Educator **Evaluations & Effectiveness** in Michigan

An analysis of 2011-2012 and 2012-2013 educator evaluation systems surveys and educator effectiveness data

MICHIGAN Department of Education

Executive Summary

The purpose of this policy brief is to provide information about K-12 educator evaluation systems in use across the State of Michigan and to relate information about these systems to other measures of accountability collected by the State. The key findings are:

- There is considerable variation across districts in the factors that inform teacher and administrator effectiveness ratings, in the types of observational tools used, and in the types of measures and amount of student growth data used in year-end evaluations.
- Statewide, 97% of all educators are rated “effective” or “highly effective.”
- The amount of student growth data used in year-end evaluations has little or no relationship to the likelihood that a teacher would be found “ineffective,” “minimally effective,” “effective,” or “highly effective.” However, the variation in educator effectiveness ratings increased if no information was provided by the district on growth data usage in the educator evaluation system.
- Overall, educator effectiveness ratings appear to have little relationship to school accountability labeling. All educator effectiveness ratings are proportionately represented at Reward schools, and a teacher rated “ineffective” is as likely to teach in a Reward school as in a Priority school.
- Over 75% of districts responding to the survey report that annual year-end evaluations are used to determine professional development needs for educators, and over half report that the results inform individualized development plans and coaching support.

Other findings include:

- In 2012-2013, over half (69.1%) of the districts surveyed reported using 20% or more student growth data in annual year-end evaluations.
- For K-8, over half of the districts surveyed report using state assessments or local assessments as student growth measures in year-end evaluations.
- In addition to student growth data, districts report using instructional practices, classroom management practices, pedagogical knowledge, and content knowledge as factors in annual evaluations.
- “Ineffective” and “minimally effective” administrators are much more likely to be present in Priority schools, while Reward schools reported no “ineffective” administrators using their district evaluation system.
- Female teachers, along with those who have spent several years in their district, are professionally certificated in the state of Michigan, hold a Master’s degree or higher, or have full-time status are more likely to be rated “highly effective.” ELA and art teachers are more likely to achieve a “highly effective” rating than elementary teachers,[1] who are in turn more likely to be highly rated than are teachers in mathematics, science, social science, special education, and world languages.
- Minority teachers are more likely to be given a “highly effective” rating than their white counterparts.

[1] Elementary certification is in all subject areas.

Educator Evaluations and Effectiveness in Michigan:

AN ANALYSIS OF 2011-2012 AND 2012-2013 EVALUATION FACTOR SURVEYS AND EDUCATOR EFFECTIVENESS DATA

Introduction

In 2011-2012, Michigan school districts began the development of educator evaluation systems in order to meet the requirements of important new legislation (MCL 380.1249) passed by the Michigan state legislature. The goal of this legislation, as of other educator quality initiatives, is to evaluate all educators (both teachers and administrators) so that all students can be exposed to high quality educators and graduate from high school career, college, and community ready. The law requires all public school and charter school districts in Michigan to evaluate all of their educators (both teachers and administrators) using a rigorous, transparent, and fair performance evaluation system and to report the results of those evaluations in the state's Registry of Educational Personnel (REP), maintained by the Center for Educational Performance and Information (CEPI). The new law also requires evaluations to be based "in significant part" on student growth and to be used to inform decisions regarding instructional leadership abilities, teacher and pupil attendance, professional contributions, training, progress report achievement, school improvement plan progress, peer input, and pupil and parent feedback.[2]

In the two years of educator evaluation implementation since the legislation (2011–2012 and 2012–2013), districts were allowed to redesign, revise, or maintain their existing systems, as long as the systems met basic requirements. The variation in systems makes direct comparison of district effectiveness ratings and systems extremely difficult, because the criteria used to determine "effectiveness" in District A may be very different than those used in District B.

While there still is work to be done, the fact that every Michigan educator is now evaluated, and that the evaluations must be based in part on student achievement data, represents a remarkable accomplishment for Michigan's educational system.

THE 2012-2013 EDUCATOR EVALUATION SYSTEM SURVEY RESULTS

In 2011–2012, and again in 2012–2013, districts were required to respond to a Michigan Department of Education (MDE) developed survey regarding their K-12 Educator Evaluation Systems. The survey asked district administrators to report how their teachers and administrators are evaluated, and was sent to all districts in Michigan, including intermediate school districts (ISDs), local education agencies (LEAs), and public school academies (PSAs). Each district was asked to report on the tools used to evaluate teaching practices, the amount of student growth data incorporated into evaluations, and the factors used to evaluate teachers and administrators. Of the districts that were asked to participate in the K-12 Educator Evaluation Survey, 770 districts provided meaningful information on the content and structure of educator evaluation systems statewide.

Our results are broken down into five distinct sections. The first four focus on the components of the local evaluation systems used by districts to evaluate their educators. The last section describes the types of decisions made by district administrators that are informed by the annual year-end evalauations.

Factors of Professional Practice Used in Teacher and Administrator Evaluations

Districts were asked to identify the most common factors used in evaluating elementary, middle, and high school teachers. An analogous set of questions was asked regarding administrator evaluations at each level.

[2] The legislation (MCL 380.1249) also allows districts on a prior contract to delay full implementation of this legislation until the contract expires. Districts were still required to report effectiveness ratings, but the content of those evaluations could be based on the prior system.

The common factors used in evaluating elementary and middle school teachers were examined separately from high school teachers. Not surprisingly, the two most common factors used in teacher evaluations on all levels were instructional practices (including the use of technology), and classroom management. Student achievement was the third most used factor at the elementary and middle school levels and the fourth most-used factor at the high school level, after principal/supervisor evaluations.

Figures 1-2 display the factors used in teacher evaluation.

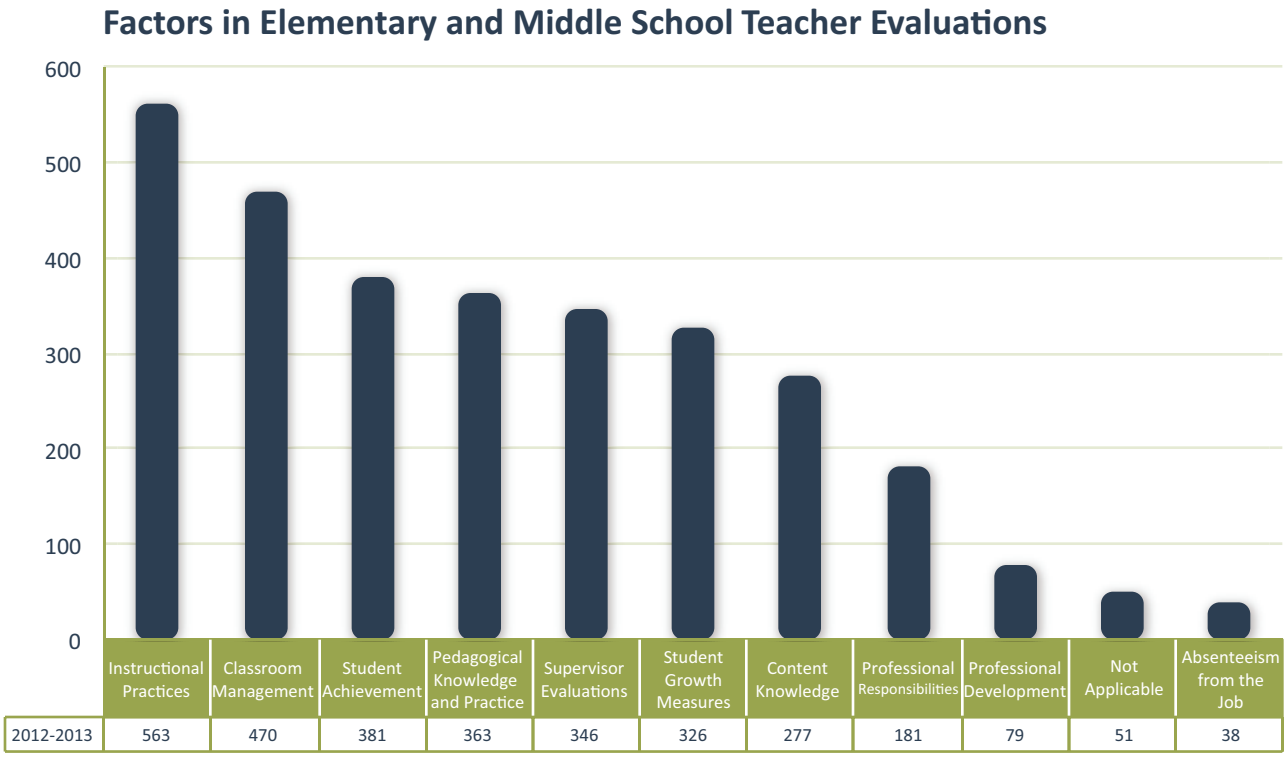


Figure 1: Factors Used in Elementary and Middle School Teacher Evaluations

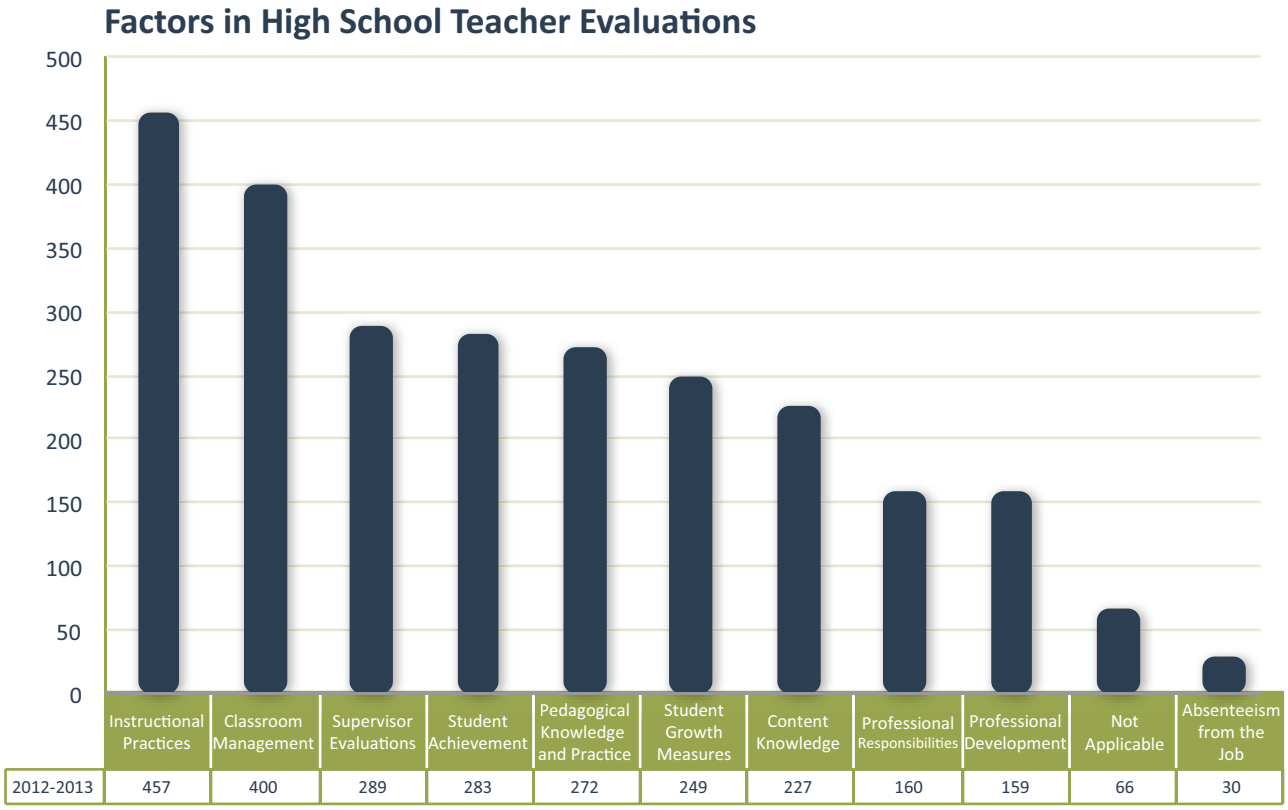


Figure 2: Factors Used in High School Teacher Evaluations

The two most common factors reported in administrator evaluations were instructional practices/leadership (including use of technology) and professional responsibilities. The next two most common factors were growth/decline of student achievement and ability to conduct valid and reliable teacher evaluations. While 285 districts used provision of appropriate support of minimally effective and ineffective teachers (which is required by law) as part of administrator evaluations at the elementary and middle school levels, only 217 did at the high school level.

