rule results in 41.92% to 127.03% average excess MSE (average=75.61%).

Lastly, we consider the correlation between the EBLP and simple average estimates among the misclassified cases (averaged over the 500 replications). For this evaluation criterion, as shown in **Table 12**, there is little distinction between using the N\*+150 rule, N\*+0 rule, or the EBLP only rule. For all three cases, the correlations range from about 0.87 to 1.00 with an average of 0.95. These correlation values for individual student groups generally are within +/-0.01 of each other. Not surprisingly, the results for the simple average rule, in contrast, are noticeably worse with correlations in the 0.75 to 0.94 range (average of 0.83), indicating that cases for which the EBLP should have been assigned instead of the simple average, the simple average and EBLP estimates are not as strongly related.

**Table 10. Probability Misclassification by Student Group**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Student Group** | **Subject** | **N\* + 150** | **N\* + 0** | **EBLP** | **Simple Average** |
| AA | ELA | 0.04 | 0.03 | 0.13 | 0.87 |
| AA | Mathematics | 0.02 | 0.03 | 0.09 | 0.91 |
| AI | ELA | 0.00 | 0.00 | 0.00 | 1.00 |
| AI | Mathematics | 0.00 | 0.00 | 0.00 | 1.00 |
| AS | ELA | 0.17 | 0.13 | 0.34 | 0.66 |
| AS | Mathematics | 0.09 | 0.06 | 0.26 | 0.74 |
| FI | ELA | 0.04 | 0.02 | 0.08 | 0.92 |
| FI | Mathematics | 0.10 | 0.08 | 0.13 | 0.87 |
| HI | ELA | 0.04 | 0.05 | 0.40 | 0.60 |
| HI | Mathematics | 0.09 | 0.11 | 0.35 | 0.65 |
| PI | ELA | 0.03 | 0.02 | 0.03 | 0.97 |
| PI | Mathematics | 0.02 | 0.01 | 0.02 | 0.98 |
| WH | ELA | 0.08 | 0.06 | 0.30 | 0.70 |
| WH | Mathematics | 0.17 | 0.19 | 0.19 | 0.81 |
| MR | ELA | 0.03 | 0.03 | 0.06 | 0.94 |
| MR | Mathematics | 0.16 | 0.15 | 0.19 | 0.81 |
| EL | ELA | 0.07 | 0.11 | 0.28 | 0.72 |
| EL | Mathematics | 0.12 | 0.17 | 0.22 | 0.78 |
| ELO | ELA | 0.09 | 0.07 | 0.28 | 0.72 |
| ELO | Mathematics | 0.04 | 0.07 | 0.21 | 0.79 |
| RFP | ELA | 0.08 | 0.05 | 0.29 | 0.71 |
| RFP | Mathematics | 0.09 | 0.14 | 0.13 | 0.87 |
| EO | ELA | 0.13 | 0.14 | 0.37 | 0.63 |
| EO | Mathematics | 0.25 | 0.28 | 0.23 | 0.77 |
| SED | ELA | 0.14 | 0.18 | 0.28 | 0.72 |
| SED | Mathematics | 0.14 | 0.16 | 0.32 | 0.68 |
| SWD | ELA | 0.11 | 0.10 | 0.22 | 0.78 |
| SWD | Mathematics | 0.13 | 0.12 | 0.21 | 0.79 |
| FOS | ELA | 0.00 | 0.00 | 0.01 | 0.99 |
| FOS | Mathematics | 0.00 | 0.00 | 0.01 | 0.99 |
| HOM | ELA | 0.08 | 0.05 | 0.14 | 0.86 |
| HOM | Mathematics | 0.03 | 0.02 | 0.09 | 0.91 |

**Table 11. Percentage Excess MSE by Student Group1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Student Group** | **Subject** | **N\* + 150** | **N\* + 0** | **EBLP** | **Simple Average** |
| AA | ELA | 0.04 | 0.03 | 0.24 | 74.71 |
| AA | Mathematics | 0.01 | 0.04 | 0.16 | 104.79 |
| AI | ELA | 0.00 | 0.00 | 0.00 | 113.25 |
| AI | Mathematics | 0.00 | 0.00 | 0.00 | 127.03 |
| AS | ELA | 0.35 | 0.27 | 0.91 | 53.14 |
| AS | Mathematics | 0.20 | 0.10 | 1.12 | 72.07 |
| FI | ELA | 0.06 | 0.02 | 0.16 | 79.11 |
| FI | Mathematics | 0.30 | 0.18 | 0.47 | 77.22 |
| HI | ELA | 0.03 | 0.06 | 1.27 | 73.08 |
| HI | Mathematics | 0.07 | 0.11 | 0.31 | 68.56 |
| PI | ELA | 0.05 | 0.04 | 0.05 | 68.48 |
| PI | Mathematics | 0.03 | 0.02 | 0.03 | 99.63 |
| WH | ELA | 0.11 | 0.06 | 0.63 | 62.60 |
| WH | Mathematics | 0.11 | 0.14 | 0.11 | 49.20 |
| MR | ELA | 0.02 | 0.01 | 0.08 | 75.06 |
| MR | Mathematics | 0.34 | 0.28 | 0.46 | 57.84 |
| EL | ELA | 0.09 | 0.22 | 0.77 | 90.40 |
| EL | Mathematics | 0.30 | 0.66 | 0.61 | 112.86 |
| ELO | ELA | 0.12 | 0.07 | 0.47 | 68.30 |
| ELO | Mathematics | 0.04 | 0.09 | 0.60 | 96.02 |
| RFP | ELA | 0.12 | 0.06 | 0.75 | 73.35 |
| RFP | Mathematics | 0.17 | 0.43 | 0.26 | 101.94 |
| EO | ELA | 0.13 | 0.16 | 0.38 | 52.52 |
| EO | Mathematics | 0.17 | 0.24 | 0.11 | 41.92 |
| SED | ELA | 0.14 | 0.24 | 0.26 | 53.75 |
| SED | Mathematics | 0.11 | 0.14 | 0.20 | 47.29 |
| SWD | ELA | 0.10 | 0.09 | 0.22 | 57.43 |
| SWD | Mathematics | 0.12 | 0.11 | 0.19 | 49.92 |
| FOS | ELA | 0.00 | 0.00 | 0.00 | 95.49 |
| FOS | Mathematics | 0.00 | 0.00 | 0.01 | 105.08 |
| HOM | ELA | 0.08 | 0.04 | 0.21 | 55.26 |
| HOM | Mathematics | 0.03 | 0.01 | 0.15 | 62.23 |

