THE OHIO STATE UNIVERSITY

# Ohio's English Language Proficiency Gap Closing Metric: An Early Assessment

**Dr. Sara E. Watson**

Department of Political Science

The Ohio State University

2140 Derby Hall

154 N. Oval Mall

Columbus, OH 43210-1393

February 2020

This brief assessed Ohio's new Gap Closing report card metric: English Language Proficiency (ELP). It describes the new ELP accountability regime and provides early data on how Ohio schools are progressing. It highlights three features of Ohio's ELP gap closing measure that deserve more critical scrutiny: the use of thresholds to determine both student and school progress toward ELP proficiency; the inconsistent opportunity for schools to earn partial credit on the ELP gap closing measure; and the heavy weighting of the ELP measure. It shows how, taken together, these features create a number of perverse outcomes—namely, that very small numbers of students can have disproportionately large outcomes on not only the ELP sub-component grade, but on the overall gap closing grade and, indeed, on the *entire* report card grade. It concludes by offering suggestions for how Ohio might improve its assessment of ELP.

## Ohio's English Language Proficiency Gap closing Metric: An Early Assessment[1]

**Dr. Sara E. Watson**

**The Ohio State University**

Nearly one-third of children in the United States currently live in a household where a language other than English is spoken, and approximately 5 million students in American public schools are classified as English Language Learners (ELLs). An important change associated with the Every Student Succeeds Act (ESSA)--the 2015 reauthorization of the 1965 Elementary and Secondary Education Act—was the inclusion of a number of new requirements for the education of ELLs. For the first time, states were required to create standardized criteria for identifying ELLs and to include English proficiency as an indicator of school quality in their accountability system. In contrast to its predecessor the No Child Left Behind (NCLB) Act, however, ESSA allowed individual states to make important decisions such as how to define English proficiency and how quickly ELL students should achieve proficiency, among other factors.

Ohio's version of ESSA's English Language Proficiency (ELP) requirement— known as ELP Gap closing—was introduced into the school accountability system two years ago. Last academic year, some 14 percent of Ohio schools had the ELP gap closing measure included in their report card score calculation. Perhaps because of the relative newness of the ELP measure, however, it has not received much attention.[2] This is unfortunate, as English language learners are one of the state's fastest growing student populations, and progress toward English proficiency is an important prerequisite for meaningful participation in their educational programs and services.

This policy brief provides an evaluation of Ohio's new ELP regime. It proceeds in three steps. First, it describes the new ELP accountability system in greater detail and provides some preliminary data on how Ohio schools are progressing under the new regime. Second, it highlights three features of Ohio's ELP gap closing measure that deserve more critical scrutiny: the use of thresholds to determine both student and school progress toward ELP proficiency; the inconsistent opportunity for schools to earn partial credit on the ELP gap closing measure; and the heavy weighting of the ELP measure. It shows how, taken together, these features create a number of perverse outcomes—namely, that very small numbers of students can have disproportionately large outcomes on not only the ELP sub-component grade, but on the overall gap closing grade and, indeed, on the *entire* report card grade.

The brief concludes with some suggestions on how Ohio might improve its assessment of ELP so as to continue to incentivize attention to some of the state's most vulnerable students without unfairly disadvantaging schools educating English learners.

[1] Thanks to Vlad Kogan and Raj Arunachalam for encouraging me to undertake this study and for feedback on a first draft; and to Marianne Motley at the ODE for help in understanding the report card scoring system. The author is also grateful to Ricardo Garcia-Graiff and Cam Gallentine for helpful research assistance.

[2] For example, the December 2019 report of the Ohio Report Card Study Committee, even in its discussion of possible reforms to the gap closing component, made no explicit mention of any potential recommendations relating to the ELP sub-component.

**Ohio's New Metric For Assessing Progress Toward English Language Proficiency**

Assessments of English Language learners are included in Ohio School Report Card's Gap Closing category, one of the six major report card components.[3] Broadly speaking, the purpose of the gap closing category is to assess progress toward achieving the goals defined in Ohio's ESSA plan, which called for cutting certain achievement gaps by 50 percent over the first decade. Prior to 2017, Ohio schools had been assessed on the gap closing measure based on how various sub-groups—including African-Americans, Hispanic, special needs, and English Language Learners --performed in three areas: English Language Arts, Math and Graduation Rates. If a school had a sufficiently large number of eligible students (30 or more African-American or Hispanic students, for example), those students were included as one of several sub-groups whose performance on these assessments were benchmarked to expected standards.

Two years ago, the calculation of the Gap closing measure underwent an important series of changes.[4] First, 2017-2018 saw the addition of a new gap closing category on which schools were evaluated: progress among English language learners in achieving English proficiency. Second, the 'minimum-N' size—the number of students in a given subgroup required for a school to be assessed based on their performance—was changed. Whereas in 2016 the minimum N used for gap closing was 30, in 2017-18 it was 25, and in 2018-19 it was 20. In 2020 it will be 15. This reduction in the minimum-N size applied to all sub-groups analyzed in the Gap closing measure, not just English language learners.

Figures 1 shows trends in the percent of Ohio schools subject to English Language Proficiency (ELP) gap closing accountability and in the profiles of those schools. As illustrated in Figure 1, in 2017-18, when the minimum-N size was 25 students, 9 percent of Ohio schools (299 schools) had the ELP gap closing measure included in their report card score calculation. As the minimum-N for subgroup accountability was lowered to 20 in 2018-19, the percent of schools evaluated on ELP gap closing rose to approximately 14 percent (481 schools). According to the ODE, if the 2020 N-size of 15 had been in effect in the 2018-19 academic year, 18 percent of Ohio schools (615 buildings) would have been subject to ELP accountability.[5]