Figures 3-4 display the factors used in administrator evaluation.

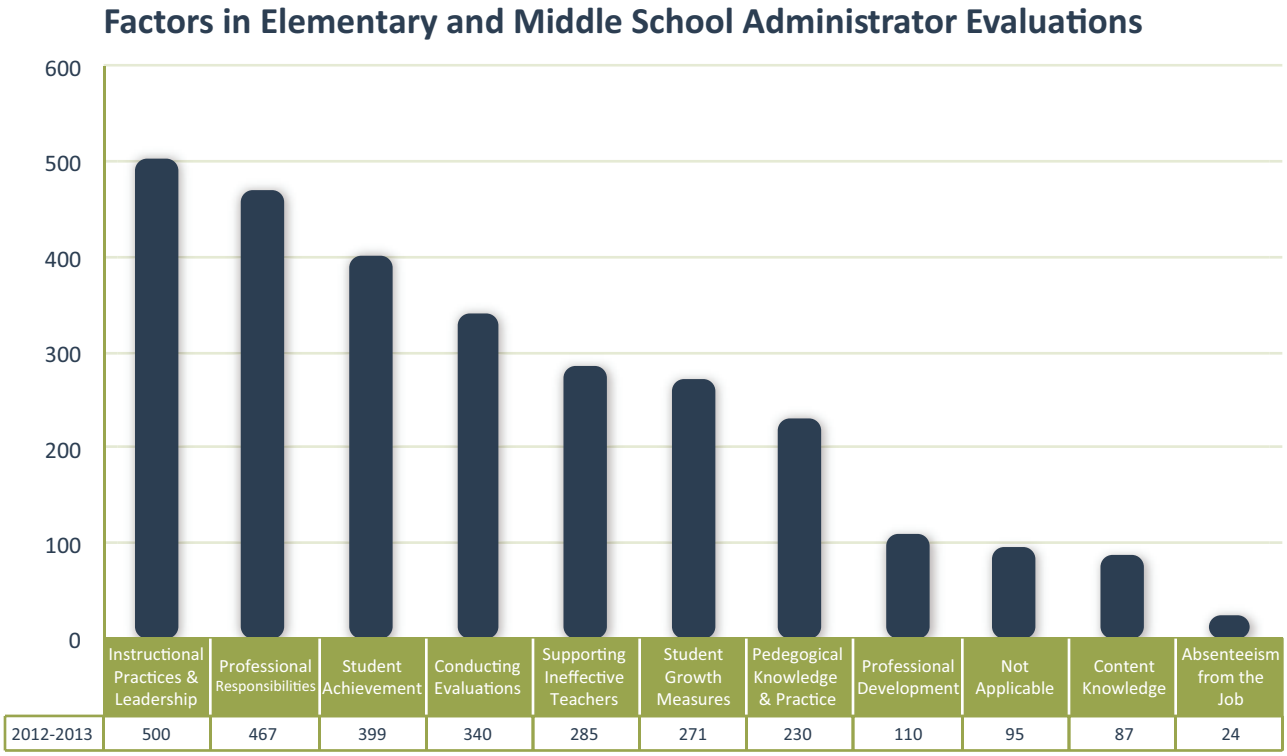


Figure 3: Factors Used in Elementary and Middle School Administrator Evaluations

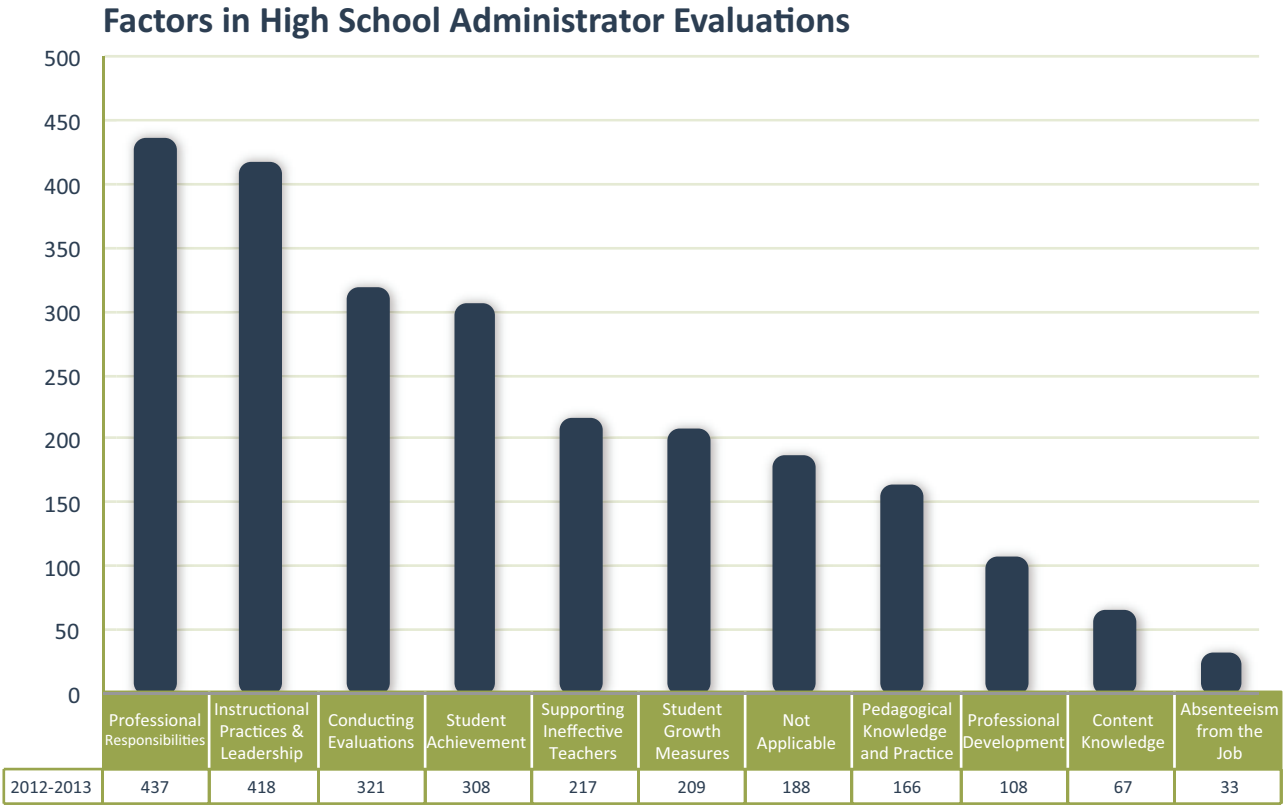


Figure 4: Factors Used in High School Administrator Evaluations

Observation Tools and Frameworks Used to Evaluate Instructional Practice

In 2012-2013, as in 2011-2012, districts were asked to report on the frameworks or tools used as part of their local evaluation system. As Figure 5 below shows, 488 districts across the state of Michigan report using Charlotte Danielson's *Enhancing Professional Practice for Performance of Teaching* or Danielson's *Framework for Teaching Proficiency Test Instrument (Teachscape)* as one of the primary tools for the observation of instructional practice. Many districts reported using Danielson and local measures combined, which is reported in the "Other" category, and whenever Danielson's framework was identified as being among the primary observation tools used, those instances were included in the distribution. Similarly, a district using components of both Danielson's and Marzano's frameworks would be listed as using each method.

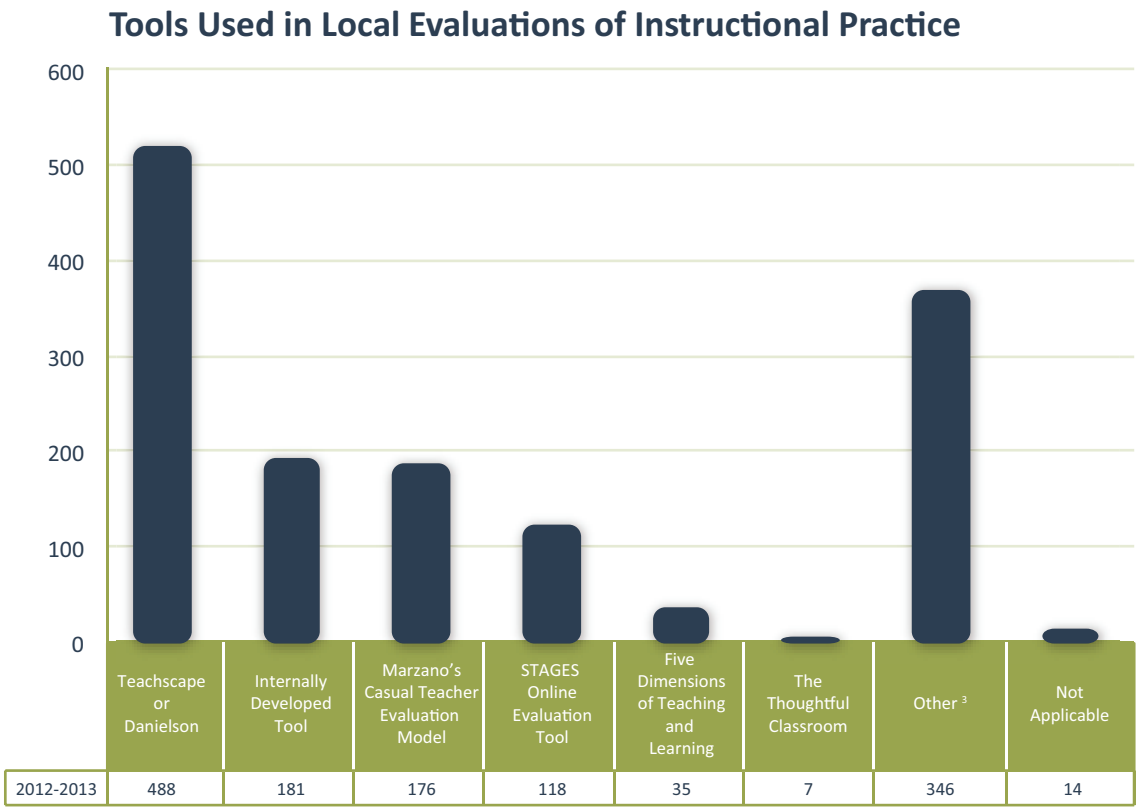


Figure 5: Tools Used in Local Evaluations of Instructional Practice

In 2012-2013, 181 districts reported using an internally developed observation tool, which is an increase from the previous year, when 132 districts reported using an internally developed tool. Internally developed tools, determined by contractual agreements, ranged from a combination of several evidence-based tools and frameworks. Additonallly 346 districts reported using "other" systems.