1. All values are expressed as percentages. For example, “0.04” in the first cell is “0.04%”.

**Table 12. Correlation between the Simple Average and EBLP estimates among Misclassified Cases by Student Group**

| **Student Group** | **Subject** | **N\* + 1501** | **N\* + 01** | **EBLP1** | **Simple Average** |
| --- | --- | --- | --- | --- | --- |
| AA | ELA | 0.94 | 0.94 | 0.96 | 0.83 |
| AA | Mathematics | 0.87 | 0.88 | 0.90 | 0.78 |
| AI | ELA | NA | NA | NA | 0.76 |
| AI | Mathematics | NA | NA | NA | 0.75 |
| AS | ELA | 0.97 | 0.97 | 0.97 | 0.87 |
| AS | Mathematics | 0.92 | 0.93 | 0.91 | 0.82 |
| FI | ELA | 0.91 | 0.91 | 0.92 | 0.80 |
| FI | Mathematics | 0.88 | 0.88 | 0.88 | 0.80 |
| HI | ELA | 0.96 | 0.96 | 0.97 | 0.82 |
| HI | Mathematics | 0.99 | 0.99 | 0.99 | 0.86 |
| PI | ELA | 0.93 | 0.92 | 0.93 | 0.84 |
| PI | Mathematics | NA | NA | NA | 0.78 |
| WH | ELA | 0.97 | 0.97 | 0.98 | 0.86 |
| WH | Mathematics | 0.99 | 0.99 | 1.00 | 0.92 |
| MR | ELA | 0.92 | 0.91 | 0.93 | 0.82 |
| MR | Mathematics | 0.95 | 0.95 | 0.95 | 0.86 |
| EL | ELA | 0.95 | 0.94 | 0.95 | 0.78 |
| EL | Mathematics | 0.92 | 0.91 | 0.91 | 0.75 |
| ELO | ELA | 0.97 | 0.97 | 0.98 | 0.84 |
| ELO | Mathematics | 0.91 | 0.90 | 0.87 | 0.76 |
| RFP | ELA | 0.94 | 0.93 | 0.95 | 0.81 |
| RFP | Mathematics | 0.93 | 0.91 | 0.93 | 0.78 |
| EO | ELA | 0.99 | 0.99 | 0.99 | 0.90 |
| EO | Mathematics | 1.00 | 1.00 | 1.00 | 0.94 |
| SED | ELA | 0.99 | 0.99 | 0.99 | 0.90 |
| SED | Mathematics | 1.00 | 1.00 | 1.00 | 0.92 |
| SWD | ELA | 0.98 | 0.98 | 0.99 | 0.88 |
| SWD | Mathematics | 0.99 | 0.99 | 0.99 | 0.91 |
| FOS | ELA | NA | NA | NA | 0.76 |
| FOS | Mathematics | NA | NA | NA | 0.76 |
| HOM | ELA | 0.97 | 0.96 | 0.97 | 0.87 |
| HOM | Mathematics | 0.96 | 0.97 | 0.97 | 0.87 |

1. A value of “NA” for “not applicable” or “not defined” is used whenever there were 0 or near 0 cases in which an LEA’s AGM was misclassified by the rule of interest. In those cases, there are no or too few Simple Average and EBLP estimates to correlate.

**Conclusions**

The analysis of real growth data from California suggests that for LEA-level reporting of AGMs for student groups, the EBLP weighted average is often, but not always, more accurate than the simple average. The results of the simulation study are consistent with this empirical finding: reporting the EBLP weighted average for all AGMs, or reporting the simple average for all AGMs, each appears to be inferior to decision rules that choose to report one or the other in some principled manner. In absolute terms, with respect to accuracy, reporting the simple average for all AGMs is vastly inferior to all other decision rules. This is not surprising given the preponderance of small-to-modest reporting-year sample sizes among LEA-level AGMs for student groups, particularly for less populous student groups. The EBLP procedure is designed to improve accuracy for these cases, and does so. Both the sample-size and estimated-accuracy decision rules are highly likely to report the EBLP weighted average for such cases, and thus provide large improvements compared to reporting the simple average for all AGMs. They also provide modest improvements over assigning the EBLP weighted average for all AGMs.