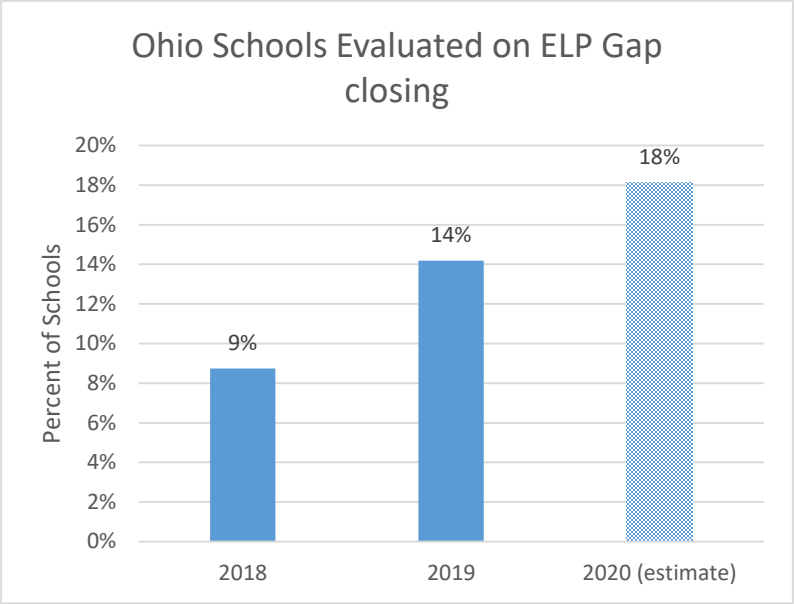
Ohio's new ELP gap closing regime is based on student performance on Ohio's English Language Proficiency Assessment (OELPA). The OELPA measures progress toward English proficiency in four domains: listening, speaking, reading, and writing. For each of these domains, the OELPA defines five ranges of performance levels: beginning; early intermediate; intermediate; early advanced; and advanced. Proficiency on the OELPA is met by scoring in the early advanced and/or advanced categories across all four domains.

[3] Gap closing is also known as Annual Measurable Objectives (AMOs).

[4] There were two other important changes implemented to gap closing scores beginning in 2018. First, scores for ELA and math began to be calculated using a Performance Index score rather than the percent of students scoring 'proficient' or above. Second, rather than assess all sub-groups of students based on a single goal, the ELA, math and graduation rate goals were based on the statewide performance of each subgroup.

[5] Personal communication.

*Figure 1: Percent of Ohio Schools Evaluated on English Language Proficiency Gap Closing, 2018-2020*




Source: ODE Report Card Data

For students that have not yet met proficiency, Ohio sets improvement targets that vary by their initial point tally, as shown in Figure 1. For elementary and middle school students with an initial score of 4 to 11 points, their expected progress is two points per year. That is, they must move up a performance level in two domains each year in order to meet the ELP gap closing target. Because research suggests that young English learners progress more quickly than older students, the expectation for improvement among more advanced students is lower. For students who have an initial score of 12 to 18 points, the expectation is an increase of 1 point per year—improvement in only one domain.

*Table 1: Annual Student Level English Language Proficiency Improvement Targets for English Learners*

| **Student's Grade Level when Identified as English Learner** | **Sum of Student's Initial OELPA Domain Score (point range)** | **Student Level Target for Annual Improvement (points/year)** |
|---|---|---|
| Kindergarten – 8th Grade | 4 points – 11 points | Increase of **2 points** per year |
| Kindergarten – 8th Grade | 12 points – 18 points | Increase of **1 point** per year |
| 9th Grade – 12th Grade | 4 points – 7 points | Increase of **2 points** per year |
| 9th Grade – 12th Grade | 8 points – 18 points | Increase of **1 point** per year |

Source: ODE 2019

There are many strengths of this system for determining student progress in achieving English language proficiency. One of the clear strengths in Ohio's "growth-to-standard" model is the explicit linkage between the state's long-term goals and measures of interim progress (Goldschmidt and Hakuta 2017). This model sets annual student-level targets aligned to Ohio's 7-year timeline for reaching proficiency.

Another clear strength of Ohio's model is the recognition that English-language acquisition is non-linear, and of the "lower is faster, higher is slower" rule for English development.[6] Finally, although ESSA permits states to include former (reclassified) ELLs in the ELP measure for up to four years after existing ELL status, Ohio's measure excludes former ELLs. It therefore provides a clear assessment of the progress of current ELLs that is uncontaminated by the inclusion of those who are already English proficient.

At the same time, Ohio's ELP gap closing measure contains some important drawbacks. One is the measure's reliance on a series of thresholds for assigning points. At a fundamental level, thresholds like this are problematic because a substantial amount of important information is lost when students and/or schools are simply counted as meeting or not meeting a target, rather than assessing actual achievement or growth.[7] But Ohio's ELP thresholds, in tandem with other features of the ELP gap closing measure—most notably, the limited ability of many schools to earn partial points and the sometimes heavy weighting of ELP—are also problematic because they can impose severe penalties for relatively small differences across schools. The following sections highlights these problems in greater detail.

***Problem 1: Thresholds for Assessing Progress and All-or-Nothing Point Assignment***

The Ohio report card scoring system for ELP gap closing relies on two sets of thresholds. The first threshold relates to assessing *student* progress in learning English. As discussed above, the state's assessment system requires that students make annual progress toward proficiency, where growth is based on advancement across performance levels rather than improvement in scaled scores. For example, students in the beginning category in year 1 are expected to move to the early intermediate category in year 2. What this means, however, is that depending on the initial location of a student score in the overall distribution, a school that achieves higher growth in ELP—but whose students started further from the threshold—may easily receive no credit for improvement while a school whose student experience lower growth, but which started closer to the threshold, receives full credit.[8] This type of threshold is also problematic in terms of incentives: it can encourage schools to focus their efforts on students "on the bubble" rather than promoting learning across the entire spectrum.[9]

Unfortunately, due to a lack of publicly available data, we cannot currently assess how common it is for higher-growth schools to receive lower ELP grades than lower-performing peers, nor the allocation of effort toward "bubble" students. But given that Ohio has moved away from thresholds in other areas of the report card due to similar concerns, its continued use in ELP gap closing merits closer attention.

