Understanding Educator Evaluations in Michigan
Results from Year 1 of Implementation

Venessa A. Keesler, Ph.D.
Carla Howe
Michigan Department of Education

What do we know about educator evaluations in Michigan?

In 2011-2012, Michigan school districts implemented significant new legislation regarding educator evaluations. Every district in Michigan evaluated all of their educators (including teachers and administrators) and reported the results of those evaluations in the state's Registry of Educational Personnel maintained by the Center for Educational Performance and Information (CEPI). This new law requires evaluations to be based "in significant part" on student growth and to be used to inform decisions regarding placement, promotion, compensation, and retention.[1]

In the first and second years of implementation of this law (2011-2012 and 2012-2013), districts were allowed to develop, implement or maintain their own unique systems that met basic requirements.[2] This means that districts developed and implemented their own systems, following the basic requirements of the legislation.[3] It is important to underscore that in the initial year of implementation, Michigan had over 800 unique district evaluation systems. This makes direct comparison of district effectiveness ratings and systems extremely difficult, as ratings were not determined with standard rigor across districts. For example, what it takes to be determined "effective" in District A may be very different than in District B.

It is also important to note that Michigan districts undertook and completed an impressive amount of work to develop and implement these evaluation systems. While there still is work to be done, and while most districts would agree that their systems need additional refinement, the fact that *every* Michigan educator was evaluated this year represents a significant and important achievement for Michigan's educational system.

---

[1] Please note that the legislation also allows districts on a prior contract to delay implementation of this legislation fully until the contract expires. Districts were still required to report effectiveness ratings, but the content of those evaluations could still be based on the prior system.

[2] Beginning in 2013-2014, a statewide system of educator evaluations will be implemented based on the educator evaluation pilot results and recommendations of the Michigan Council for Educator Effectiveness (MCEE).

[3] For a more detailed discussion of the legislation, see Appendix A.

Understanding Educator Evaluations in Michigan   *Keesler and Howe*
*Michigan Department of Education*   *Page 1*
November 2012

What was included in local evaluation systems?

From April to August of 2012, districts were required as part of the State Fiscal Stabilization Fund to respond to a Michigan Department of Education (MDE)-developed survey regarding their K-12 Educator Evaluation Systems. There were 792 of 897 districts that participated in the K12 Educator Evaluation Survey. This provides important information on the content and structure of educator evaluation systems statewide. A summary of the results is below.

Use of Frameworks/Systems/Methods As Part of Local Evaluation Systems

Districts were asked to report on the systems, frameworks or methods used as part of their local evaluation system. As Figure 1 below shows, 50% of districts across the state of Michigan report using Charlotte Danielson's *Enhancing Professional Practice for Performance of Teaching* as the framework for their evaluation systems; many more districts reported in the "Other" category as using Danielson and local measures combined.