The sample-size rule, with various choices of sample-size cutoff, appears to perform somewhat better than the estimated-accuracy rule. This finding is perhaps surprising given that the estimated-accuracy rule makes its decisions using direct estimates of the quantities we would like to operate on (the true, rather than estimated, MSEs), whereas the sample-size rule makes its decisions using sample size as a proxy for which estimator might be preferred. Investigations into the simulation results indicated that the estimated-accuracy rule suffers from sampling errors in the estimated accuracy ratio for individual AGMs, which can be nontrivial for some AGMs, particularly for the less-populous student groups. Sampling variability of the estimated MSE for the EBLP weighted averages drives this uncertainty, because accurately estimating these MSEs with relatively few aggregation units (here, LEAs), and when most aggregation units have only modest numbers of students for particular groups, is difficult. As a result, the estimated-accuracy rule tends to make somewhat more incorrect decisions than the sample-size rule, and the cost of these wrong decisions, in terms of accuracy, tends to be somewhat higher.

The results overall suggest that the sample-size rule may be a reasonable choice. Our method for selecting the sample-size threshold from the observed data was to compute a number , defined as the threshold that minimizes the average estimated MSE of the numbers reported for each AGM, and then possibly shifting up or down to determine the actual threshold used for the sample-size rule. The value of obtained from the actual data is 446 using 2019 as the reporting year, and 447 using 2018 as the reporting year. The consistency of the two values provides some assurance that the value is not highly sensitive to the vagaries of growth data that may exist in any particular year. The overall simulation results in Table 2 indicate that sample-size rules using a threshold obtained by shifting up by a value anywhere between 0 and 250 result in extremely similar performance. This suggests that thresholds for the sample-size rule in the range of 450 to 700 (obtained by shifting the approximate value of 450 for up by a value anywhere between 0 and 250) would function similarly.

The subgroup results shown in Tables 3-5 indicate that such a rule would also function reasonably for most student groups, though admittedly a rule tailored to each specific group (e.g., separate sample-size thresholds for each of the 32 combinations of student group and subject) might have slightly better performance. An evaluation of such rules was outside the scope of this analysis, and any benefits that such rules might provide over a common sample-size threshold would have to be weighed against the extra complexity of explaining the procedures to various stakeholders.

**References**

Castellano, K. E., Lockwood, J. R., & McCaffrey, D. F. (2020). Exploring empirical best linear prediction for aggregate growth measures. Memorandum to the CDE Technical Design Group, August 4.

Lockwood, J. R. (2020). *Schoolgrowth: Functions for improving accuracy of school-level growth measures using Empirical Best Linear Prediction.* R package version 0.6-3.

Lockwood, J. R., Castellano, K. E., & McCaffrey, D. F. (2020). Improving accuracy and stability of aggregate student growth measures using best linear prediction. Unpublished manuscript.