[6] See Cook et al 2008 and Carnock 2017.
[7] This is not a new observation—and indeed, has long been part of the debate with respect to other dimensions of Ohio's gap closing measure.
[8] For a technical discussion of the perverse effects of this version of the Growth to Standard (GTS) model in ELP, see pages 21-25 of Goldschmidt and Hakuta (2017).
[9] For example, Neal and Schanzenbach (2010) find that teachers are more likely to target "bubble students" near passing cutoffs, because it gives the largest boost to class passing rates.

In addition to student-level thresholds for assessing ELP progress, there is a second type of threshold used in computing the gap closing score that is more amenable to analysis. This is a threshold for determining whether a *school* is making sufficient progress toward ELP gap closing.

Here, it is important to understand how individual student performance on the OELPA translates into building-level report card grades. This happens through one of two pathways. The first pathway involves a building-level proficiency threshold. If a sufficient proportion of ELL students in a building meet a proficiency threshold, defined as a 1- or 2-point gain on the OELPA, then a school receives full credit. In 2017-18, the threshold was set at 51 percent. If 51 percent of a school's ELL students scored at least a 2-point gain on the OELPA, the school received full credit (100 points) for the ELP portion of the gap closing grade. In 2018-19, the threshold was raised to 54 percent, and is set to rise by approximately three percentage points each year through 2025-26, when the threshold will be held steady at 75 percent. The motivation of this steady rise in the proficiency threshold is that schools are being incentivized to "close" the gap between native and English learner students.

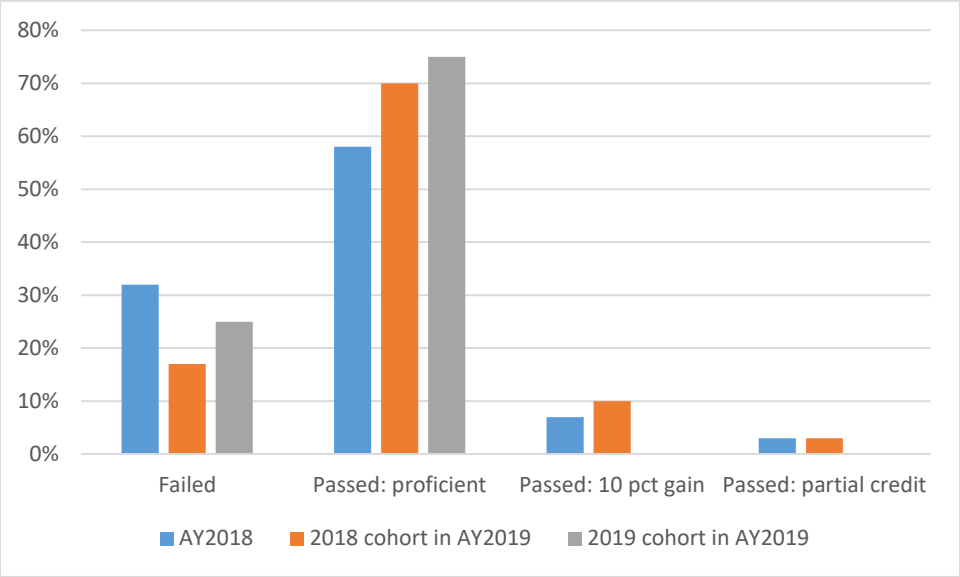
For schools that fail to meet this benchmark, there is a second possible pathway. Schools may earn full or partial credit toward the ELP gap closing for improvement over the previous year in closing the school's achievement gap.[10] Schools which fail to meet the annual proficiency threshold but which achieve a ten percent improvement vis-à-vis the previous year are granted the full 100 points, while schools that make some progress receive partial points.

Figure 2 shows the distribution of different outcomes across those Ohio schools assessed on ELP gap closing in the first two years of the new regime. It divides the sample of ELP-assessed schools into three groups, showing first the performance of schools which were assessed on ELP gap closing in 2018 (the 2018 cohort); the performance of this same 2018 cohort in 2019; and, finally, the performance of schools that were assessed on ELP gap closing for the first time in 2019 (the 2019 cohort).

The overall results from the first two years of assessment are somewhat encouraging. Among the schools that were assessed on the ELP gap closing benchmark in 2018 and 2019, a majority of schools met the proficiency standard as it was defined in that year. In 2018, for example, 58 percent of assessed schools met the proficiency standard of 51 percent. And in 2019, despite the increase in the proficiency threshold to 54 percent, approximately 70 percent of assessed schools met the threshold. Moreover, when we compare the performance of the 2018 cohort across the two years of data, we see a substantial improvement in the second year of assessment: whereas only 58 percent of the 2018 cohort met the threshold in 2018, some 70 percent met it in 2019.

[10] For a given subgroup of a school, this calculation is: one-year improvement in proficiency / ELP AMO gap.

***Figure 2: Outcomes on ELP Gap Closing, 2018-2019***




Source: ODE Report Card Data. *Note: For each year, Figure 2 shows percent of schools assessed on ELP gap closing that failed the threshold (awarded zero points); passed the threshold (awarded 100 points); failed the threshold but achieved ten percent improvement in gap closing (100 points); and failed the proficiency threshold but received partial credit (1-99 points).*

Nevertheless, there remain a number of concerns. The first is that outcomes are less rosy if one assesses schools against the target 2026 benchmark of 75 percent: in 2019, only 22 percent of affected schools met the long-term proficiency threshold. To be sure, this may be less a concern if Ohio schools continue to post year-on-year improvements, as shown in Figure 2. But with only one year of assessment data to draw from, it is difficult to know if the increase is a trend or a merely a one-year bump.

Another concern is that, of the schools failing to meet the proficiency threshold, very few end up receiving partial credit on ELP gap closing. In 2019, only 20 percent of such schools received points through this alternative pathway. Of 9 schools receiving partial points in 2019, their average score was an F, with the mean points awarded 52 out of 100.