Figure 1: Frameworks/Systems in Local Evaluation Systems
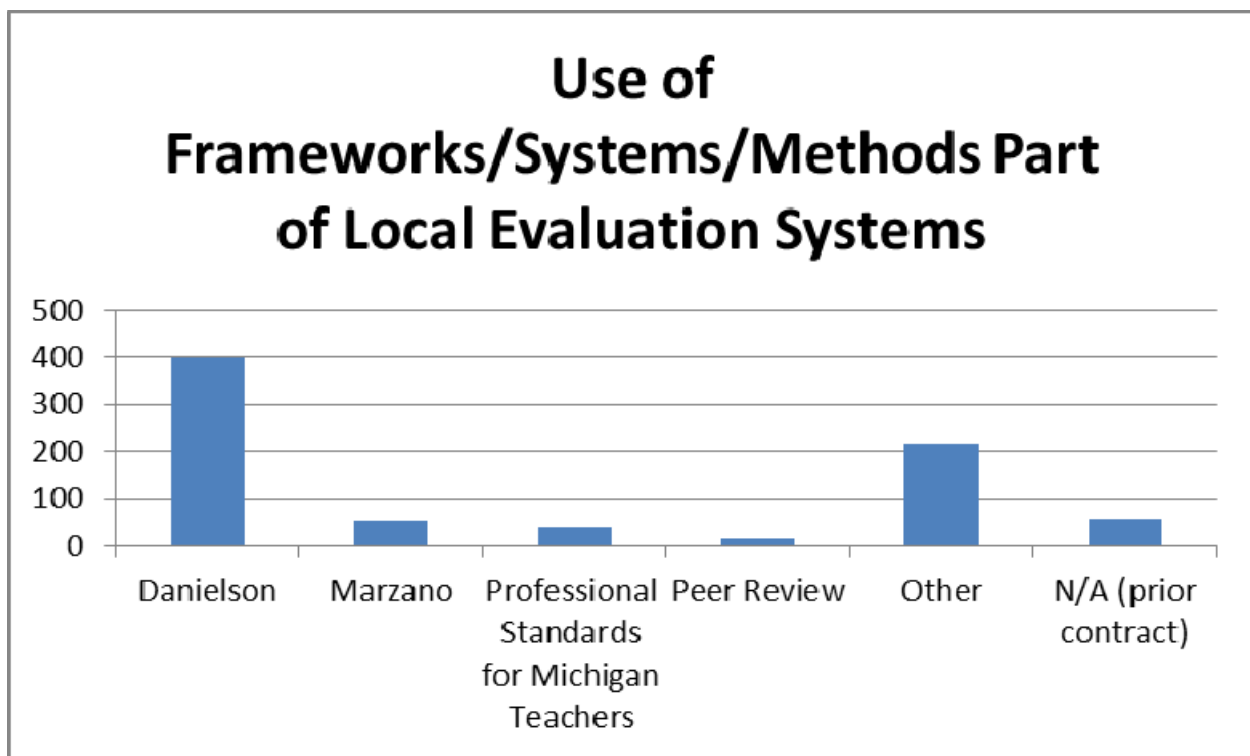

The "Other" Frameworks reported include: Charlotte Danielson Framework AND a local component, Teacher Advancement Program, My Learning Plan, 5 Dimensions of Teaching and Learning, Local District or ISD framework, McREL, STAGES, Kim Marshall Rubrics.

Percentage of Student Growth in Local Evaluation Systems

As indicated above, the educator evaluation legislation required that districts include student growth as a "significant part" of the final evaluation. However, state law does not define "significant," which means local districts had to determine this for themselves.

Figure 2 displays the percentage of student growth in the local evaluation systems. Nearly 400 districts had between 11 and 30% of their evaluation based on student growth, and an additional 200 had more than 30% based on student growth. This is reasonable for the initial year of implementation, particularly as districts and the state as a whole grapple with the issue of appropriate metrics and measures of growth for all teachers and students. These districts also are on track to make the transition to the statewide evaluation system (when implemented), as the law requires that student growth will be 25% of each evaluation beginning in 2013-2014.