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

**Appendix A1: School-Level Stability and Accuracy Results**

**Table A1. Cross Year Stability of the 2018–19 Two-Year EBLP Weighted Average and the Simple Average for Schools**

| **Student Group** | **Subject** | **2018–19 School Size1** | **Number of Schools** | **Correlation between the 2017–18 and 2018–19 Simple Average** | **Correlation between the 2017–18 and 2018–19 2-‍year EBLP** |
| --- | --- | --- | --- | --- | --- |
| ALL | ELA | 11–29 | 154 | 0.32 | 0.65 |
| ALL | ELA | 30–149 | 2142 | 0.30 | 0.47 |
| ALL | ELA | ≥150 | 4989 | 0.42 | 0.49 |
| ALL | Mathematics | 11–29 | 148 | 0.38 | 0.66 |
| ALL | Mathematics | 30–149 | 2146 | 0.41 | 0.53 |
| ALL | Mathematics | ≥150 | 4986 | 0.58 | 0.62 |
| AA | ELA | 11–29 | 1245 | 0.16 | 0.58 |
| AA | ELA | ≥30 | 955 | 0.28 | 0.56 |
| AA | Mathematics | 11–29 | 1244 | 0.20 | 0.61 |
| AA | Mathematics | ≥30 | 955 | 0.32 | 0.60 |
| AI | ELA | ≥11 | 93 | 0.12 | 0.58 |
| AI | Mathematics | ≥11 | 94 | 0.05 | 0.59 |
| AS | ELA | 11–29 | 1124 | 0.07 | 0.51 |
| AS | ELA | 30–149 | 1211 | 0.30 | 0.53 |
| AS | ELA | ≥150 | 279 | 0.46 | 0.55 |
| AS | Mathematics | 11–29 | 1124 | 0.29 | 0.61 |
| AS | Mathematics | 30–149 | 1212 | 0.52 | 0.69 |
| AS | Mathematics | ≥150 | 278 | 0.73 | 0.77 |
| FI | ELA | 11–29 | 617 | 0.03 | 0.45 |
| FI | ELA | ≥30 | 368 | 0.33 | 0.53 |
| FI | Mathematics | 11–29 | 617 | 0.30 | 0.61 |
| FI | Mathematics | ≥30 | 368 | 0.48 | 0.68 |
| HI | ELA | 11–29 | 764 | 0.17 | 0.56 |
| HI | ELA | 30–149 | 3453 | 0.27 | 0.49 |
| HI | ELA | ≥150 | 2631 | 0.40 | 0.48 |
| HI | Mathematics | 11–29 | 767 | 0.25 | 0.58 |
| HI | Mathematics | 30–149 | 3453 | 0.36 | 0.52 |
| HI | Mathematics | ≥150 | 2630 | 0.52 | 0.58 |
| PI | ELA | 11–149 | 86 | 0.08 | 0.40 |
| PI | Mathematics | 11–149 | 86 | 0.19 | 0.69 |
| WH | ELA | 11–29 | 1216 | 0.12 | 0.51 |
| WH | ELA | 30–149 | 2754 | 0.29 | 0.50 |
| WH | ELA | ≥150 | 775 | 0.43 | 0.49 |
| WH | Mathematics | 11–29 | 1218 | 0.18 | 0.49 |
| WH | Mathematics | 30–149 | 2752 | 0.41 | 0.55 |
| WH | Mathematics | ≥150 | 774 | 0.61 | 0.65 |
| MR | ELA | 11–29 | 1323 | 0.11 | 0.53 |
| MR | ELA | ≥30 | 610 | 0.21 | 0.51 |
| MR | Mathematics | 11–29 | 1324 | 0.30 | 0.61 |