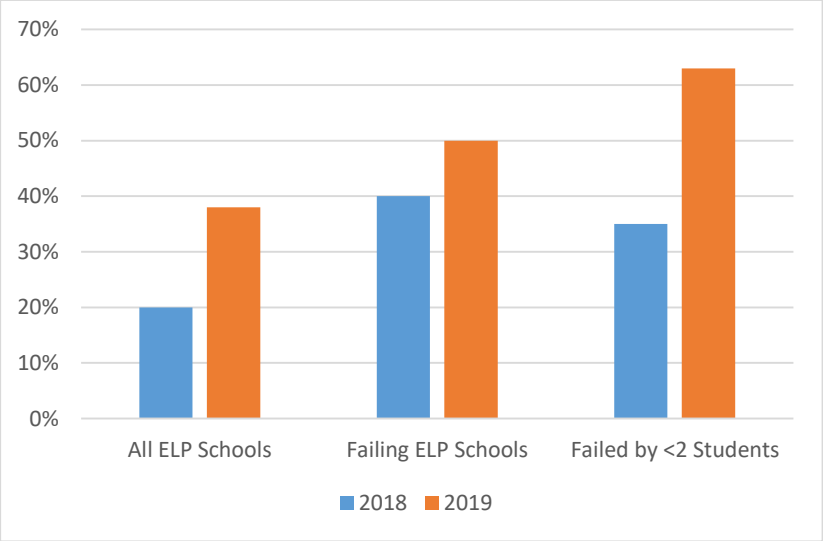
Closer inspection of the data also indicate that a substantial number of ELP-failing schools are entirely ineligible to earn partial points through the second pathway. This phenomenon is driven by the combination of changing minimum-N sizes and participation rules for "first year" schools.[11] Over time, the minimum number of ELL students needed to trigger assessment on ELP gap closing has fallen, making more schools subject to ELP gap closing accountability. But during the first year in which a school is assessed on ELP gap closing, it has no prior scores to draw from—and hence has no opportunity to gain points through the "improvement" pathway. For these schools, ELP gap closing

[11] The low rate of awarding partial credit may also be due to the fact that some failing schools have a very large gap to close. For example the mean gap to be closed in 2019 for failing schools was 32 points—and as high as 75 points. For some schools, then, closing the gap is quite an ambitious task.

points are effectively all-or-nothing. Schools either meet the annual proficiency threshold, earning the full 100 points, or they do not—and are assigned 0 points.

Figure 3 shows the proportion of ELP-assessed schools which have been ineligible to earn partial points. In 2019, for example, nearly 40 percent of schools were ineligible to earn points through the alternative pathway. Among failing schools, fifty percent had no alternative pathway available, and were consigned to receiving zero points on the sub-component.

***Figure 3: Percent of Schools Assessed on ELP Gap closing Ineligible to Earn Partial Points Due to Lack of Prior Year Score.***




*Source:* ODE Report Card Data.

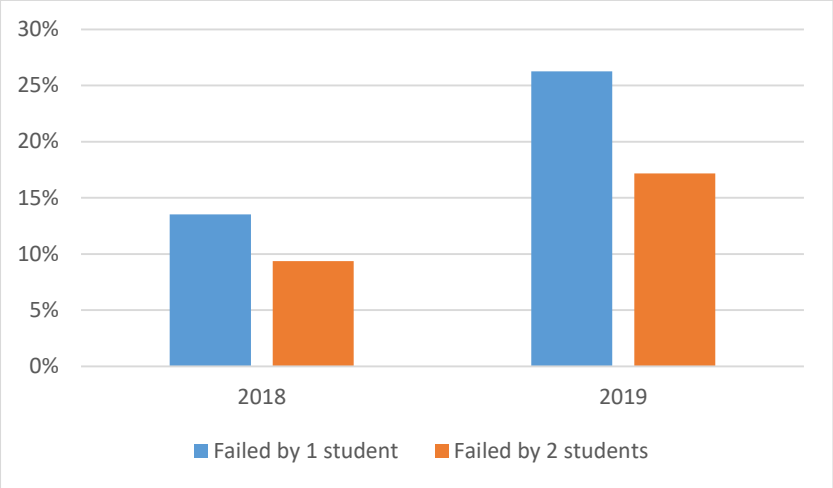
The lack of opportunity for "first year" ELP schools to earn partial points is especially troubling given that a substantial portion of schools failing the ELP benchmark missed the proficiency threshold by a very small number of students.[12] Figure 4 shows that, in 2018 approximately 25 percent of schools which failed the ELP proficiency threshold did so by only one or two students; in 2019, this figure rose to 40 percent.[13] Similarly, of schools which failed to meet the ELP proficiency threshold by only one or two students—well within the confidence intervals for meeting the threshold, one might add—nearly two-thirds had no possibility to earn partial points through the alternative pathway. [14]

[12] For eighty percent of these schools, the 95% confidence intervals include values higher than the proficiency threshold for that year.

[13] Visual inspection of the data suggest that the spike here is more the result of the threshold being raised than by a change in schools' performance.

[14] One might argue that this problem of absent previous-year scores may diminish over time, as more schools become subject to ELP gap closing accountability and establish a performance record on ELP assessment. But this misses the fact that the population of ELP students in a given school is likely to vary substantially from year to year because ELLs progress out of the academic status and because of other factors such as high rates of geographic mobility. We only have one year of data to

*Figure 4: Percent of Failing Schools Missing ELP Proficiency Threshold by 1 or 2 Students*




*Source:* ODE Report Card Data

**Problem 2: Heavy Weighting of ELP Sub-Component Leads to Inflated Penalties**

The problems described above—thresholds which discard substantial information about actual student growth and a highly punitive all-or-nothing point system for schools lacking a prior ELP record --might be less troubling if the ELP gap closing measure played only a modest role in report card scores. But now we come to a second major problem with the ELP gap closing component: the fact that for some schools it weighs very heavily on the overall gap closing measure--and can have dramatic effects on the entire report card grade.