**Figure 2: Percentage of Student Growth in Local Evaluation Systems**
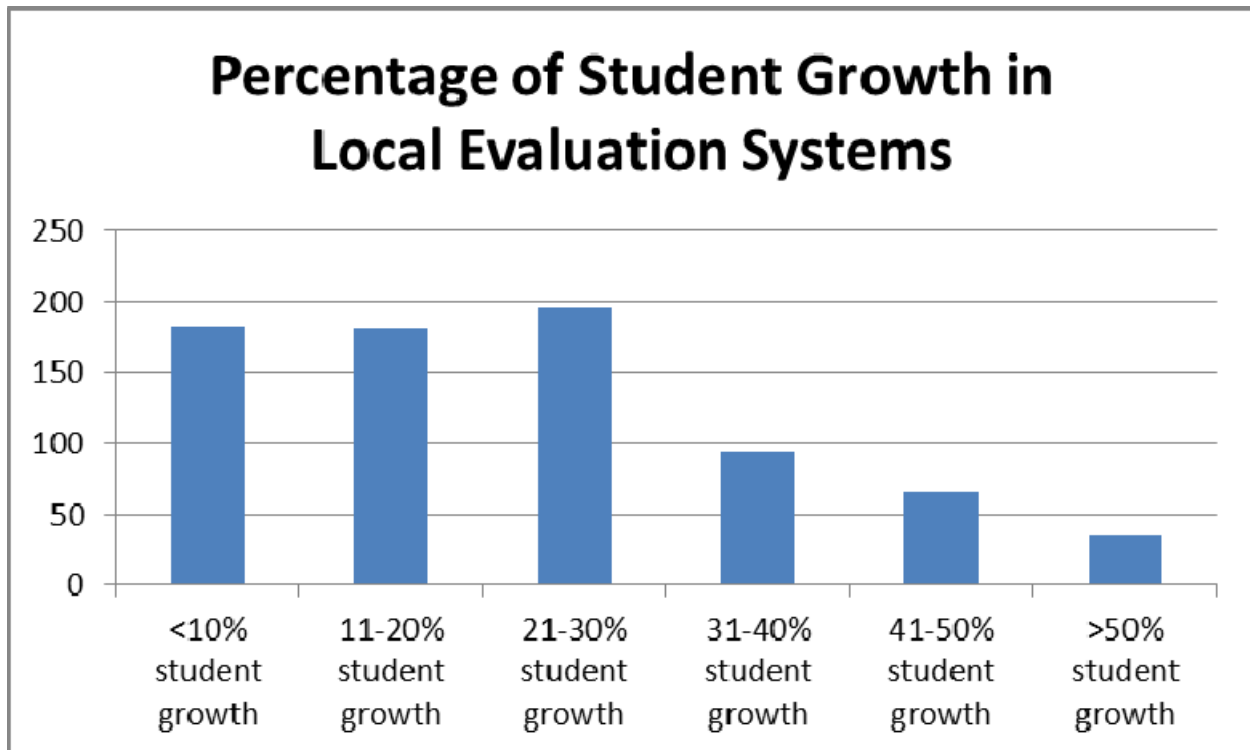
Type of Assessments Used

Districts were asked to indicate which types of assessments they used to determine student growth. In Figures 3-6, the most common types of assessments indicated are displayed by various levels (early elementary, elementary, middle school, and high school).
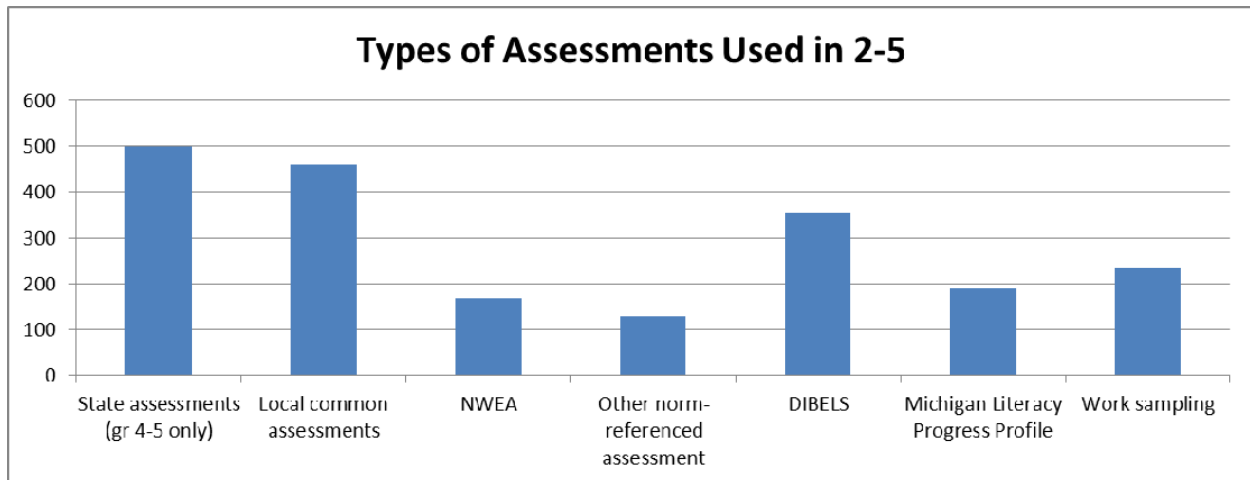
Several things to note:

- Local common assessments are a frequent option utilized by districts at all levels.

- There are more types of assessments available for measuring growth in the younger grades than in the high school grades.

- In the early elementary and elementary grades, the most popular choices for assessments to determine growth are local common assessments, followed by DIBELS and work sampling (see Figures 3 and 4).