[3] Among the "other" frameworks reported were: Lenawee ISD's "Framework for Teaching: Supporting Professional Learning," Jackson County ISD's "Effective Evaluation for Educators," Bay-Arenac ISD's "Instructional Leadership Series for Principals and Teacher Leaders," Airport Community Schools' "Evaluation, Collaboration, and Feedback Training to be Consistent to Support Teachers," Clarkston Community Schools' "Educator Evaluation Program," and Imlay City Community Schools' "Training for Observers/Evaluators."

Student Growth Measures Used to Determine Student Growth

Michigan legislation requires that assessments and measures be "reliable and valid," and that student growth be measured in all subjects, not just in mathematics and reading. In both the 2011-2012 and 2012-2013 surveys, districts were asked to indicate which types of assessments they used to determine student growth.

In elementary and middle schools, locally developed common assessments were a frequent option utilized by districts for educator evaluations.[4] In elementary grades, over half of responding districts reported using primarily Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in their evaluations.[5] State assessments are still primarily used to determine student growth in educator evaluations in districts for the grades in which they are available (grades 4-8) (see Figures 6 and 7);[6] however, there is a decreased reliance on state assessments in 2012-2013. In addition to state assessments, most districts base educator evaluations on multiple measures, which is consistent with what we know about best practices in educator evaluation systems.[7] In 2012-2013, 33% of districts report using the Northwest Evaluation Association (NWEA) as a measure of student growth, whereas in 2012, 24% used NWEA.[8] ACT Explore is the most common choice (after state and local common assessments) used for evaluations at the middle school level.

In Figure 6 below, the most common types of assessments mentioned by the elementary and middle school levels are displayed.[9]

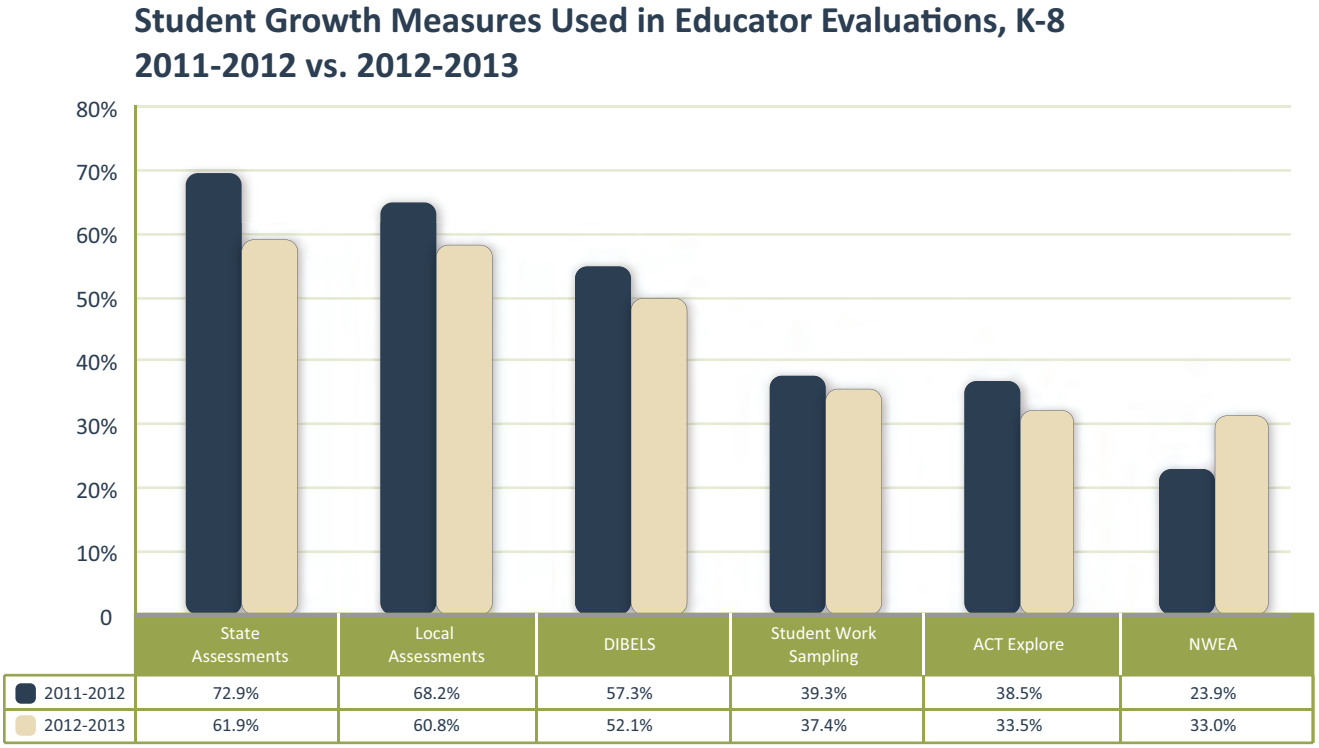


Figure 6: Types of Assessments Used to Determine Student Growth in Grades K-8, 2011-2012 vs. 2012-2013 Comparison

[4] In 2012-2013, 353 districts at the early elementary level (K-1), 355 districts at the elementary level (2-5), and 409 districts at the middle school level (6-8) used some kind of common pre-post assessment.

[5] In 2012-2013, 391 districts at the early elementary level (K-1), 301 districts at the elementary level (2-5), and 69 districts at the middle school level (6-8) used used a locally developed common assessment..

[6] In 2012-2013, 17 districts at the early elementary level (K-1), 398 districts at the elementary level (2-5), and 448 districts at the middle school level (6-8) used DIBELS.

[7] See A Practical Guide to Designing Comprehensive Teacher Evaluation Systems: A Tool to Assist in the Development Teacher Evaluation Systems, published by the National Comprehensive Center for Teacher Quality (2011), American Institutes for Research.

[8] In 2012-2013, 213 districts at the early elementary (K-1) level, 238 districts at the elementary (2-5) level, and 233 districts at the middle school level (6-8) used NWEA.

[9] Due to differences in the sets of response options provided in the 2011-2012 and 2012-2013 surveys, we were unable to make a district comparison for the secondary level factors. Frequencies lower than 100 were not included in these displays.

At the high school level, 419 districts report using state assessments (namely MEAP and MME), while 368 use the ACT Plan, and 226 use the ACT (college entrance exam). It is worth noting that districts report using these summative assessments as indicators of student growth. However, these summative assessments provide only a single data point, while true measures of student growth require two or more data points.

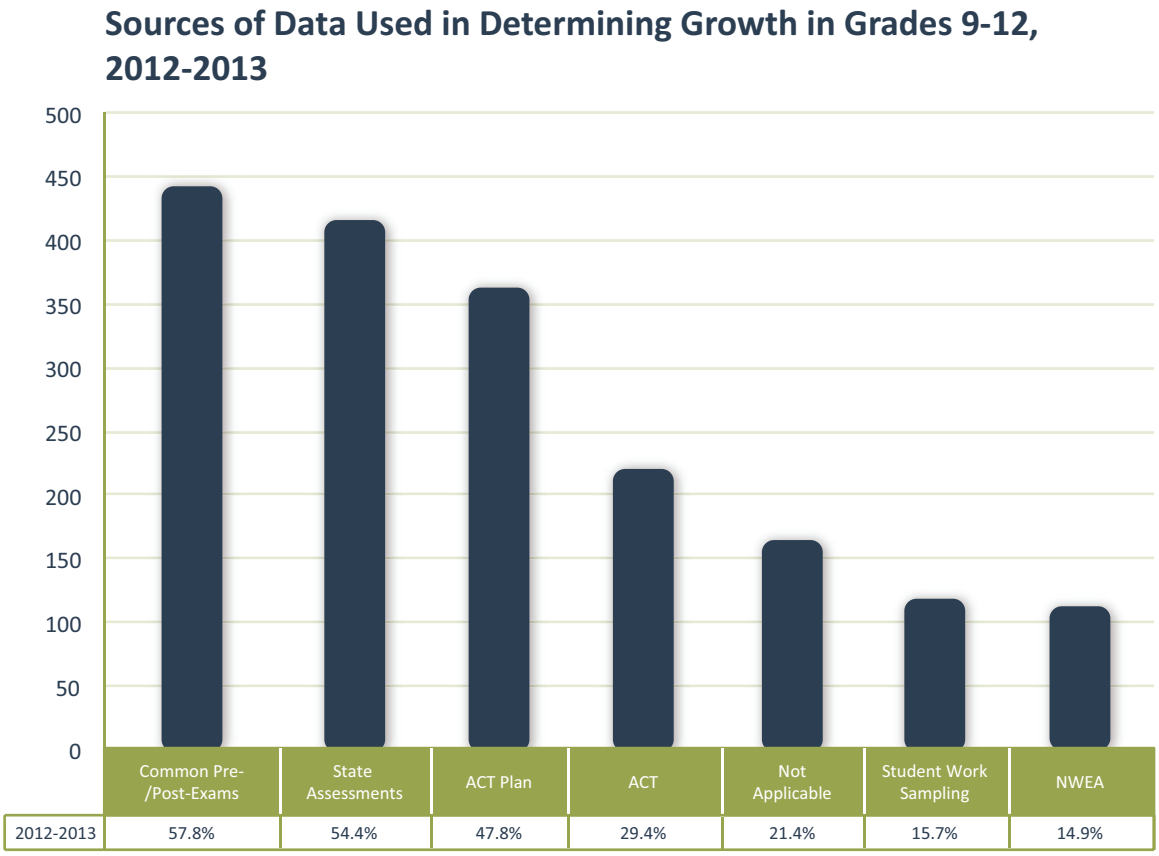


Figure 7: Types of Assessments Used to Determine Student Growth in Grades 9-12 in 2012-2013.

Reported Percentage of Student Growth Component in Local Evaluation Systems

In the 2011-2012 and 2012-2013 school years, the educator evaluation state legislation (MCL 380.1249) required that districts include student growth and assessment as a significant part of the annual year-end evaluation.[10] Figure 8 (on the following page) displays the percentage of student growth component in local evaluation systems over a two-year period. In 2012-2013, 69.1% (526) of the districts reported basing 20% or more of their evaluations on student growth, an increase from the previous year's survey, when only 49.4% (386) of districts reported basing 20% or more of their annual evaluation results on student growth data. This increase is to be expected as districts make progress towards implementing the new requirements. It should also be noted that 24.0% of the districts reported that 40% or more of their year-end evaluations are based on student growth data. Some districts (7.8%) report that student growth data are not yet used in local evaluations, and a few districts (1.3%) did not respond to this question on the survey.