Unlike the other measures of gap closing, which score schools based on the average performance of multiple subgroups (white, African-American, Hispanic, economically disadvantaged, disabled, etc) in math, ELA and graduation, the entire ELP gap closing grade is based on the performance of a single subgroup: English learners.[15] This means that sub-par performance by EL learners cannot be counter-balanced by the performance of another group.

assess, but in 2019, even as the minimum subgroup-N size was *reduced* from 25 to 20 (making it harder for schools to be exempted from ELP assessment), some 4 percent of the original 2018 cohort nevertheless "exited" ELP status. Should these schools re-enter the sample in the future, they will once again be treated as "first year" schools. After 2020, when the minimum-N size stabilizes at 15, it seems likely that schools with smaller populations of ELL students could easily shift in and out of eligibility for accountability, and therefore have spotty 'prior-year' records. More broadly, the current system for assigning partial points within ELP gap closing is also worth re-evaluating due to concerns about statistical validity associated with small N-sizes in certain settings (ie, thresholds), a problem mentioned only in passing in ODE's ESSA documentation and its webinar around minimum-N sizes.

[15] In both 2018 and 2019, the mean and median number of subgroups assessed in ELA and math gap closing across all Ohio schools was 4. In ELP gap closing, however, there is only one assessed subgroup.

Moreover, as highlighted in Table 2, for most schools being held accountable for ELP gap closing, the performance of English learners constitutes one-third to one-quarter of the entire gap closing grade—but in some instances constitutes as much as 100 percent. And because the Ohio report card system weights the gap closing component differently depending on a school's population and grade configuration[16], the gap closing measure itself can account for anywhere between 4. 5 and 47 percent of a school's overall report card grade. Indeed, there is currently no correlation between the weighting of ELP in a school's overall report card grade and the percent ELP students in a given school population.

***Table 2: Contribution of ELP to Gap closing Score and to Overall Report Card Score***

| Type of School | Contribution of ELP to Gap closing Score | Contribution of ELP to Overall Report Card |
|---|---|---|
| High Schools | 25% | 4.5% |
| Middle Schools | 33% | 10% |
| K-5 Elementary School | 33% | 7.5% - 11.5% |
| K-3 Elementary School | 33% | 10.5%-15.8% |
| K-2 Elementary School | 100% | 32%-47.5% |

Ohio's reliance on thresholds and (in many instances) all-or-nothing points allocation in ELP assessment, when combined with the heavy weighting of ELP sub-component, creates a highly punitive system which permits dramatic declines in gap closing scores and, indeed, in overall report card scores based on the performance of a very small number of students. [17] In the past two years, more than 100 Ohio schools have been penalized four letter grades on the *entire* Gap Closing measure because they failed ELP. More shocking, dozens of these schools received a final report card grade *one to two grades lower than otherwise equally performing schools* due to the performance of very small numbers of English-language learner**.**

To flesh out what is at stake, consider the experience of Harmon Middle School in Pickerington Local School District in 2019. Harmon Middle School was not assessed on English Language Progress in 2018 because it did not have a sufficient number of English language learners to meet the minimum accountability threshold of 25 students. In 2019, however, the school reported having 25 English language learners in its student body, and thus for the first time the progress of its English language learners was included in its gap closing measure. According to ODE data, 13 of Harmon Middle School's 25 ELL students (52 percent) met the ELP proficiency threshold. But, because the threshold for giving a school credit for having passed was set at 54 percent, the school was judged as not achieving proficiency

[16] For example, most middle schools are only graded on three report card components: achievement, progress and gap closing. Because they have no elementary school or high school students, they are not eligible to be scored on K-3 literacy, graduation rates or 'Prepared for Success'. Likewise, K-5 elementary schools with high numbers of grade-level readers would be graded on progress, achievement, and gap closing, while K-5 schools with lower-achieving readers would additionally be assessed on K-3 literacy. Finally, schools educating only K-2 students might only be graded on the achievement and gap closing components.

[17] It's also worth noting that the inclusion of ELP into the report card does benefit some schools, 'inflating' their final report card score. In the first two years of the ELP regime, approximately 210 schools saw their gap closing scores rise (vis-à-vis the baseline of no ELP) and about 190 saw their scores decline. This analysis focuses on the score declines because the point penalties imposed on schools have been, on average, about three times as large as any 'bonuses.'

and was awarded zero points. Its overall gap closing grade, which would have been a 95 (an A) had it been offered a grace period this year, was reduced to a 63.3—a D.

There are two notable features of the Harmon Middle School example. First, the school missed the proficiency threshold by a single student. Had one more ELL student been made sufficient progress in achieving English proficiency, the school would have met the ELP threshold and been assigned the full 100 points. Moreover, because this was Harmon Middle School's first year of being assessed on ELP gap closing, the school had no previous data from which to assess "annual improvement" and was therefore ineligible to earn partial points for ELP gap closing.

This was not the end of the matter for Harmon Middle School, however. Harmon, like most Ohio middle schools, is scored on only three report card components: achievement, progress/value-added and gap closing. It received C ratings on achievement and progress. Thus, had the school been exempted from assessment on ELP gap closing due to the absence of prior scores, it would have received an overall grade of C. But due to its ELP gap closing score of zero, which had downstream effects on its overall gap closing grade, the school's final report card grade was lowered a full grade, to a D.

Harmon Middle School provides one recent example of the perversity of the gap closing scoring system, but it represents only the tip of the iceberg. Curious about much a single student's performance on the gap closing ELP grade might affect a school's overall report card grade (compared to an otherwise identical school), I created a dataset containing all possible combinations of report card scores for the achievement, progress, K-3 literacy, graduation rates and Prepared for Success components. I then simulated the final report card grade for two schools that were identical except for their ELP gap closing measure.[18] Results are summarized in Table 3.