- The state assessments are used heavily in districts for the grades in which they are available (grades 4-8) (see Figures 4 and 5).

- ACT Explore and ACT Plan are the most common choices (after common assessments and/or end-of-course assessments) in the middle and high school grades.

Figure 3: Types of Assessments Used to Determine Student Growth in Grades K-1 (early elementary)
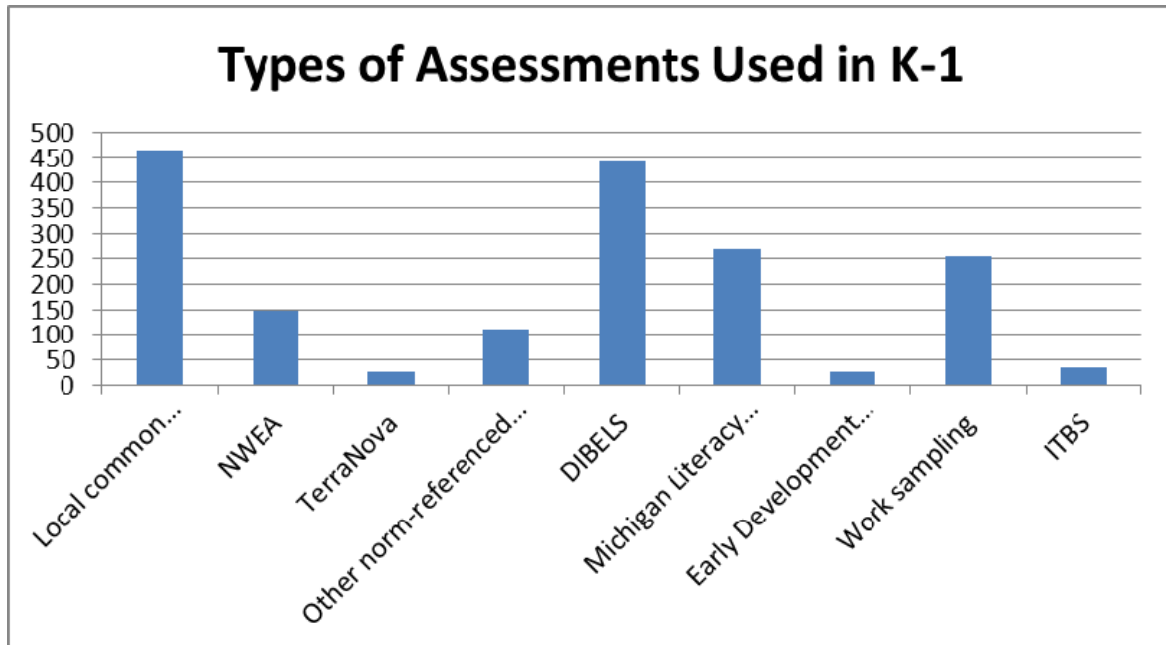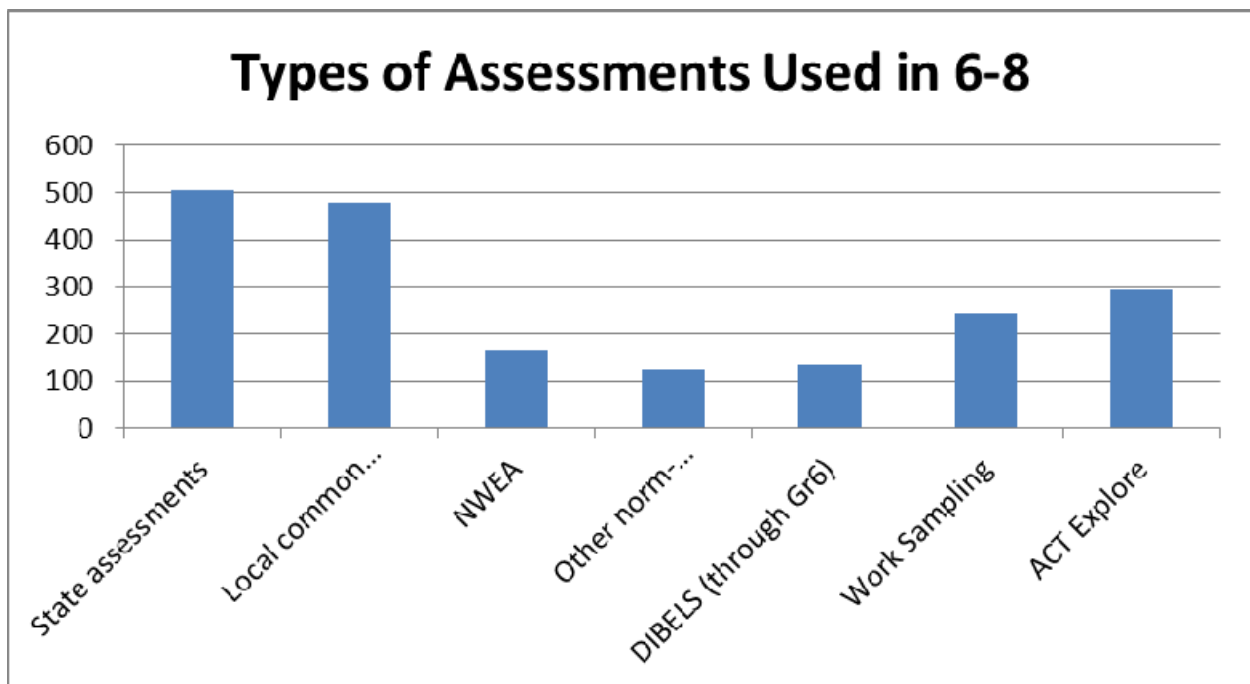