[10] As defined by federal policy, student growth means a change (usually one grade level in an academic year) for an individual student between two or more points in time. (Secretary's Priorities for Discretionary Grant Priorities, 2010, p. 47290)

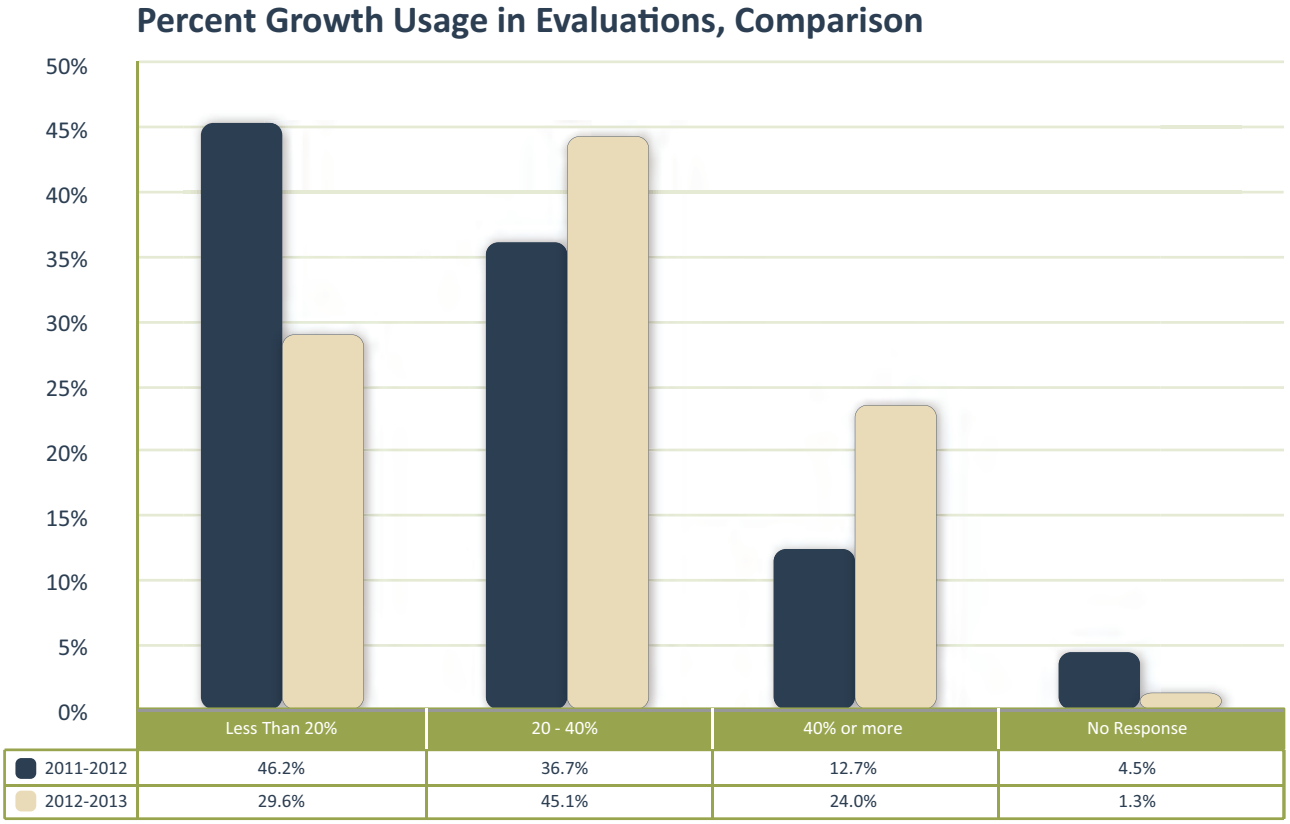


Figure 8: Percentage of Student Growth Component in Local Evaluation Systems in 2011–2012 and 2012–2013

Decisions Informed by Evaluation Results

Districts were also asked to indicate how they use the results of the evaluations (see Figure 9 on the following page). Over 600 districts indicated they use the results to determine the professional development needs and requirements for educators, which was one of the key desired outcomes of the state law (MCL 380.1249). Additionally, over 500 districts used the results to inform individualized development plans or to drive coaching efforts for teachers. This is an important positive step and represents an increase in the amount of professional development, instructional and leadership coaching, and support that Michigan educators are receiving.

In contrast to the 2011-2012 survey results, where over 400 districts used the evaluation results to provide induction support for new teachers in 2012-2013, only 119 districts report using the evaluation results to inform new teacher training and support. Further, in 2012-2013, 475 districts report using evaluation results to recommend the removal of teachers after providing time and resources for improvement. Even though districts are primarily concerned with providing quality coaching and professional development support and resources, many report using the evaluation process as a way of removing “ineffective” teachers from the classroom.

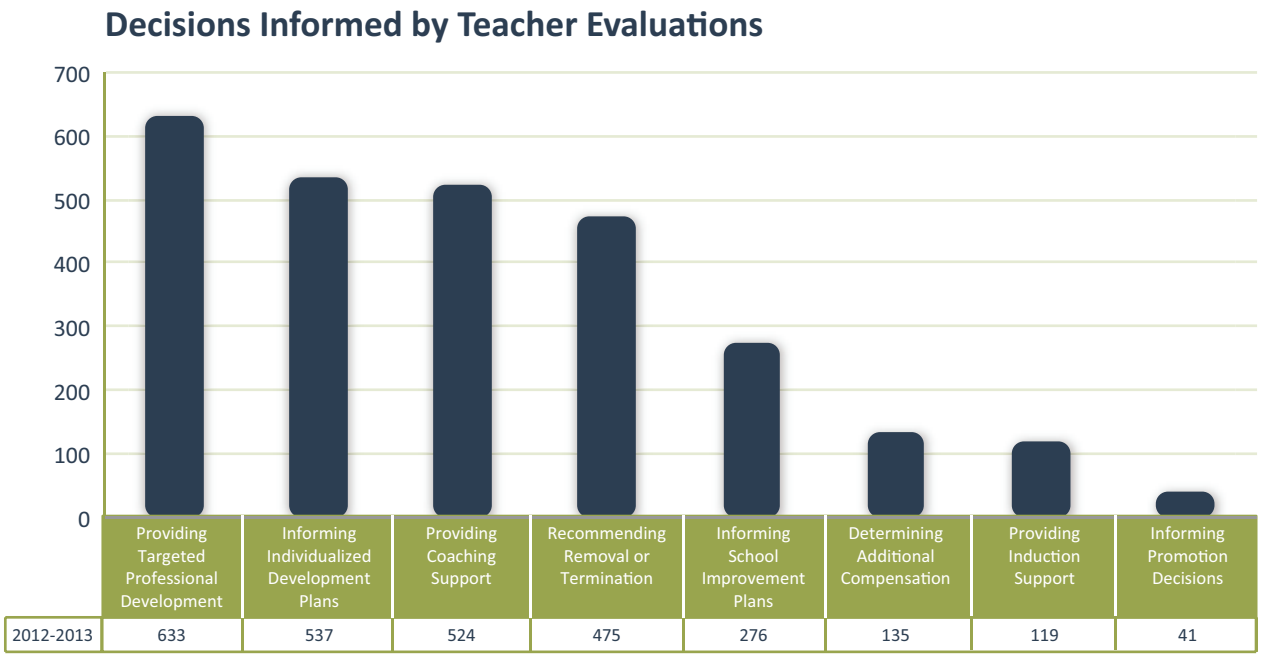


Figure 9: Decisions Informed by Evaluation Results Across Michigan Districts (Teachers)

Common decisions based on administrator evaluations were providing leadership coaching support, determining types of professional development, and informing school improvement plans (see Figure 10). The third most common decision, however, was the recommendation for removal or termination, with 437 districts reporting. Another surprising finding was that districts were less likely to use the evaluations to inform professional development support for new teachers and administrators than to use the evaluations for termination and/or removal.

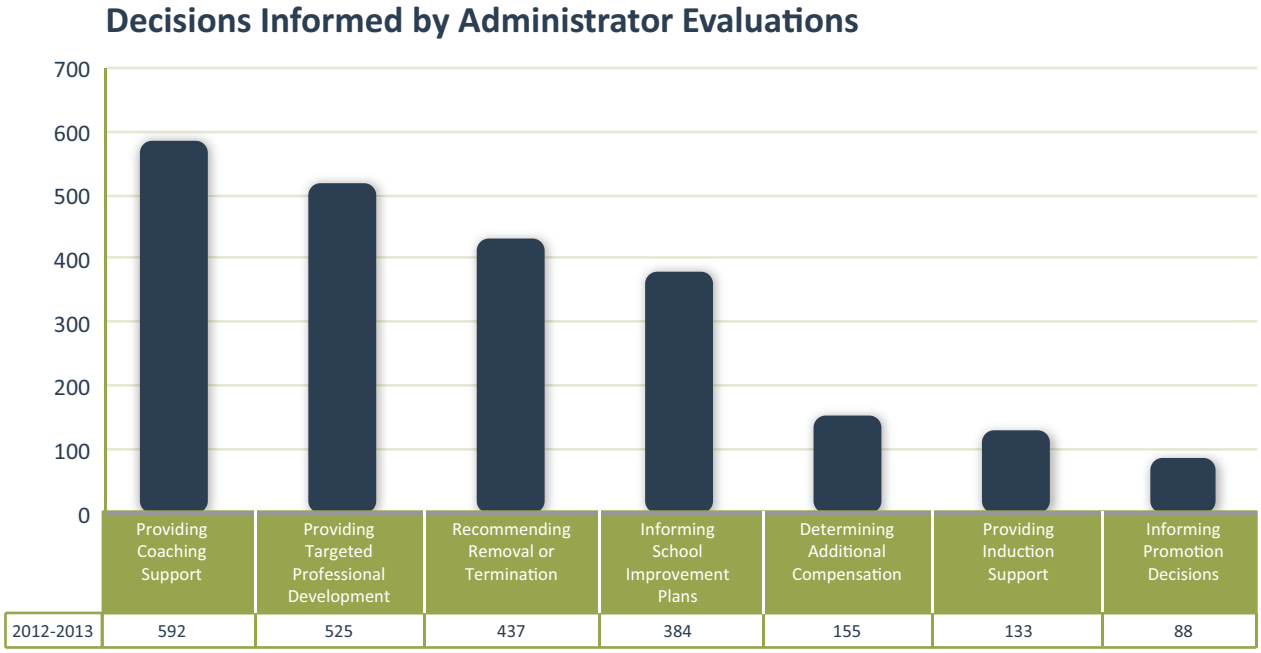


Figure 10: Decisions Informed by Evaluation Results Across Michigan Districts (Administrators)

STATEWIDE DISTRIBUTION OF EDUCATOR EFFECTIVENESS RATINGS[11]

An analysis of Registry of Education Personnel (REP) data as it relates to educator evaluation system characteristics and accountability measures was completed. For the purposes of this brief, an analysis was conducted to determine how closely related educator effectiveness is to the amount (percentage) of student growth measures used in evaluations.[12] Analysis of the relationship between educator effectiveness and school accountability ratings (Reward, Focus, and Priority designations) and of teacher characteristics as they relate to educator effectiveness ratings were also conducted.

Although ratings are not directly comparable across districts, it is important to understand information about ratings statewide, keeping in mind that ratings were based on local evaluation systems.[13]

In Figure 11, the statewide distribution of teachers in each of the four effectiveness ratings is presented.

- In comparing 2011-2012 ratings with 2012-2013, more teachers were reported as being "highly effective" in 2012-2013, while fewer teachers were reported as being "effective" as determined by their local evaluation systems.
- Overall, 97.0% of Michigan teachers were reported as "effective" or "highly effective" as determined by their local evaluation systems.
- 2.4% of Michigan teachers were reported as "minimally effective" as determined by their local evaluation systems.
- 0.6% of Michigan teachers were reported as "ineffective" as determined by their local evaluation systems.