| MR | Mathematics | ≥30 | 608 | 0.47 | 0.66 |
| EL | ELA | 11–29 | 1091 | 0.15 | 0.57 |
| EL | ELA | 30–149 | 3913 | 0.28 | 0.52 |
| EL | ELA | ≥150 | 1350 | 0.43 | 0.52 |
| EL | Mathematics | 11–29 | 1087 | 0.37 | 0.68 |
| EL | Mathematics | 30–149 | 3916 | 0.45 | 0.62 |
| EL | Mathematics | ≥150 | 1348 | 0.57 | 0.64 |
| ELO | ELA | 11–29 | 1620 | 0.06 | 0.48 |
| ELO | ELA | 30–149 | 3302 | 0.28 | 0.51 |
| ELO | ELA | ≥150 | 324 | 0.32 | 0.44 |
| ELO | Mathematics | 11–29 | 1621 | 0.24 | 0.60 |
| ELO | Mathematics | 30–149 | 3301 | 0.35 | 0.57 |
| ELO | Mathematics | ≥150 | 321 | 0.40 | 0.51 |
| RFP | ELA | 11–29 | 1854 | 0.14 | 0.53 |
| RFP | ELA | 30–149 | 3241 | 0.32 | 0.55 |
| RFP | ELA | ≥150 | 447 | 0.43 | 0.52 |
| RFP | Mathematics | 11–29 | 1853 | 0.34 | 0.63 |
| RFP | Mathematics | 30–149 | 3243 | 0.50 | 0.67 |
| RFP | Mathematics | ≥150 | 445 | 0.62 | 0.69 |
| EO | ELA | 11–29 | 515 | 0.24 | 0.60 |
| EO | ELA | 30–149 | 4145 | 0.29 | 0.49 |
| EO | ELA | ≥150 | 2519 | 0.50 | 0.56 |
| EO | Mathematics | 11–29 | 515 | 0.26 | 0.56 |
| EO | Mathematics | 30–149 | 4153 | 0.42 | 0.56 |
| EO | Mathematics | ≥150 | 2510 | 0.64 | 0.68 |
| SED | ELA | 11–29 | 530 | 0.13 | 0.54 |
| SED | ELA | 30–149 | 3423 | 0.27 | 0.46 |
| SED | ELA | ≥150 | 3067 | 0.40 | 0.48 |
| SED | Mathematics | 11–29 | 532 | 0.23 | 0.54 |
| SED | Mathematics | 30–149 | 3428 | 0.37 | 0.51 |
| SED | Mathematics | ≥150 | 3060 | 0.53 | 0.58 |
| SWD | ELA | 11–29 | 3205 | 0.04 | 0.42 |
| SWD | ELA | ≥30 | 3025 | 0.19 | 0.44 |
| SWD | Mathematics | 11–29 | 3210 | 0.04 | 0.36 |
| SWD | Mathematics | ≥30 | 3019 | 0.16 | 0.36 |
| FOS | ELA | 11–29 | 29 | -0.02 | 0.35 |
| FOS | Mathematics | 11–29 | 29 | -0.35 | 0.11 |
| HOM | ELA | 11–29 | 767 | 0.21 | 0.56 |
| HOM | ELA | ≥30 | 617 | 0.21 | 0.44 |
| HOM | Mathematics | 11–29 | 768 | 0.30 | 0.61 |
| HOM | Mathematics | ≥30 | 614 | 0.38 | 0.59 |

1. The size bins refer to the number of students within schools with growth scores in grades four through eight in *both* 2017–18 and 2018–19 for the student group and subject of interest (indicated in the first two columns). The number of schools in each bin is smaller than in Table A2– given that not all schools have estimates in both years.