*Note: NWEA = Northwest Evaluation Association; DIBELS = Dynamic Indicators of Basic Early Literacy Skills

Figure 4: Types of Assessments Used to Determine Student Growth in Grades 2-5 (Elementary)

Types of Assessments Used in 2-5

A bar chart showing the following approximate values:
- State assessments (gr 4-5 only): 500
- Local common assessments: 460
- NWEA: 170
- Other norm-referenced assessment: 130
- DIBELS: 355
- Michigan Literacy Progress Profile: 190
- Work sampling: 235

*Note: NWEA = Northwest Evaluation Association; DIBELS = Dynamic Indicators of Basic Early Literacy Skills

Figure 5: Types of Assessments Used to Determine Student Growth in Grades 6-8

Types of Assessments Used in 6-8

A bar chart showing the following approximate values:
- State assessments: 505
- Local common...: 480
- NWEA: 165
- Other norm-...: 125
- DIBELS (through Gr6): 135
- Work Sampling: 245
- ACT Explore: 295

*Note: NWEA = Northwest Evaluation Association; DIBELS = Dynamic Indicators of Basic Early Literacy Skills

Figure 6: Types of Assessments Used to Determine Student Growth in High School Grades

Types of Assessments Used in HS
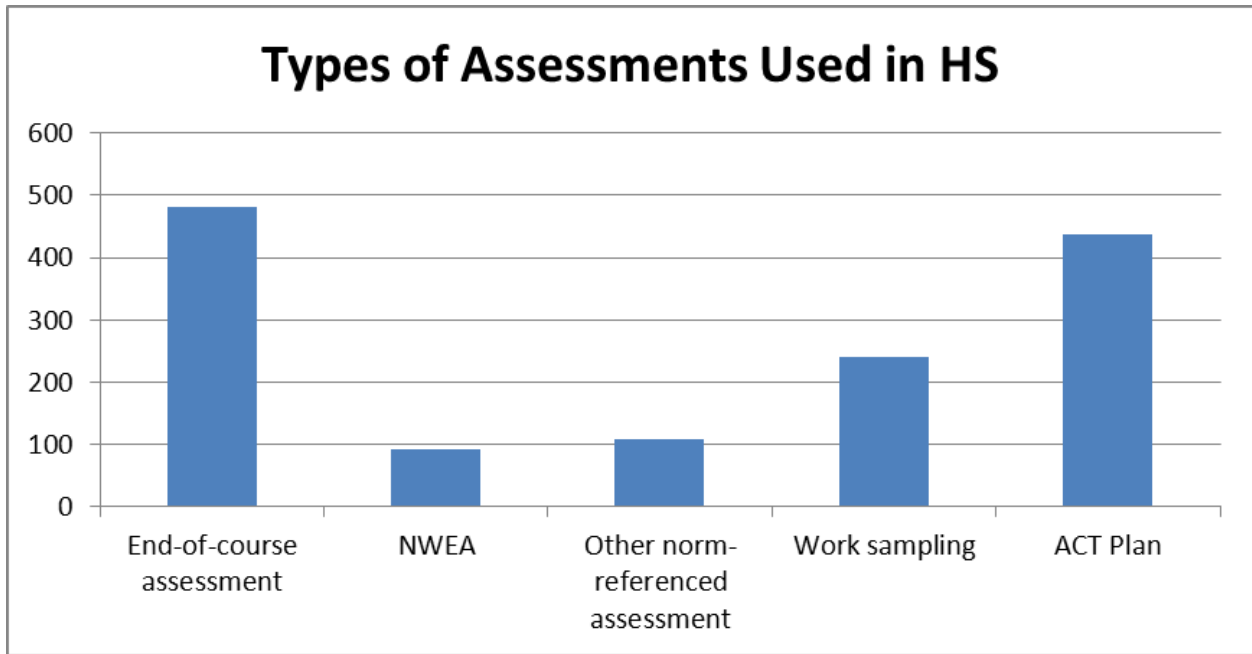
Decisions Informed by Evaluation Results

Districts were asked to indicate how they are using the results of the evaluations (see Figure 7). Over 700 districts indicated they are using the results to make determinations about the professional development needs and requirements for educators, which was one of the key desired outcomes of the new state law. Additionally, over 600 districts use it to drive coaching efforts for teachers, and over 400 used it to determine and provide induction support for new teachers. This is an important positive step and represents an increase in the amount of professional development, coaching (e.g. instructional, leadership), and support that Michigan educators are receiving.

**Figure 7: Decisions Informed by Evaluation Results Across Michigan Districts**
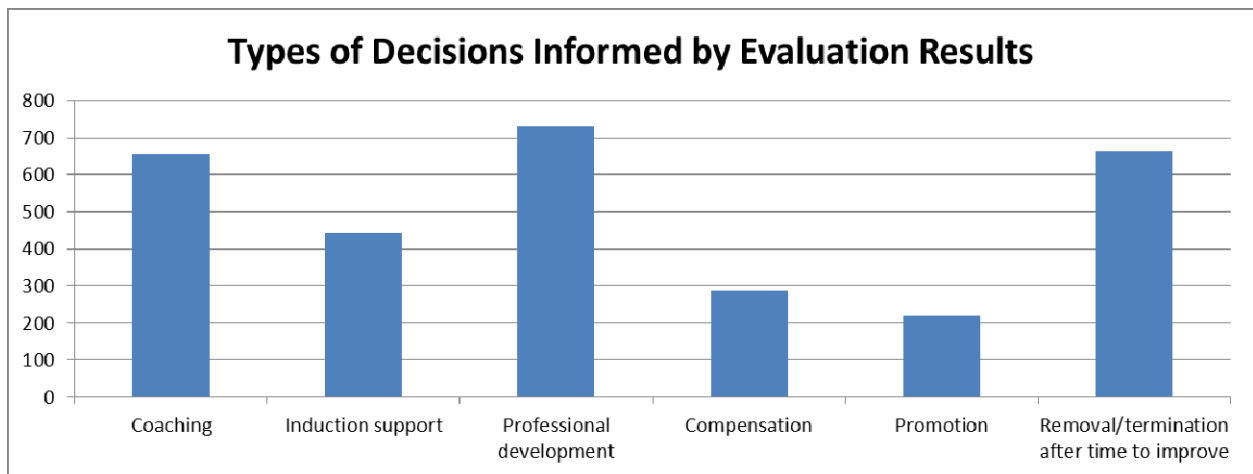
Others types of decisions include: Assignment to committees or roles beyond the classroom, classroom support and assistance, layoff/recall/transfer, mentoring, staff placement, scheduling, setting improvement goals, merit pay.