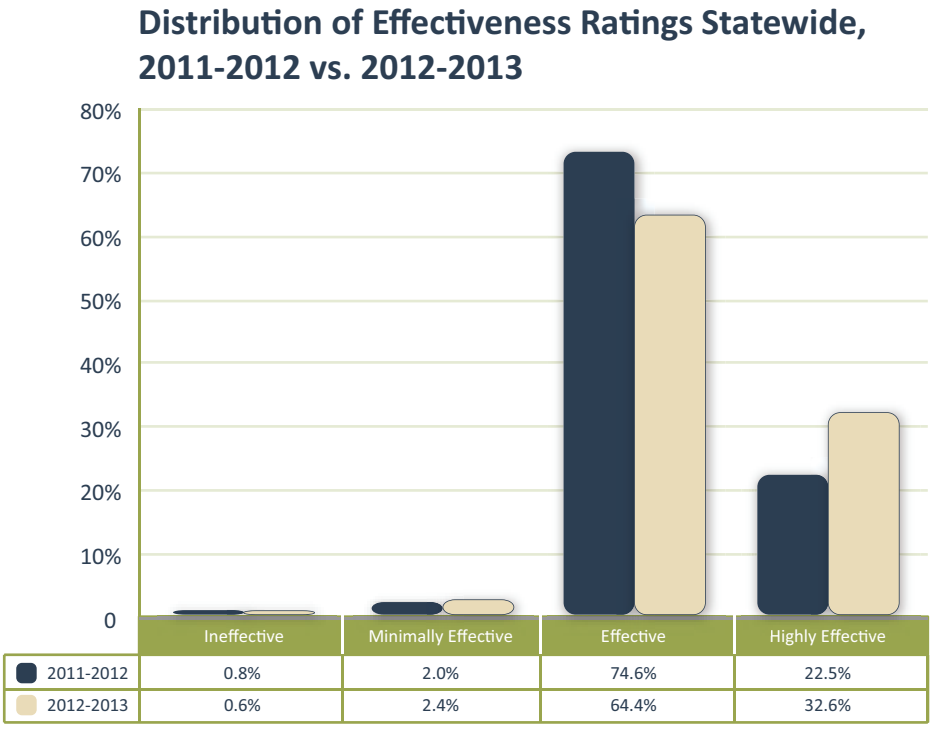


Figure 11: Percent of Michigan Teachers Reported in Each of the Four Effectiveness Ratings

[11] At the present, MDE does not endorse or identify one particular definition of "educator effectiveness" and the corresponding rating levels, although an agreed-upon definition would be helpful in creating and implementing an educator effectiveness policy. It is recommended that a common definition be identified from research and best practices and subsequently adopted by the state of Michigan to strengthen the reliability and validity of the educator evaluation system.

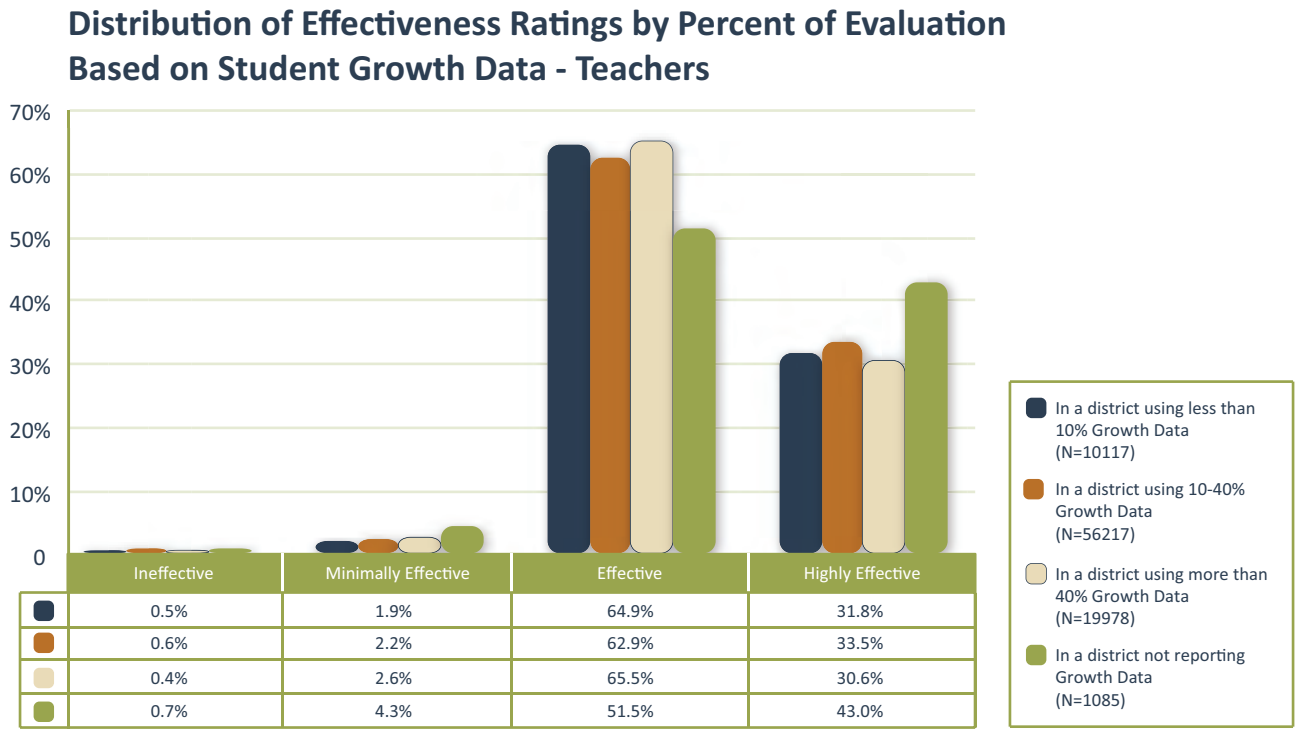
[12] In 2011-2012 and 2012-2013, the educator evaluation legislation required that districts include student growth and assessment as a "significant part" of the annual year-end evaluations.

[13] Revised School Code Act 451 of 1976 Section 380.1249 provides for some local control over the evaluation tool used to evaluate teachers.

Distribution of Effectiveness Ratings By Weighting of Student Growth in Evaluations

Because student growth provides what is perceived as an objective measure in the evaluation system, it is hoped to yield a more even distribution of effectiveness ratings and a more realistic indicator of teacher effectiveness.[14] In the second year of implementation of the educator evaluation legislation (MCL 380.1249), as the percentage of the evaluation based on student growth increased, the number of teachers reported as "effective" (the most common category) decreased and the number of teachers reported as "highly effective" increased. As stated earlier in this brief, districts overall reported a 20% increase in the amount of student growth data incorporated into teacher and administrator evaluations, but reported little change in the number of teachers rated "effective" or "highly effective."

Figures 12 (below) and 13 (following page) illustrate the distribution of effectiveness ratings by amount of student growth used in year-end evaluations for teachers and administrators, respectively.



	Ineffective	Minimally Effective	Effective	Highly Effective
In a district using less than 10% Growth Data (N=10117)	0.5%	1.9%	64.9%	31.8%
In a district using 10-40% Growth Data (N=56217)	0.6%	2.2%	62.9%	33.5%
In a district using more than 40% Growth Data (N=19978)	0.4%	2.6%	65.5%	30.6%
In a district not reporting Growth Data (N=1085)	0.7%	4.3%	51.5%	43.0%

Figure 12: Distribution of Effectiveness Ratings by Percent of Evaluation Based on Student Growth—Teachers

In the 2012-2013 survey results, there are some findings worth noting:

- 1,085 teachers are in districts that did not provide information on the percentage of growth data used in their evaluation system.
- The percentage of student growth data used in year-end teacher evaluations appears to have little or no relationship to a rating of "ineffective."
- The percentage of student growth data incorporated in year-end teacher evaluations appears to have little or no relationship to the likelihood a teacher would be found "minimally effective," *unless the district did not respond to this survey item*.
- The percentage of student growth data incorporated in year-end teacher evaluations appears to have little or no relationship to the likelihood that a teacher would be found "effective" or "highly effective." However, it is less likely that a teacher would be rated "effective" and more likely that a teacher would be rate "highly effective" *if the district did not respond to this survey item*.

[14] Note: MDE does not hold the position that student growth data are more objective or reliable in educator evaluations than other factors such as observations. We do, however, recognize that student growth data are often *perceived* to be more concrete and defensible when explaining the results of an evaluation decision. We also expect that as more principals are trained to use the approved evaluation tools, the perception of other factors as "subjective," and therefore less reliable, will greatly diminish.

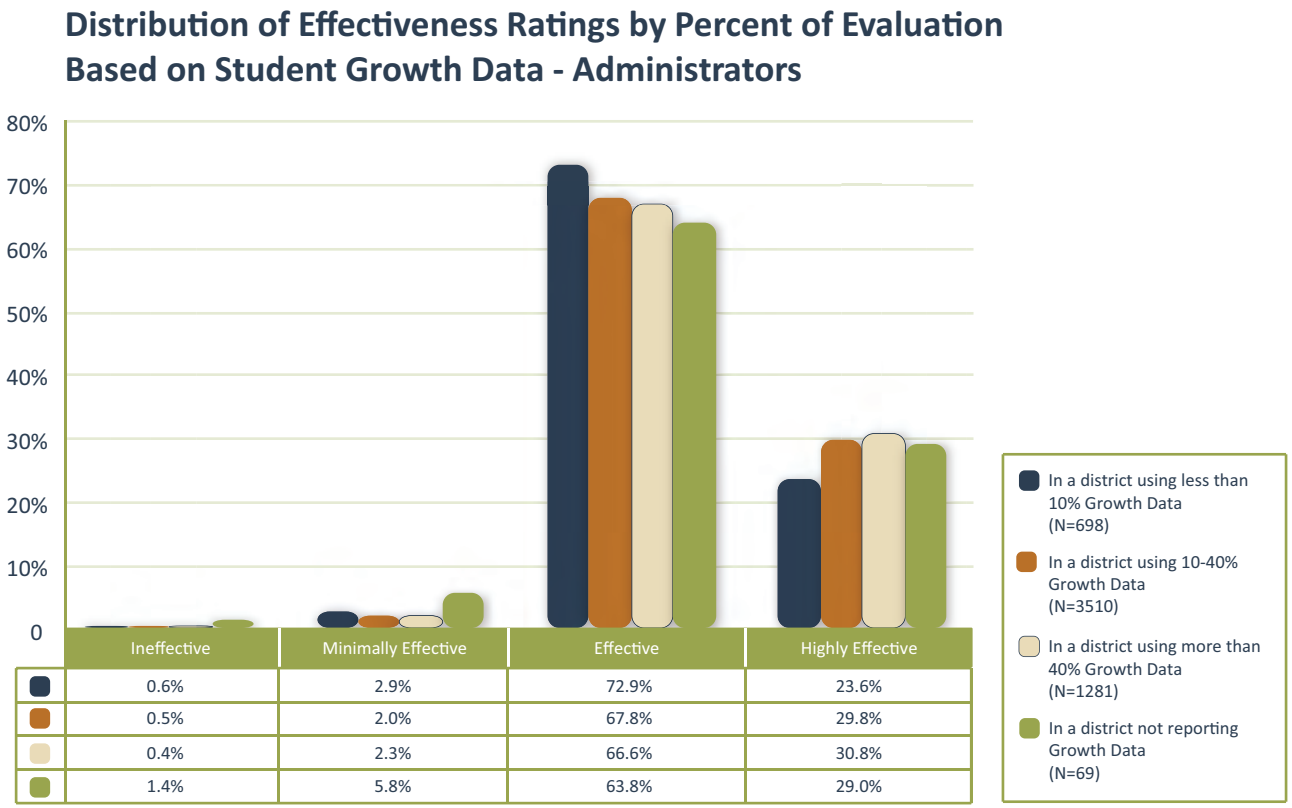


Figure 13: Distribution of Effectiveness Ratings by Percent of Evaluation Based on Student Growth—Administrators

For administrators, a slightly different picture emerges in relation to the amount of student growth data used in administrator evaluations. Some key takeaways are:

- 69 Administrators are employed in districts that did not provide information on the percentage of student growth data.
- In both "ineffective" and "minimally effective" categories, the amount of student growth data used in year-end evaluations appears to have little relationship to the assignment of administrator effectiveness ratings, *among districts reporting the percent attributed to growth data*. Since the number of administrators employed in districts that did not respond to this survey item is low, its relationship is probably insignificant.
- The percentage of student growth data used in year-end administrator ratings appears to be negatively related to the likelihood that an administrator would be rated "effective." It appears to be less likely that an administrator would be rated as "effective" if their district did not respond to this survey item, but since the number of administrators employed in such districts is low, the relationship is probably insignificant.
- Conversely, in the "highly effective" category, the increased use of data appears to increase the likelihood that an administrator will be assigned a "highly effective" rating.