**Table A2. Improvement in Accuracy of Growth Estimates Using Two-Year EBLP Weighted Averages Versus Simple Averages at the School Level**

| **Student Group** | **Subject** | **2018–19 School Size1** | **Number of Schools** | **Mean Accuracy Ratio for 2-‍year EBLP vs. Simple Average** | **Percentage of Schools with Improved Accuracy for 2-‍year EBLP vs. Simple Average** |
| --- | --- | --- | --- | --- | --- |
| ALL | ELA | 11–29 | 186 | 1.56 | 100% |
| ALL | ELA | 30–149 | 2172 | 1.18 | 100% |
| ALL | ELA | ≥150 | 5019 | 1.08 | 100% |
| ALL | Mathematics | 11–29 | 179 | 1.49 | 100% |
| ALL | Mathematics | 30–149 | 2176 | 1.14 | 100% |
| ALL | Mathematics | ≥150 | 5016 | 1.06 | 100% |
| AA | ELA | 11–29 | 1421 | 1.67 | 100% |
| AA | ELA | ≥30 | 967 | 1.34 | 99% |
| AA | Mathematics | 11–29 | 1417 | 1.72 | 100% |
| AA | Mathematics | ≥30 | 967 | 1.41 | 99% |
| AI | ELA | ≥11 | 113 | 1.66 | 99% |
| AI | Mathematics | ≥11 | 113 | 1.84 | 98% |
| AS | ELA | 11–29 | 1300 | 1.61 | 100% |
| AS | ELA | 30–149 | 1219 | 1.28 | 100% |
| AS | ELA | ≥150 | 279 | 1.05 | 92% |
| AS | Mathematics | 11–29 | 1296 | 1.56 | 100% |
| AS | Mathematics | 30–149 | 1220 | 1.27 | 100% |
| AS | Mathematics | ≥150 | 278 | 1.07 | 95% |
| FI | ELA | 11–29 | 816 | 1.52 | 100% |
| FI | ELA | ≥30 | 372 | 1.28 | 100% |
| FI | Mathematics | 11–29 | 817 | 1.51 | 100% |
| FI | Mathematics | ≥30 | 372 | 1.26 | 100% |
| HI | ELA | 11–29 | 847 | 1.61 | 100% |
| HI | ELA | 30–149 | 3478 | 1.26 | 100% |
| HI | ELA | ≥150 | 2645 | 1.09 | 100% |
| HI | Mathematics | 11–29 | 848 | 1.53 | 100% |
| HI | Mathematics | 30–149 | 3478 | 1.21 | 100% |
| HI | Mathematics | ≥150 | 2644 | 1.08 | 100% |
| PI | ELA | 11–149 | 114 | 1.56 | 100% |
| PI | Mathematics | 11–149 | 113 | 1.90 | 100% |
| WH | ELA | 11–29 | 1395 | 1.55 | 100% |
| WH | ELA | 30–149 | 2778 | 1.24 | 100% |
| WH | ELA | ≥150 | 782 | 1.07 | 100% |
| WH | Mathematics | 11–29 | 1398 | 1.43 | 100% |
| WH | Mathematics | 30–149 | 2776 | 1.17 | 100% |
| WH | Mathematics | ≥150 | 781 | 1.06 | 100% |
| MR | ELA | 11–29 | 1600 | 1.59 | 100% |
| MR | ELA | ≥30 | 617 | 1.32 | 100% |
| MR | Mathematics | 11–29 | 1598 | 1.47 | 100% |
| MR | Mathematics | ≥30 | 615 | 1.26 | 100% |
| EL | ELA | 11–29 | 1194 | 1.63 | 100% |
| EL | ELA | 30–149 | 3937 | 1.29 | 100% |
| EL | ELA | ≥150 | 1357 | 1.11 | 100% |
| EL | Mathematics | 11–29 | 1189 | 1.57 | 100% |
| EL | Mathematics | 30–149 | 3940 | 1.25 | 100% |
| EL | Mathematics | ≥150 | 1355 | 1.09 | 100% |
| ELO | ELA | 11–29 | 1867 | 1.59 | 100% |
| ELO | ELA | 30–149 | 3322 | 1.31 | 100% |
| ELO | ELA | ≥150 | 326 | 1.12 | 100% |
| ELO | Mathematics | 11–29 | 1869 | 1.56 | 100% |
| ELO | Mathematics | 30–149 | 3321 | 1.29 | 100% |
| ELO | Mathematics | ≥150 | 323 | 1.12 | 100% |
| RFP | ELA | 11–29 | 2024 | 1.57 | 100% |
| RFP | ELA | 30–149 | 3260 | 1.30 | 100% |
| RFP | ELA | ≥150 | 447 | 1.11 | 100% |
| RFP | Mathematics | 11–29 | 2024 | 1.50 | 100% |
| RFP | Mathematics | 30–149 | 3262 | 1.28 | 100% |
| RFP | Mathematics | ≥150 | 445 | 1.12 | 100% |
| EO | ELA | 11–29 | 571 | 1.54 | 100% |
| EO | ELA | 30–149 | 4182 | 1.23 | 100% |
| EO | ELA | ≥150 | 2535 | 1.08 | 100% |
| EO | Mathematics | 11–29 | 568 | 1.45 | 100% |
| EO | Mathematics | 30–149 | 4190 | 1.17 | 100% |
| EO | Mathematics | ≥150 | 2526 | 1.06 | 100% |
| SED | ELA | 11–29 | 612 | 1.54 | 100% |
| SED | ELA | 30–149 | 3455 | 1.22 | 100% |
| SED | ELA | ≥150 | 3085 | 1.08 | 100% |
| SED | Mathematics | 11–29 | 612 | 1.46 | 100% |
| SED | Mathematics | 30–149 | 3460 | 1.17 | 100% |
| SED | Mathematics | ≥150 | 3078 | 1.07 | 100% |
| SWD | ELA | 11–29 | 3436 | 1.49 | 100% |
| SWD | ELA | ≥30 | 3041 | 1.28 | 100% |
| SWD | Mathematics | 11–29 | 3440 | 1.41 | 100% |
| SWD | Mathematics | ≥30 | 3034 | 1.22 | 100% |
| FOS | ELA | 11–29 | 49 | 1.46 | 100% |
| FOS | Mathematics | 11–29 | 48 | 1.48 | 100% |
| HOM | ELA | 11–29 | 1093 | 1.47 | 100% |
| HOM | ELA | ≥30 | 632 | 1.24 | 100% |
| HOM | Mathematics | 11–29 | 1095 | 1.48 | 100% |
| HOM | Mathematics | ≥30 | 629 | 1.24 | 100% |