***Table 3: How Much Can One Student's Performance in ELP Gap closing Affect a School's Overall Report Card Grade?***

| Type of School | Point Reduction (out of 5 points) | Letter Grade Penalty |
|---|---|---|
| High School | .59 points | 0 to 1 Letter Grade |
| Middle School | .98 points | 0 to 1 Letter Grade |
| K-5 Elementary School | .73 to .98 points | 0 to 1 Letter Grade |
| K-3 Elementary School | 1.02 to 1.54 points | 1 to 2 Letter Grades |
| K-2 Elementary School | 1.54 points | 1 to 2 Letter Grades |

The simulation suggests that two otherwise identical schools can--due to the failure of a single English language learner to make sufficient progress -- earn an overall report card grade that is as much as two letter grades lower. That is, a school that would otherwise earn an A could see its entire report card score reduced to a B, or even a C. Consistent with what we might expect from Table 2, early learning

[18] I assume that School A received 100 points on its ELA and math gap closing sub-components and also assume that it passed the ELP gap closing threshold, earning the full 100 points for that sub-component. School B, like Harmon Middle School, is assumed to have missed the threshold by a single student, lacks a prior year score on ELP gap closing, and has no way to earn partial points in this category. Because it receives zero points for the ELP gap closing sub-component, its final gap closing score is reduced to a 66. For each possible combination of report card component scores, I compare the final grades of School A and School B.

elementary schools are the hardest hit. These schools primarily educate K-2 or K-3 students. Because state testing in traditional subjects (math and ELA) does not begin until third grade, they are often assessed on only two report card components: achievement[19] and gap closing. And for these schools, the ELP sub-component comprises the entire gap closing grade. Similarly situated middle and K-5 elementary schools, which are graded on three report card components, can easily see their report card grade fall by an entire letter grade. High schools are the least vulnerable because they are assessed on the largest range of report card components, reducing the weight of ELP gap closing in the overall report card grade. Even so, depending on how close they are to the grade threshold, some high schools can see their grades reduced by a full letter grade.

**Avenues for Reform**

According to the ODE, in 2018 Ohio served approximately 55,000 English learners who collectively make up 3 percent of Ohio's total student population. Recent efforts by Ohio lawmakers to incorporate progress toward English proficiency into the report card are a move in the right direction toward providing educational equity for a population that has been historically marginalized in K-12 policy discussions.

At the same time, school report card scores are high-stakes accountability mechanisms. Parents use them as informational shortcuts in making decisions about where to purchase homes and where to send their kids to school. There are financial repercussions as well: schools which receive low report card grades potentially put their districts on the hook for millions of dollars of EdChoice voucher subsidies. Finally, it is also worth noting that nearly 80 percent of ELP-assessed schools are located in districts adversely affected by Ohio's arbitrary funding caps. In 2019, the funding cap deprived ELP-assessed districts of $4.5 million in categorical funds specifically aimed at educating "Limited English Proficiency" students—and hundreds of millions of dollars overall.[20]

For all these reasons, it is incumbent on Ohio lawmakers to design an ELP measure which balances the need to prioritize the needs of English learners with a system which does not unfairly penalize schools educating substantial ELL populations.

How might ELP assessment in Ohio be productively reformed? Here I highlight three suggested changes: (1) revising the system for assigning partial points in ELP gap closing; (2) moving away from raw threshold measures in assessing ELL progress; and (3) shifting ELP accountability out of gap closing, into either the progress or a standalone category. These reforms would maintain incentives for districts and schools to attend to the needs of English Language Learners while at the same time avoiding inflated penalties for small differences across schools.

[19] It's something of a stretch to claim that K-2 schools are really assessed on achievement, as the only achievement indicator included in these schools' report cards is attendance, something that is not clearly in the school's remit.

[20] Eighty percent of schools assessed on ELP gap closing in 2018 and 2019 were in districts subject to the cap.

*Reform the All-or-Nothing Point Assignment Mechanism for 'first-year' schools.*

A first, and arguably short-term, change would be to reform the system for assigning points for "first-year schools." As we saw, schools in their first year of assessment are punished by a system which provides them with no alternative pathway for earning points, which can have highly punitive consequences. One can imagine several possible strategies to remedy this problem. Schools in their first year might be granted a "grace period", in which they are informed of their under-performing score but have it excluded from the calculation of their grade in that year. Alternatively, these schools might be assigned points based on how close they are to the annual or even the long-term benchmark. If the annual benchmark is 54 percent but the school achieved only 51 percent, they could be assigned 94 points (ie, 94 percent). If this same school were assessed vis-à-vis the 2026 benchmark of 75 percent, it would receive a 68. Note that this latter grade, a D, may be disappointing but is far less punitive than receiving a zero, which is what the current system prescribes.

*Move Away from the Use of Thresholds.*

The solution described above is something of a temporary Band-Aid, however, because it ignores the more fundamental flaws of Ohio's ELP measure. A more ambitious solution for many of the problems outlined above would involve abandoning the use of stark thresholds for measuring student and school progress. Instead, Ohio might consider modifying its current "Growth-to-Standard" model so that each student contributes some amount to a school's rating, points awarded are proportional to the degree of progress, and information regarding actual growth is not lost by focusing on the binary dichotomy of 'sufficient vs insufficient' progress.

*Move ELP out of Gap closing.*

Finally, lawmakers and stakeholders should re-consider whether including ELP progress in the Gap Closing component of the report card is really the best way to assess the progress of Ohio's English learners. Although ESSA requires that ELP be incorporated into state systems of accountability and also requires that states give "substantial weight" to ELP in the overall summative assessment, it does *not* mandate that it be included in a gap-closure category. Indeed, Ohio is only one of six states to embed ELP within a broader metric (rather than creating a discrete, standalone assessment category for ELP). Ohio is also the only state to include ELP in a gap closing measure (Achieve/Unidos 2018). Ohio's gap closing measure has long been criticized for its opaqueness, and the ELP sub-component is no exception. Any layperson opening their school report card would be hard-pressed to locate where information on ELP performance was located, let alone interpret the information provided.

One possibility would be to move the ELP assessment metric into the Progress component of the school report card system. This is a fairly intuitive solution, as the EL progress measure is fundamentally asking whether EL learners are making adequate *progress* toward English language proficiency. If policymakers move in this direction, they might consider using a value-added model (VAM) for assessing ELP, so as to remain consistent with how progress is calculated for other sub-groups.