Priority, Focus and Reward Schools: Teacher and Administrator Effectiveness Ratings

Teacher and administrator effectiveness ratings and school-level accountability are not the same thing. For example, it is possible for a school to have low overall levels of student proficiency but for many teachers or administrators within that school to have positive student growth results and be "effective" in their roles. Teacher and administrator effectiveness ratings are based on the ability to move students forward and to help them show growth. School-level accountability is based on a mixture of achievement (which is strongly related to prior achievement levels), growth, and achievement gaps. Therefore, it cannot be assumed there is a causal relationship between these two metrics.

Given these concerns, MDE did analyze the teacher and administrator effectiveness ratings in our three categories of schools: Priority, Focus, and Reward.[15] The results are presented in Figures 14 (below) and 15 (following page).

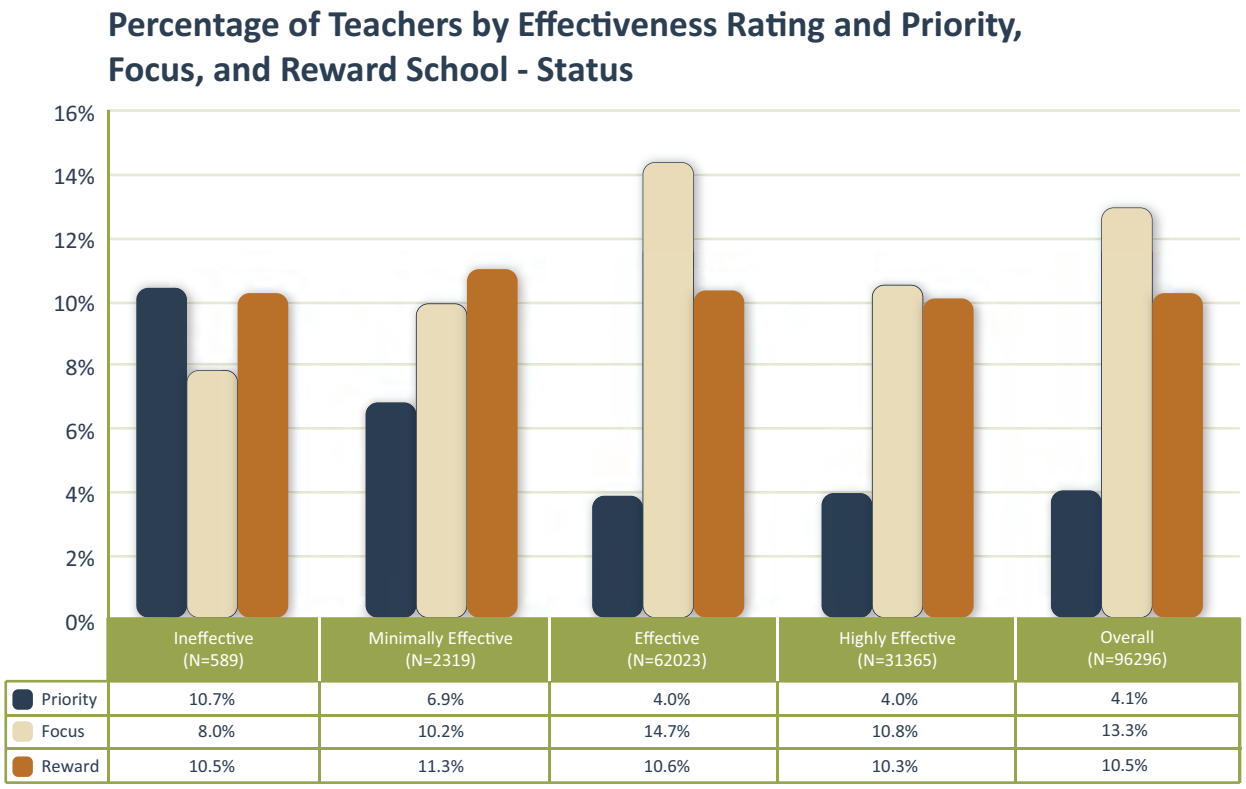


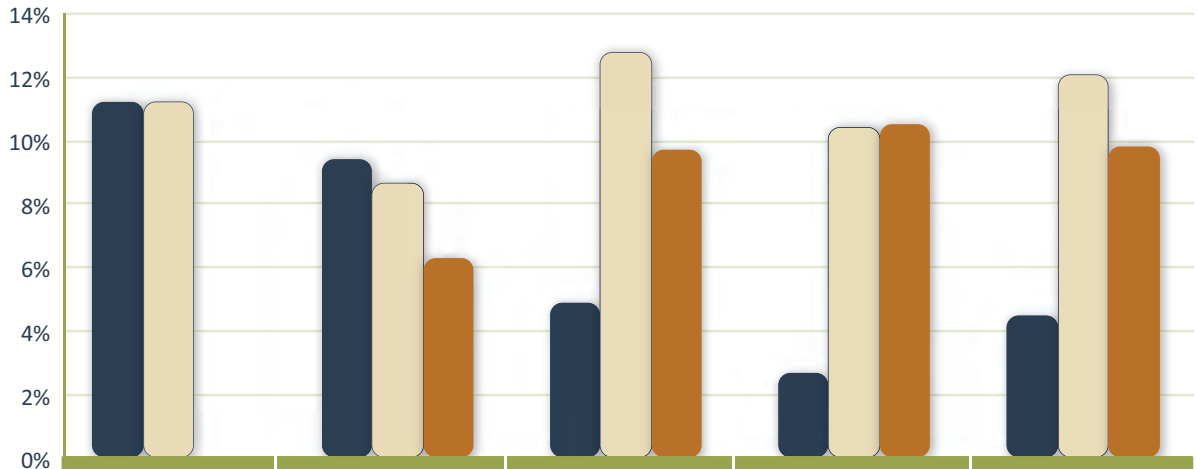
Figure 14: Percentage of Teachers by Effectiveness Rating and Priority, Focus and Reward Schools

As indicated in Figure 14, 10.7% of "ineffective" teachers are located in Priority schools—a rate two and a half times higher than their share of the overall teacher pool. Similarly, "minimally effective" teachers are overrepresented at Priority schools by over 70%. Despite this, Priority schools have numbers of "effective" and "highly effective" teachers roughly in line with their overall share of the state's teacher population. It must be stressed that no causal relationship exists in these data. It cannot be inferred that any Priority school will have more "ineffective" teachers than any other school, or that schools are given the Priority status because they contain more "ineffective" teachers than any others. This may be a result of the requirements placed on districts with Priority Schools—these districts must implement a transformation model for school improvement and are required to undertake certain staffing changes after being named to the Priority List.

"Ineffective" teachers, meanwhile, are half as likely to appear in Focus schools as are teachers selected at random. The percentages of "ineffective," "minimally effective," "effective," and "highly effective" teachers in Reward schools appear to be roughly the same as that for all teachers.[16]

[15] During the 2012-2013 school year there were 2,561 schools that did not receive a school status and therefore are not reflected in these analyses.

[16] The fact that Reward schools contain relatively high proportions of "ineffective" teachers is a puzzle. One possible explanation is that Reward schools are better able to identify "ßineffective" teachers than are schools in general. Another is that Reward schools may have stronger teacher retention policies. Many other possibilities exist; MDE does not presently take a stance on what exactly causes this relationship.



	Ineffective (N=26)	Minimally Effective (N=124)	Effective (N=3420)	Highly Effective (N=1307)	Overall (N=4877)
Priority	11.5%	9.7%	5.0%	2.8%	4.6%
Focus	11.5%	8.9%	13.1%	10.7%	12.4%
Reward	0.0%	6.5%	10.0%	10.8%	10.1%

Figure 15: Percentage of Administrators by Effectiveness Rating and Priority, Focus and Reward Schools

Similarly to the previous analysis,[17] "ineffective" and "minimally "effective" administrators are disproportionately likely to be employed at Priority schools. Conversely, Reward schools appear to have no "ineffective" administrators, and are more likely to employ administrators who are rated "effective" and "highly effective." Again, no causal relationship can be inferred, but it is expected that schools that are doing better than predicted (Beating the Odds schools), or are demonstrating growth overall, may have more "effective" leadership in place than schools that are persistently low achieving or have wide achievement gaps. In comparison to teacher effectiveness, administrator effectiveness appears to be more closely related to school accountability, and matters most in Reward and Focus schools.

Understanding Teacher Characteristics and Evaluation Ratings

In order to better understand the correlations between teachers' characteristics and their effectiveness ratings, MDE performed an analysis that allowed us to predict the likelihood of a teacher appearing in various effectiveness categories. Please note that this analysis only describes relationships—it does not attempt to determine what *causes* different effectiveness ratings. Similarly, MDE does not take any stance on why certain traits are correlated with higher effectiveness ratings. We use only the data available in state systems. Finally, school and district characteristics are not taken into account. This analysis therefore is useful in providing a basic picture of teacher effectiveness, but should be viewed strictly as *descriptive* rather than *prescriptive*.[18]

[17] During the 2012-2013 school year there were 2,561 schools that did not receive a school status and therefore are not reflected in these analyses.

[18] As a technical note, the following analysis is based on an ordered logistic regression. This model allows us to take data with a ranked set of categories and determine the predictors of appearing in a given category. Its key feature is that while we are able to rank the different effectiveness categories, we cannot quantify exactly how "good" a "highly effective" teacher is or how "bad" an "ineffective" teacher is using the effectiveness ratings alone. By looking at the odds of appearing in any given category, ordered logistic regressions are able to sidestep this issue. Appendix B contains detailed logistic regression output; additional output is available upon request.

Who is more or less likely to be rated a "highly effective" in Michigan?[19]

The analysis includes several sets of descriptors. The first set—gender, age, and minority status—controls for teachers' demographic characteristics. The second contains the number of years a teacher has spent in a district, whether he or she is a "new" teacher (defined as having three or fewer years in the district), and an interaction term designed to measure whether newer teachers have different returns to experience (i.e., whether they face an early learning curve). A third set of descriptors looks at a teacher's qualifications—whether he or she is professionally certified, has a Master's or higher degree, majored in his or her area of certification, or is certified in an area where a major does not apply. A fourth set of descriptors looks at various subject assignments and at whether a teacher has full or part-time status. In examining the relationship between subject area and effectiveness ratings, we use elementary education as the omitted category (determining whether teachers in any particular subject are more likely to get higher ratings than elementary teachers are).