1. The size bins refer to the number of students within schools with growth scores in grades four through eight in 2018–19 for the student group and subject of interest (indicated in the first two columns).

**Appendix A2: LEA-Level Stability and Accuracy Results**

**Table B1. Cross Year Stability of the 2018–19 Two-Year EBLP Weighted Average and the Simple Average for LEAs**

| **Student Group** | **Subject** | **2018–19 LEA Size1** | **Number of LEAs** | **Correlation between the 2017–18 and 2018–19 Simple Average** | **Correlation between the 2017–18 and 2018–19 2-‍year EBLP** |
| --- | --- | --- | --- | --- | --- |
| ALL | ELA | 11–149 | 200 | 0.32 | 0.59 |
| ALL | ELA | 150–1499 | 323 | 0.37 | 0.47 |
| ALL | ELA | ≥1500 | 299 | 0.65 | 0.66 |
| ALL | Mathematics | 11–149 | 200 | 0.33 | 0.57 |
| ALL | Mathematics | 150–1499 | 324 | 0.43 | 0.51 |
| ALL | Mathematics | ≥1500 | 298 | 0.85 | 0.85 |
| AA | ELA | 11–149 | 226 | 0.13 | 0.59 |
| AA | ELA | ≥150 | 110 | 0.67 | 0.73 |
| AA | Mathematics | 11–149 | 226 | 0.08 | 0.64 |
| AA | Mathematics | ≥150 | 110 | 0.63 | 0.84 |
| AI | ELA | 11–1499 | 217 | 0.17 | 0.60 |
| AI | Mathematics | 11–1499 | 216 | 0.11 | 0.61 |
| AS | ELA | 11–149 | 212 | -0.03 | 0.48 |
| AS | ELA | ≥150 | 184 | 0.53 | 0.59 |
| AS | Mathematics | 11–149 | 212 | 0.33 | 0.72 |
| AS | Mathematics | ≥150 | 184 | 0.80 | 0.86 |
| FI | ELA | ≥11 | 305 | 0.06 | 0.50 |
| FI | Mathematics | ≥11 | 305 | 0.18 | 0.54 |
| HI | ELA | 11–149 | 256 | 0.20 | 0.61 |
| HI | ELA | 150–1499 | 305 | 0.41 | 0.61 |
| HI | ELA | ≥1500 | 185 | 0.63 | 0.68 |
| HI | Mathematics | 11–149 | 257 | 0.22 | 0.62 |
| HI | Mathematics | 150–1499 | 306 | 0.44 | 0.57 |
| HI | Mathematics | ≥1500 | 184 | 0.72 | 0.74 |
| PI | ELA | 11–1499 | 169 | 0.06 | 0.43 |
| PI | Mathematics | 11–1499 | 169 | 0.13 | 0.51 |
| WH | ELA | 11–149 | 316 | 0.21 | 0.51 |
| WH | ELA | ≥150 | 422 | 0.48 | 0.60 |
| WH | Mathematics | 11–149 | 316 | 0.28 | 0.53 |
| WH | Mathematics | ≥150 | 422 | 0.59 | 0.65 |
| MR | ELA | 11–29 | 102 | 0.18 | 0.63 |
| MR | ELA | 30–149 | 197 | 0.23 | 0.59 |
| MR | ELA | ≥150 | 139 | 0.52 | 0.69 |
| MR | Mathematics | 11–29 | 102 | 0.13 | 0.67 |
| MR | Mathematics | 30–149 | 198 | 0.47 | 0.66 |
| MR | Mathematics | ≥150 | 138 | 0.70 | 0.76 |
| EL | ELA | 11–149 | 246 | 0.06 | 0.62 |
| EL | ELA | 150–1499 | 304 | 0.50 | 0.71 |
| EL | ELA | ≥1500 | 120 | 0.70 | 0.80 |
| EL | Mathematics | 11–149 | 246 | 0.30 | 0.72 |
| EL | Mathematics | 150–1499 | 304 | 0.67 | 0.86 |
| EL | Mathematics | ≥1500 | 120 | 0.82 | 0.90 |
| ELO | ELA | 11–149 | 277 | 0.05 | 0.49 |
| ELO | ELA | ≥150 | 334 | 0.44 | 0.58 |
| ELO | Mathematics | 11–149 | 276 | 0.29 | 0.72 |
| ELO | Mathematics | ≥150 | 334 | 0.60 | 0.77 |
| RFP | ELA | 11–149 | 259 | 0.15 | 0.63 |
| RFP | ELA | ≥150 | 330 | 0.51 | 0.64 |
| RFP | Mathematics | 11–149 | 260 | 0.23 | 0.70 |
| RFP | Mathematics | ≥150 | 330 | 0.76 | 0.86 |
| EO | ELA | 11–149 | 262 | 0.31 | 0.62 |
| EO | ELA | 150–1499 | 345 | 0.46 | 0.56 |
| EO | ELA | ≥1500 | 205 | 0.69 | 0.70 |
| EO | Mathematics | 11–149 | 262 | 0.24 | 0.48 |
| EO | Mathematics | 150–1499 | 346 | 0.60 | 0.64 |
| EO | Mathematics | ≥1500 | 204 | 0.86 | 0.86 |
| SED | ELA | 11–149 | 257 | 0.15 | 0.55 |
| SED | ELA | 150–1499 | 330 | 0.43 | 0.57 |
| SED | ELA | ≥1500 | 199 | 0.58 | 0.62 |
| SED | Mathematics | 11–149 | 257 | 0.25 | 0.54 |
| SED | Mathematics | 150–1499 | 330 | 0.42 | 0.52 |
| SED | Mathematics | ≥1500 | 199 | 0.74 | 0.75 |
| SWD | ELA | 11–29 | 114 | -0.06 | 0.37 |
| SWD | ELA | 30–149 | 232 | 0.05 | 0.46 |
| SWD | ELA | ≥150 | 322 | 0.51 | 0.60 |
| SWD | Mathematics | 11–29 | 113 | -0.18 | 0.27 |
| SWD | Mathematics | 30–149 | 232 | 0.04 | 0.35 |
| SWD | Mathematics | ≥150 | 322 | 0.49 | 0.56 |
| FOS | ELA | 11–1499 | 180 | 0.02 | 0.61 |
| FOS | Mathematics | 11–1499 | 176 | 0.07 | 0.51 |
| HOM | ELA | 11–29 | 100 | -0.03 | 0.58 |
| HOM | ELA | 30–149 | 149 | 0.09 | 0.53 |
| HOM | ELA | ≥150 | 101 | 0.52 | 0.64 |
| HOM | Mathematics | 11–29 | 100 | 0.04 | 0.49 |
| HOM | Mathematics | 30–149 | 149 | 0.05 | 0.44 |
| HOM | Mathematics | ≥150 | 101 | 0.65 | 0.77 |