Alternatively, Ohio might consider creating a stand-alone category for EL progress—something that 44 other US states have done. This strategy would have the benefit of allowing Ohio to adopt a multi-dimensional approach to measuring and assessing ELP. Here, policymakers might productively learn from the experience of other states with larger EL populations who have carefully considered questions of data and appropriate metrics. Consider Oregon which, like Ohio, relies on the ELPA21 assessment tool. Oregon has created a standalone ELP assessment category, and within that category reports two separate English language proficiency indicators. The first is a norm-referenced measure which directly assesses student growth while controlling for grade level, prior achievement and whether a student has had a disrupted education, among other factors. The second indicator, like Ohio's, is a criterion-referenced "growth-to-standard" measure which assesses whether students are "on track to ELP." But rather than rating schools as meeting or not meeting a threshold, it instead assigns them to one of several categories (such as exemplary, commendable, under-performing or lowest-performing). Finally, Oregon has created a new category of "ever-English" learners—that is, students who were once but have now exited EL status. This allows policymakers to monitor former ELLs across their entire K-12 career, even after re-classification, and to diagnose needs and assign supports as required.

Creating a standalone ELP category within Ohio's school accountability system would also open up the space to have thoughtful, evidence-based discussions about factors that should be considered in setting academic goals for ELL students, not just for achieving English language proficiency but across a broader range of content areas. Among questions worth considering are: What student-level and program-level factors should be considered when setting goals for ELL progress?[21] Should ELL status be included as a prior assessment result included in *content* growth models?[22] Finally, given that ELPs represent a relatively small proportion of school populations in many ELP-assessed schools, how heavily should ELP be weighted in the overall summative assessment? Although any weighting system is necessarily subjective, it is especially important to discuss the question of weights in the context of EL. If policymakers weight the indicator too low, schools will have no incentive to invest in EL students; but given the small minimum-N sizes at play, weighting it too high risks unfairly penalizing schools.

Continued dialogue with stakeholders and policymakers on these and other important questions can only improve our understanding of whether, and to what degree, Ohio schools are meeting the needs of their English language learners.

***Dr. Sara Watson*** *is an Associate Professor of Political Science at The Ohio State University. Her research and teaching interests lie in comparative public policy and political economy. She can be reached at watson.584@osu.edu.*

[21] To its credit, Ohio already takes into account a number of *student-level* factors (such as grade-level and disability status), but might also consider other factors that slow ELP such as a history of disrupted education and economic disadvantage. Ohio also currently discounts *program-level* factors such as the language instructional context. Importantly, however, longitudinal studies offer substantial evidence that while ELL students in dual language programs on average take longer to attain English language proficiency and to achieve academically in ELA—*they ultimately do so at higher rates than those in other language instructional models.* See also Thompson 2015; and Steele et al 2017.

[22] Pages 6-10 of Goldschmidt and Hakuta (2017) offer a useful discussion of how such a strategy ensures that schools are not advantaged or disadvantaged based solely on the distribution of language proficiency among students.

*Appendix:*

***Figure A1: English Language Learners as a Percent of School District Enrollment (2019)***

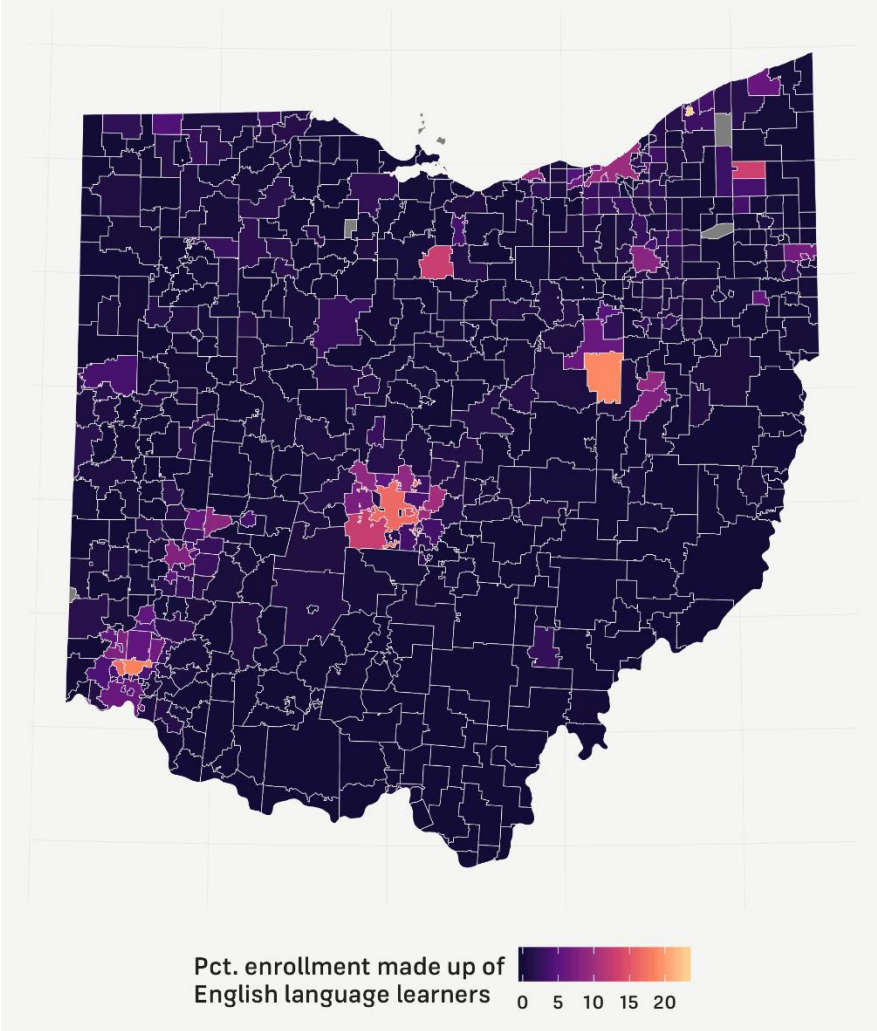



Figure A1 shows that districts with the highest concentration of English Language learners are those located in major urban areas (ie, Columbus and Cincinnati). N=608 school districts.