It appears that a female teacher (holding all else equal) is more likely to receive a "highly effective" rating. The same appears to be true if the teacher is minority, if she or he has a longer district tenure, if she or he is professionally certified, if she or he holds a Master's degree or higher, or if she or he has a full time assignment. ELA teachers and art teachers are more likely to receive higher ratings than elementary teachers in all subject areas, while teachers of mathematics, science, social science, special education, and world languages appear to have lower effectiveness ratings than elementary teachers. New teachers appear to get more of a boost from additional time in their district than experienced teachers do—this could suggest either that there is a substantial learning curve for new teachers or that the first several years are instrumental in determining who is "effective" and who is not. Similarly, older teachers are less likely to be rated "highly effective," but experienced teachers who have taught in the same district for a number of years are more likely to be given a highly effective rating. In addition, teachers who hold a major in their certification area or who are in an area without a corresponding major appear to be less likely to be given a "highly effective" rating than teachers who are outside of their major. A likely explanation is that the teachers who teach outside of their major may be more effective to begin with—if a school has an area of need but lacks a teacher who is certified (or is endorsed) in that area, it is likely that they would fill that need with a more "effective" teacher, assuming that pedagogical skill may make up for a lack of specific content knowledge.

The results examine correlations *holding all else equal*. This fact may explain why minority teachers are more likely to receive higher effectiveness ratings than white teachers while also being overrepresented among "minimally effective" and "ineffective" teachers.[20] Minority teachers are less likely to be professionally certified than white teachers, for instance, which would work to counteract the "boost" we observe in our regressions.[21]

[19] Based on local evaluation systems and 2012-2013 data

[20] Of the evaluated teachers, 74.3% are female, while 91.6% are White.

[21] We ran separate regressions by minority status (not shown here) to determine whether certain factors are more strongly correlated with effectiveness ratings among different subgroups. We find that possession of professional certification or of a Master's degree has a very strong impact on minority teachers' effectiveness ratings.

APPENDIX A: EDUCATOR/ADMINSTRATOR EVALUATION SYSTEMS SURVEY

K-12 Educator/Administrator Evaluation Systems

WELCOME TO THE MICHIGAN DEPARTMENT OF EDUCATION'S
K-12 EDUCATOR/ADMINISTRATOR EVALUATION SYSTEMS SURVEY

o Bureau of Assessment and Accountability
o Center for Educational Performance and Information
o Office of Educational Improvement and Innovation
o Office of Professional Preparation Services

INTRODUCTION

This survey is designed to collect information about your district's educator and administrator evaluations. It is critically important districts respond to this survey in a timely manner in order to help the Michigan Department of Education (MDE)comply with Federal requirements. The survey also helps the MDE understand how districts are conducting evaluation and where the MDE might provide strategic technical support and information.

To review and/or download the survey prior to completing, copy and paste the following URL into any browser:

www.michigan.gov/documents/mde/2012-13_K-12_Educator_Administrator_Evaluation_Systems_Survey_418516_7.pdf

Page 1

K-12 Educator/Administrator Evaluation Systems

Please provide the following demographic information.

***1. Choose your District Name/Code:**

***2. Your Name:**

***3. Position/Title:**

- District Superintendent
- District Assistant Superintendent
- District level designee

Page 2

K-12 Educator/Administrator Evaluation Systems

4. Which, if any, of the following systems, frameworks, or methods are your local evaluations mostly based on? Please check UP TO FOUR of the following:

- ☐ Not Applicable
- ☐ A Framework for Teaching: Supporting Professional Learning (Lenawee ISD)
- ☐ Danielson's Framework for Teaching Proficiency Test Instrument (Teachscape)
- ☐ Enhancing Professional Practice for Performance of Teaching (Danielson)
- ☐ Marzano's Causal Teacher Evaluation Model
- ☐ Teacher Supervision and Evaluation Framework (Kim Marshall)
- ☐ 5 Dimensions of Teaching and Learning (University of Washington Center for Educatic adership)
- ☐ The Thoughtful Classroom Teacher Effectiveness Framework (Silver Strong Associates)
- ☐ Teacher Evaluation System(s) CUES Model or Standard-Based Model (McREL)
- ☐ Professional Standards for Michigan Teachers (PSMT)
- ☐ Educator Evaluation: Together We Make Each Other Better (MASSP)
- ☐ STAGES online evaluation tool
- ☐ Effective Evaluation for Educators (JCISD)
- ☐ Educator Evaluation Program (Clarkston Community Schools)
- ☐ Evaluation Collaboration and Feedback Training to be Consistent and Support Teachers (Airport Community Schools)
- ☐ Instructional Leadership Series for Principals and Teacher Leaders (Bay-Arenac ISD)
- ☐ Supporting Teacher Growth Through Evaluation (KISD)
- ☐ Training for Observers/Evaluators (Imlay City Community Schools)
- ☐ Peer Review
- ☐ Internally Developed Tool / or Other (please specify)

5. Does the district conduct different evaluations for teachers based on content area and/or grade level taught?

- ○ Yes
- ○ No

Page 3

K-12 Educator/Administrator Evaluation Systems

6. For any evaluations made public, how are results mostly reported? Please check ONE of the following:

- ◯ Not applicable
- ◯ Results are not made public
- ◯ On the agency's website
- ◯ REP reporting
- ◯ Written notice to the general public
- ◯ Annual written report
- ◯ District Board meeting
- ◯ Other, please specify

Page 4

K-12 Educator/Administrator Evaluation Systems

The following questions pertain to **STUDENT GROWTH MEASURES** within evaluations.

7. How is student growth data mostly used in evaluations? Please check TWO of the following:

- ☐ Not yet used
- ☐ Statistics from State test data
- ☐ In a prescribed way (e.g.formula driven)
- ☐ Evidence from local assessments (interim and formative assessments)
- ☐ Subjective rating from supervising administrator
- ☐ Multiple measures used
- ☐ Other, please specify

8. What percentage of evaluations is based on student achievement growth data?

- ○ 0%
- ○ <10%
- ○ 10 - 19%
- ○ 20 - 29%
- ○ 30 - 39%
- ○ 40 - 49%
- ○ 50 or >%

9. The State reports for each student of grades 4-8 a Performance Level Change (a measure of student growth) in reading and mathematics on MEAP and MI-Access FI. Does your district make use of the Performance Level Change (PLC) designation by the State for the purpose of educator evaluations?

- ○ Yes
- ○ No

10. Did your district make use of the MDE weighted PLC tool for measuring improvement/growth?

- ○ Yes
- ○ No

K-12 Educator/Administrator Evaluation Systems

11. Which sources of assessment data are mostly used for determining student growth at the early elementary (grades K-1) level? Please check UP TO FOUR of the following:

- ☐ Not Applicable
- ☐ Locally developed common assessments
- ☐ Northwest Evaluation Association (NWEA)
- ☐ Diagnostic Reading Assessments (DRA)
- ☐ AIMSweb
- ☐ Scholastic Reading Inventory (SRI)
- ☐ DIBELS
- ☐ Running Records
- ☐ Star Reading and Math
- ☐ Scantron Performance Series
- ☐ Fountas & Pinnell Leveled Literacy Intervention
- ☐ Student work sampling
- ☐ Curriculum-based assessment (CBA)
- ☐ Other, please specify

Page 6

K-12 Educator/Administrator Evaluation Systems

12. Which sources of assessment data are mostly used for determining student growth at the elementary (grades 2-5) level? Please check UP TO FOUR of the following:

- ☐ Not Applicable
- ☐ State assessments (in grades 4-5)
- ☐ Locally developed common assessments
- ☐ Northwest Evaluation Association (NWEA)
- ☐ Diagnostic Reading Assesments (DRA)
- ☐ Other Norm-referenced assessment
- ☐ AIMSweb
- ☐ Scholastic Reading Inventory (SRI)
- ☐ Discovery Education
- ☐ Star Reading and Math
- ☐ Scantron Performance Series
- ☐ Fountas & Pinnell Leveled Literacy Intervention
- ☐ DIBELS
- ☐ Student work sampling
- ☐ Curriculum-based assessments (CBA)
- ☐ Other, please specify

K-12 Educator/Administrator Evaluation Systems

13. Which sources of assessment data are mostly used for determining student growth at the middle school (grades 6-8) level? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] State assessments
- [ ] Locally developed common assessments
- [ ] Northwest Evaluation Association (NWEA)
- [ ] Other Norm-referenced assessment
- [ ] AIMSweb
- [ ] Scholastic Reading Invenetory (SRI)
- [ ] Discovery Education
- [ ] Star Reading and Math
- [ ] Scantron Performance Series
- [ ] DIBELS (through grade 6)
- [ ] Student work sampling
- [ ] ACT Explore
- [ ] Other, please specify

Page 8

K-12 Educator/Administrator Evaluation Systems

14. Which sources of assessment data are mostly used for determining student growth at the secondary (grades 9-12) level? Please check UP TO FOUR of the following:

- ☐ Not Applicable
- ☐ Common pre- and post-assessments
- ☐ End of course common assessments
- ☐ Common interim assessments
- ☐ Northwest Evaluation Association (NWEA)
- ☐ Other norm-referenced assessment
- ☐ Student work sampling
- ☐ Scantron Performance Series
- ☐ ACT Plan
- ☐ ACT College Entrance Exam
- ☐ MME
- ☐ MEAP (9th grade Social Studies only)
- ☐ Other, please specify

15. For which subject areas are local measures of student growth determined? (check all that apply)

- ☐ Not Applicable
- ☐ Reading
- ☐ Writing
- ☐ Mathematics
- ☐ Science
- ☐ Social Studies
- ☐ Fine Arts
- ☐ World Languages
- ☐ Health/Physical Education
- ☐ Family and Consumer Science
- ☐ Career and Technical Education
- ☐ Other, please specify

K-12 Educator/Administrator Evaluation Systems

16. For which grades have measures of student growth been determined/developed? (check all that apply)

- ☐ K
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8
- ☐ 9
- ☐ 10
- ☐ 11
- ☐ 12

17. If you would like to provide additional information about how student growth is measured and incorporated into evaluations, please do so here.

Page 10

K-12 Educator/Administrator Evaluation Systems

The following questions pertain to College and Career Readiness (CCR).

18. Do you have a locally defined measure of CCR?

- ◯ Yes
- ◯ No

Page 11

K-12 Educator/Administrator Evaluation Systems

19. Please indicate whether the locally defined measure for CCR impacts your educator and adminstrator evaluations.

- ◯ Yes, it impacts our evaluations
- ◯ No, it does not impact our evaluations

20. How is the locally defined measure of CCR mostly determined? Please check UP TO FOUR of the following:

- ☐ HS Diploma attained
- ☐ MME scores (Proficient vs. Partially Proficient)
- ☐ ACT Plan scores
- ☐ ACT College Entrance Exam scores
- ☐ AP exam scores
- ☐ Common pre- and post-assessments
- ☐ Other (please specify)

Page 12

K-12 Educator/Administrator Evaluation Systems

The following questions pertain to **TEACHER** evaluations.