1. The size bins refer to the number of students within LEAs with growth scores in grades four through eight in *both* 2017–18 and 2018–19 for the student group and subject of interest (indicated in the first two columns). The number of LEAs in each bin is smaller than in [Table 3](#Tabke3), given that not all LEAs have estimates in both years.

**Table B2. Improvement in Accuracy of Growth Estimates Using Two-Year EBLP Weighted Averages Versus Simple Averages at the LEA Level**

| **Student Group** | **Subject** | **2018–19 LEA Size1** | **Number of LEAs** | **Mean Accuracy Ratio for 2-‍year EBLP vs. Simple Average** | **Percentage of LEAs with Improved Accuracy for 2-‍year EBLP vs. Simple Average** |
| --- | --- | --- | --- | --- | --- |
| ALL | ELA | 11–149 | 205 | 1.44 | 100% |
| ALL | ELA | 150–1499 | 323 | 1.09 | 100% |
| ALL | ELA | ≥1500 | 299 | 1.01 | 100% |
| ALL | Mathematics | 11–149 | 205 | 1.39 | 100% |
| ALL | Mathematics | 150–1499 | 324 | 1.07 | 100% |
| ALL | Mathematics | ≥1500 | 298 | 1.01 | 100% |
| AA | ELA | 11–149 | 240 | 1.68 | 100% |
| AA | ELA | ≥150 | 110 | 0.99 | 55% |
| AA | Mathematics | 11–149 | 240 | 2.08 | 100% |
| AA | Mathematics | ≥150 | 110 | 1.36 | 85% |
| AI | ELA | 11–1499 | 233 | 2.17 | 100% |
| AI | Mathematics | 11–1499 | 232 | 2.33 | 100% |
| AS | ELA | 11–149 | 218 | 1.45 | 100% |
| AS | ELA | ≥150 | 184 | 1.01 | 63% |
| AS | Mathematics | 11–149 | 218 | 1.59 | 99% |
| AS | Mathematics | ≥150 | 184 | 0.66 | 24% |
| FI | ELA | ≥11 | 321 | 1.64 | 97% |
| FI | Mathematics | ≥11 | 320 | 1.49 | 87% |
| HI | ELA | 11–149 | 264 | 1.73 | 100% |
| HI | ELA | 150–1499 | 305 | 1.09 | 72% |
| HI | ELA | ≥1500 | 185 | 0.90 | 28% |
| HI | Mathematics | 11–149 | 265 | 1.65 | 100% |
| HI | Mathematics | 150–1499 | 306 | 1.07 | 78% |
| HI | Mathematics | ≥1500 | 184 | 0.94 | 34% |
| PI | ELA | 11–1499 | 176 | 1.58 | 95% |
| PI | Mathematics | 11–1499 | 176 | 1.98 | 99% |
| WH | ELA | 11–149 | 324 | 1.58 | 100% |
| WH | ELA | ≥150 | 422 | 1.09 | 95% |
| WH | Mathematics | 11–149 | 324 | 1.43 | 100% |
| WH | Mathematics | ≥150 | 422 | 1.06 | 100% |
| MR | ELA | 11–29 | 114 | 1.88 | 100% |
| MR | ELA | 30–149 | 197 | 1.57 | 100% |
| MR | ELA | ≥150 | 139 | 1.19 | 83% |
| MR | Mathematics | 11–29 | 113 | 1.74 | 100% |
| MR | Mathematics | 30–149 | 198 | 1.36 | 96% |
| MR | Mathematics | ≥150 | 138 | 0.98 | 48% |
| EL | ELA | 11–149 | 252 | 1.88 | 100% |
| EL | ELA | 150–1499 | 304 | 1.06 | 59% |
| EL | ELA | ≥1500 | 120 | 0.57 | 11% |
| EL | Mathematics | 11–149 | 251 | 2.13 | 100% |
| EL | Mathematics | 150–1499 | 304 | 1.11 | 60% |
| EL | Mathematics | ≥1500 | 120 | 0.35 | 3% |
| ELO | ELA | 11–149 | 289 | 1.59 | 100% |
| ELO | ELA | ≥150 | 334 | 0.91 | 43% |
| ELO | Mathematics | 11–149 | 289 | 1.87 | 99% |
| ELO | Mathematics | ≥150 | 334 | 0.86 | 39% |
| RFP | ELA | 11–149 | 273 | 1.71 | 100% |
| RFP | ELA | ≥150 | 330 | 1.01 | 54% |
| RFP | Mathematics | 11–149 | 273 | 2.04 | 100% |
| RFP | Mathematics | ≥150 | 330 | 1.12 | 60% |
| EO | ELA | 11–149 | 265 | 1.50 | 100% |
| EO | ELA | 150–1499 | 345 | 1.08 | 97% |
| EO | ELA | ≥1500 | 205 | 1.01 | 95% |
| EO | Mathematics | 11–149 | 265 | 1.38 | 100% |
| EO | Mathematics | 150–1499 | 346 | 1.06 | 100% |
| EO | Mathematics | ≥1500 | 204 | 1.01 | 100% |
| SED | ELA | 11–149 | 270 | 1.50 | 100% |
| SED | ELA | 150–1499 | 330 | 0.98 | 60% |
| SED | ELA | ≥1500 | 199 | 0.58 | 9% |
| SED | Mathematics | 11–149 | 270 | 1.44 | 100% |
| SED | Mathematics | 150–1499 | 330 | 1.02 | 71% |
| SED | Mathematics | ≥1500 | 199 | 0.87 | 18% |
| SWD | ELA | 11–29 | 121 | 1.78 | 100% |
| SWD | ELA | 30–149 | 232 | 1.35 | 100% |
| SWD | ELA | ≥150 | 322 | 1.06 | 84% |
| SWD | Mathematics | 11–29 | 121 | 1.66 | 100% |
| SWD | Mathematics | 30–149 | 232 | 1.27 | 100% |
| SWD | Mathematics | ≥150 | 322 | 1.05 | 90% |
| FOS | ELA | 11–1499 | 195 | 1.96 | 100% |
| FOS | Mathematics | 11–1499 | 192 | 2.13 | 100% |
| HOM | ELA | 11–29 | 130 | 1.65 | 100% |
| HOM | ELA | 30–149 | 152 | 1.32 | 98% |
| HOM | ELA | ≥150 | 101 | 0.90 | 42% |
| HOM | Mathematics | 11–29 | 129 | 1.75 | 100% |
| HOM | Mathematics | 30–149 | 152 | 1.48 | 100% |
| HOM | Mathematics | ≥150 | 101 | 1.17 | 82% |

The size bins refer to the number of students within LEAs with growth scores in grades four through eight in 2018–19 for the student group and subject of interest (indicated in the first two columns).

1. The CDE is considering a linear translation of the residual gain scores to have marginal mean of 100 rather than 0. Such a translation would have no impact on the accuracy of AGMs, and it is not considered further in this document. [↑](#footnote-ref-2)
2. This excludes one of the AGMs with a large discrepancy between EBLP and simple averages, and for which investigation of the underlying growth scores revealed highly anomalous patterns. [↑](#footnote-ref-3)