## REFERENCES[23]


Achieve and UnidosUS. *How are States Including English Language Proficiency in ESSA Plans?*

https://www.scribd.com/document/380681560/Achieve-UnidosUS-ESSA-ELP-Indicator-1-Final

Burho, Jamey. "Oregon's Use of the Ever English Learner Category for Reporting, Accountability, and Research: An AERA presentation from the Partnership." *Oregon Department of Education/Oregon State University English Language Learner Partnership.* Oregon State University, 13 May 2019.

http://blogs.oregonstate.edu/odeosuellpartnership/2019/05/13/oregons-use-of-the-ever-english-learner-category-for-reporting-accountability-and-research-an-aera-presentation-from-the-partnership/

Carnock, Janie Tankard. *Seeing Clearly: Five Lenses to Bring English Learner Data into Focus.* 16 August 2017.

https://www.newamerica.org/education-policy/policy-papers/seeing-clearly/

_____. *Pioneering Change: Leveraging Data to Reform English Learner Education in Oregon.* 16 August 2017.

https://www.newamerica.org/education-policy/policy-papers/pioneering-change/

Churchill, Aaron. "How to address 'Gap Closing' accountability under ESSA." Thomas Fordham Institute Blog, 4 April 2016.

https://fordhaminstitute.org/ohio/commentary/how-address-gap closing-accountability-under-essa

Churchill, Aaron and Jessica Poiner. "Ohio's ESSA Plan: The Gap Closing measure remains a labyrinth." Thomas Fordham Institute Blog , 9 March 2017.

https://fordhaminstitute.org/national/commentary/ohios-essa-plan-gap closing-measure-remains-labyrinth

Cook, H. Gary et al. "Issues in Development of Annual Measurable Achievement Objectives for WIDA Consortium States." Wisconsin Center for Education Research, April 2008, pp. 1-41

"Fix Ohio School Voucher Program Before it Runs Off the Rails." *The Columbus Dispatch,* 28 Jan. 2020.

Goldschmidt, Pete and Kenji Hakuta. "Incorporating English Learner Progress Into State Accountability Systems". Council of Chief State School Officers, 2017, pp. 1-44

Gross, Natalie. "Dual-Language Programs on the Rise Across the U.S." *Blog: Latino Ed. Beat.* Education Writers Association, 3 August 2016.

https://www.ewa.org/blog-latino-ed-beat/dual-language-programs-rise-across-us


[23] Unless otherwise indicated, all weblinks listed here were live as of February 2020.

Neal, Derek and Diane Whitmore Schanzenbach. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *The Review of Economics and Statistics,* vol. 92, issue 2, May 2010

Ohio Department of Education. *ESSA: Minimum N-size for Subgroup Evaluation*. 27 July 2016.

_____. *ESSA: Minimum (N) size for Subgroup Evaluation. (07/27/2016)*. Youtube video, uploaded by, 29 July 2016.

https://www.youtube.com/watch?v=umt4TL4J3uU&list=PLDB1C5-YO_ji7ZQMH9ks3FotvY4BTjcqR&index=2&t=0s

_____. *Ohio School Report Cards: Building AMO Calculations.* 2017-2018 and 2018-2019.

https://reportcard.education.ohio.gov/download

_____. *Revised State Template for the Consolidated State Plan: The Elementary and Secondary Education Act of 1965, as Amended by the Every Student Succeeds Act*. 16 Jan 2018.

http://education.ohio.gov/getattachment/Topics/Every-Student-Succeeds-Act-ESSA/OH_ESSA_SUBMISSION.pdf.aspx

_____. *English Language Proficiency (ELP) Improvement Measure, 2018-2019: Technical Documentation.*

Link has disappeared, only the 2019-20 documentation is now publicly available:
http://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Gap closing-Component/EL-Improvement-Measure_Technical-Documentation_2019-2020.pdf.aspx?lang=en-US

Siegel, Jim. "School funding cap hits central Ohio districts hard, costs $890 million statewide." *The Columbus Dispatch,* 30 Apr. 2017.

*https://www.dispatch.com/news/20170430/school-funding-cap-hits-central-ohio-districts-hard-costs-890-million-statewide*

Seastrom, Marilyn. "Best Practices for Determining Subgroup Size in Accountability Systems While Protecting Personally Identifiable Student Information." US. Department of Education, Institution of Education Sciences. 2017.

Pennsylvania Department of Education. "Classifying Language Instruction Educational Programs (LIEPs)." July 2017.

https://www.education.pa.gov/Documents/Teachers-Administrators/Curriculum/ESL/Classifying%20Language%20Instruction%20Education%20Programs.pdf

Steele, Jennifer L., et al. "Effects of Dual-Language Immersion Programs on Student Achievement: Evidence From Lottery Data." American Educational Research Journal, vol. 54, no. 1_suppl, Apr. 2017, pp. 282S-306S.

Thompson, Karen D. "English Learners' Time to Reclassification: An Analysis." *Educational Policy*, vol. 31, no. 3, May 2017, pp. 330–363.

Thompson, Karen D. et al. “Understanding Outcomes for English Learners: The Importance of the ‘Ever EL' Category.” *Inside IES Research*. Institute of Education Sciences, 16 June 2017.

https://ies.ed.gov/blogs/research/post/understanding-outcomes-for-english-learners-the-importance-of-the-ever-learner-category

Umansky, Ilana M., and Sean F. Reardon. “Reclassification Patterns Among Latino English Learner Students in Bilingual, Dual Immersion, and English Immersion Classrooms.” *American Educational Research Journal, Month XXXX, Vol. XX, No. X,* pp. 1–34, 2014.

Williams, Conor P. and Amaya Garcia. *Federal English Learners Policy Changes: Lots of Talk, Little Policy Clarity*. 15 April 2016.

https://www.newamerica.org/education-policy/edcentral/essaelsair/