21. Which factors are mostly used in evaluations for elementary teachers? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Absenteeism from the job
- [ ] Classroom management
- [ ] Content knowledge
- [ ] Instructional practices (including use of technology)
- [ ] Pedagogical knowledge and practice
- [ ] Principal/supervisor evaluations
- [ ] Professional development
- [ ] Professional responsibilities
- [ ] Growth/decline of student achievement data
- [ ] Growth/decline of student growth measures
- [ ] Other, please specify

22. Which factors are mostly used in performance evaluations for middle school teachers? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Absenteeism from the job
- [ ] Classroom management
- [ ] Content knowledge
- [ ] Instructional practices (including use of technology)
- [ ] Pedagogical knowledge and practice
- [ ] Principal/supervisor evaluations
- [ ] Professional development
- [ ] Professional responsibilities
- [ ] Growth/decline of student achievement data
- [ ] Growth/decline of student growth measures
- [ ] Other, please specify

K-12 Educator/Administrator Evaluation Systems

23. Which factors are mostly used in evaluations for secondary teachers? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Absenteeism from the job
- [ ] Classroom management
- [ ] Content knowledge
- [ ] Instructional practices (including use of technology)
- [ ] Pedagogical knowledge and practice
- [ ] Principal/supervisor evaluations
- [ ] Professional development
- [ ] Professional responsibilities
- [ ] Growth/decline of student achievement data
- [ ] Growth/decline of student growth measures
- [ ] Other, please specify

24. What types of decisions are mostly informed by teacher evaluations? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Providing coaching
- [ ] Providing induction support
- [ ] Providing targeted professional development to address specific needs
- [ ] Informing Individualized Development Plan
- [ ] Informing School Improvement Plan
- [ ] Determining additional compensation
- [ ] Determining promotion
- [ ] Recommending removal/termination after being given time to improve
- [ ] Other, please specify

K-12 Educator/Administrator Evaluation Systems

The following questions pertain to **BUILDING PRINCIPAL AND ASSISTANT PRINCIPAL** evaluations.

25. Which factors are mostly used in performance evaluations for elementary principals and assistant principals? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Absenteeism from the job
- [ ] Content knowledge
- [ ] Instructional practices/leadership (including use of technology)
- [ ] Pedagogical knowledge and practice
- [ ] Professional development
- [ ] Professional responsibilities
- [ ] Providing appropriate support for minimally effective and ineffective teachers
- [ ] Conducting evaluations validly and reliably
- [ ] Growth/decline of student achievement data
- [ ] Growth/decline of student growth measures
- [ ] Other, please specify

26. Which factors are mostly used in evaluations for middle school principals and assistant principals? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Absenteeism from the job
- [ ] Content knowledge
- [ ] Instructional practices/leadership (including use of technology)
- [ ] Pedagogical knowledge and practice
- [ ] Professional development
- [ ] Professional responsibilities
- [ ] Providing appropriate support for minimally effective and ineffective teachers
- [ ] Conducting evaluations validly and reliably
- [ ] Growth/decline of student achievement data
- [ ] Growth/decline of student growth measures
- [ ] Other, please specify

K-12 Educator/Administrator Evaluation Systems

27. Which factors are mostly used in evaluations for secondary principals and assistant principals? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Absenteeism from the job
- [ ] Content knowledge
- [ ] Instructional practices/leadership (including use of technology)
- [ ] Pedagogical knowledge and practice
- [ ] Professional development
- [ ] Professional responsibilities
- [ ] Implementation of appropriate support for minimally effective and ineffective teachers
- [ ] Conducting evaluations validly and reliably
- [ ] Growth/decline of student achievement data
- [ ] Growth/decline of student growth measures
- [ ] Other, please specify

28. Which types of decisions are mostly informed by building principal and assistant principal evaluations? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Providing leadership coaching support
- [ ] Providing induction support
- [ ] Imforming school improvement plan
- [ ] Determining appropriate professional development
- [ ] Determining additional compensation
- [ ] Determining promotion
- [ ] Recommending removal/termination after being given time to improve
- [ ] Other, please specify

K-12 Educator/Administrator Evaluation Systems

The following questions pertain to **SUPERINTENDENT** evaluations.

29. Which factors are mostly used in evaluations for the superintendent? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Absenteeism from the job
- [ ] Content knowledge
- [ ] Instructional practices/leadership (including use of technology)
- [ ] Pedagogical knowledge and practice
- [ ] Professional development
- [ ] Professional responsibilities
- [ ] Providing appropriate support for minimally effective and ineffective teachers
- [ ] Conducting evaluations validly and reliably
- [ ] Growth/decline of student achievement data
- [ ] Growth/decline of student growth measures (other than summative data)
- [ ] Other, please specify

30. Which types of decisions are mostly informed by superintendent evaluations? Please check UP TO FOUR of the following:

- [ ] Not Applicable
- [ ] Providing leadership coaching support
- [ ] Providing induction support
- [ ] Informing overall district improvement plan
- [ ] Determining appropriate professional development
- [ ] Determining additional compensation
- [ ] Determining promotion
- [ ] Recommending removal/termination after being given time to improve
- [ ] Other, please specify

APPENDIX B: ORDERED LOGISTIC REGRESSION

1: DATA MANIPULATION

To do this analysis, it was necessary to merge data from CEPI with data from the Registry of Educational Personnel (REP) and the Educational Entity Master (EEM). Data from CEPI consisted of teachers' effectiveness levels and building placements, but did not contain information on their demographics, qualifications, experience, or courses taught. As such, we needed to merge several disparate data sets into a usable whole.

This analysis required several files from the June 2013 REP—Personnel Assignment, Personnel Employment, Personnel Master, and Assignment Codes—along with the district and ISD list from the 2013 EEM. Personnel Assignment contains data on the courses that each teacher was responsible for, Personnel Employment contains data on each teacher's employment history, Personnel Master contains demographic data, Assignment Codes match course data in the Personnel Assignment file with a more convenient and easily usable set of course codes, and the district and ISD list allows us to match each building assignment to a district and ISD. We combined these data sets along sets of unique identifiers—Personnel Employment ID, Personnel Identification Code (PIC), and School Entity ID values. As the CEPI data is uniquely identified at the teacher-by-building level, but the REP data exists at the teacher-by-building-by-course level, we collapsed the REP data to the teacher-by-building level. In doing so, we aggregated the number and type of courses that each teacher taught.

We followed a standard set of procedures in generating each teacher's age and experience. If a teacher was born in January through May of a given year, we computed his age as 2013 minus his date of birth; if he was born in June through December, we subtracted a year (as he had presumably not reached that year's birthday). A similar procedure using hire dates, yielded teachers' years of experience within their districts. Minority teachers, meanwhile, are defined as belonging to any non-White race or ethnicity. While this combines disparate groups of individuals who may face distinct sets of workplace challenges, we are able to avoid small sample issues in the treatment of certain racial or ethnic groups.

2: TECHNICAL INFORMATION

To do this analysis, we utilized the ordered logistic regression function in the Stata software package. This is a type of regression performed when the outcome variable consists of ordered categories but where we cannot precisely quantify any category.

The outcome variable is effectiveness rating, where "highly effective" is the highest rating possible, and "ineffective" is the lowest rating possible. The ordered logistic model predicts the likelihood of a teacher appearing in each successively higher category of effectiveness. **The three "cut" values establish boundaries past which we would expect a teacher to be in a particular effectiveness category, and the coefficients determine where a teacher is located relative to these cut values.** Both intensive margins and extensive margins—whether a teacher is on either side of a particular cut and how far they are from that cut—matter in assessing the likelihood of appearing in a particular category. To illustrate, consider three teachers—one with a predicted value of 0, one with a predicted value of 0.75, and one with a predicted value of 3. We would expect that the teacher with a value of 0 would be listed as "effective," as he is below the highest cut point but above the next lowest one. As his value is much closer to the highest cut point than it is to the next lowest, we would also that he is somewhat likely to be listed as "highly effective" and fairly unlikely to be listed as "minimally effective." The teacher with a value of 0.75 is most likely to be "highly effective," but his proximity to the cut point makes it nearly as likely that he could be listed as "effective." The teacher with a value of 3 is so far above the highest cut point that it is unlikely that he will be listed in any category other than "highly effective."

For more information on understanding the output of an ordinal logit model from the Stata software program, please see **http://www.ats.ucla.edu/stat/stata/output/stata_ologit_output.htm**.

Variable Name	Category	Interpretation	Omitted Category
female	Demographics	Female	Male
age		Age (in years)	N/A
minority		Minority	White
newteach	Experience	New (first 3 years teaching)	Experienced Teachers
yrsindist		Years in district	N/A
newexp		newteach * yrsindist (higher returns in first 3 years?)	Experienced Teachers
profcert	Qualifications	Has professional certification	No professional certification
MAplus		Has MA or higher degree	BA or lower degree
maj_yes		Majored in area of certification	Did not major in area of certification
maj_NA		Major not applicable to certification	Did not major in area of certification
fte1plus	Assignments	Has FTE >= 1	Teacher has FTE<1
ela		Teaches ELA	Elementary Only
math		Teaches Math	
socsci		Teaches Social Sciences	
sci		Teaches Science	
sped		Teaches Special Education	
arts		Teaches Arts	
world		Teaches World Languages	
tech		Teaches Technology	
well		Teaches Wellness	
mult		Teaches Multiple Subjects	

3: RESULTS

Ordered logistic regression Number of obs = 93133
LR chi2 (21) = 3538.02
Prob > chi2 = 0.0000
Log likelihood = −70105.289 Pseudo R2 = 0.0246

effcode	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
female	.2781125	.0172341	16.14	0.000	.2443342	.3118907
age	-.0179868	.0009289	-19.36	0.000	-.0198073	-.0161662
minority	.1733546	.0253203	6.85	0.000	.1237277	.2229815
newteach	-.524115	.0340251	-15.40	0.000	-.590803	-.4574271
yrsindist	.0340047	.0011608	29.30	0.000	.0317297	.0362798
newexp	.0860519	.0140194	6.14	0.000	.0585743	.1135294
profcert	.1789792	.0157754	11.35	0.000	.1480601	.2098984
MAplus	.181832	.0157273	11.56	0.000	.1510071	.2126568
maj_yes	-.3397171	.020702	-16.41	0.000	-.3802937	-.2991404
maj_NA	-.1408221	.0303849	-4.63	0.000	-.2003754	-.0812689
fulltime	.2110261	.0205389	10.27	0.000	.1707706	.2512816
ela	.0477341	.0237151	2.01	0.044	.0012533	.094215
math	-.0907009	.0261623	-3.47	0.001	-.1419781	-.0394237
socsci	-.1734248	.0290636	-5.97	0.000	-.2303885	-.1164611
sci	-.0598877	.0282837	-2.12	0.034	-.1153227	-.0044526
sped	-.1485253	.0263058	-5.65	0.000	-.2000837	-.0969669
arts	.1132191	.0286652	3.95	0.000	.0570363	.1694019
world	-.1574759	.0375013	-4.20	0.000	-.2309772	-.0839746
tech	-.108376	.0494917	-2.19	0.029	-.2053779	-.0113741
well	-.1460406	.0337692	-4.32	0.000	-.2122271	-.0798542
mult	-.2447648	.0306359	-7.99	0.000	-.3048099	-.1847196
/cut1	-5.302876	.0635099			-5.427354	-5.178399
/cut2	-3.676642	.0515523			-3.777683	-3.575601
/cut3	.6209152	.0481528			.5265375	.7152929

APPENDIX C: MCL 380.1249 LEGISLATIVE TIMELINE

The chart below illustrates the progression from the local systems to the statewide evaluation system.

Current Legislative Timeline			
School Year	**Tool Type**	**% of evaluation based on student growth & achievement data**	**Reporting Requirement**
2011-2012	locally determined Educator Evaluation Systems	significant part	effectiveness labels in June REP collection
2012-2013	locally determined Educator Evaluation Systems & MCEE Pilot		
2013-2014	MCEE's Evaluation Tool	25%	
2014-2015		40%	
2015-2016		50%	

For more information, please see **www.michigan.gov/educatorevaluations**,

NOTES

NOTES

MICHIGAN Department of Education