STATEWIDE DISTRIBUTION OF EDUCATOR EFFECTIVENESS RATINGS

Although ratings are not directly comparable across districts, it is important to understand information about ratings statewide, keeping in mind that each rating was based on local district determinations. In Figure 8 below, the statewide distribution of teachers in each of the four effectiveness ratings is presented.

- 23% of Michigan teachers were reported as "highly effective" as determined by their local evaluation systems.
- 75% of Michigan teachers were reported as "effective" as determined by their local evaluation systems.
- 2% of Michigan teachers were reported as "minimally effective" as determined by their local evaluation systems.
- Slightly less than 1% (0.8%) were reported as "ineffective" as determined by their local evaluation systems.

Figure 8: Percent of Michigan Teachers Reported in Each of the Four Effectiveness Ratings[4]
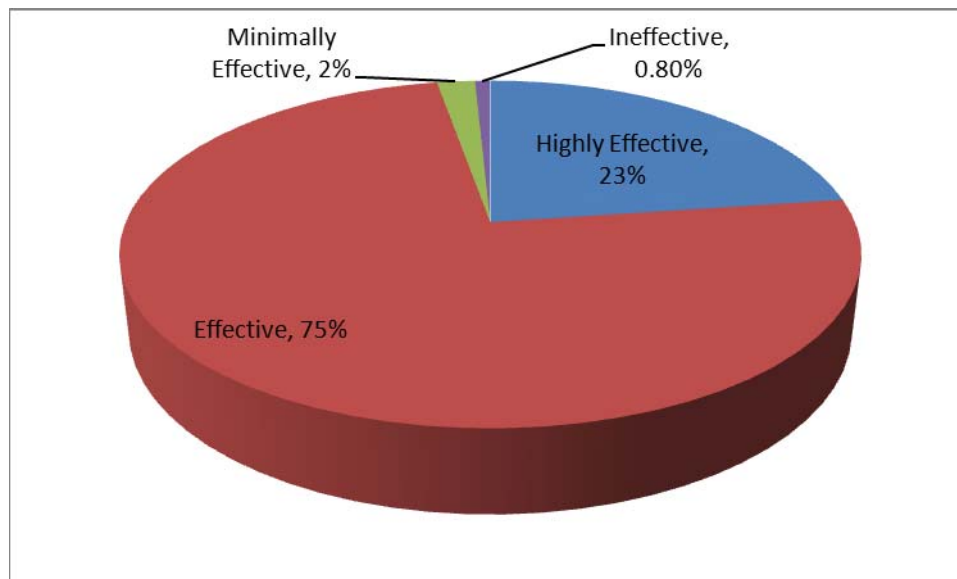
Key Takeaways:

- Evaluation ratings have expanded in most districts from two ratings previously (satisfactory and unsatisfactory) to four rating categories. The most differentiation occurs between effective and highly effective. Previously, the system would not have been able to make this distinction. This

---

[4] The distribution of labels for administrators was very similar (23% highly effective, 74% effective, 1.76% minimally effective, and 0.73% ineffective).

adds an important level of feedback and differentiation to help teachers strategically improve as professionals.

- The rating "ineffective" was given to slightly less than 1% of teachers by their local evaluation systems. This may cause some to wonder why there were not more teachers reported as "ineffective." These rating are appropriate, particularly given that this is the first year of implementation. **Educators at all levels of the system have room to improve, as do professionals in all lines of work.** It is critical that educators continue to ensure that this policy does what it is intended to do—which is provide timely and appropriate feedback to all educators and that the result of that feedback is specific improvement in areas where growth is necessary.[5]

---

[5] Two other issues with the first year of implementation that should be recognized are: 1) principals received little to no concerted training in conducting the evaluations and 2) various law firms have advised that much of the legal defensibility for assigning a given label falls solely to the principal, in the absence of a statewide system. Both of these factors made it more challenging for principals to know when an "ineffective" label was appropriate, and that they likely exercised caution in using this label.

Understanding Educator Evaluations in Michigan   *Keesler and Howe*
Michigan Department of Education   *Page 9*
November 2012

Distribution of Effectiveness Ratings By Level of Student Growth in the System

Given the challenges with assigning the more "extreme" rating labels (i.e. highly effective and ineffective), it is likely that districts with greater percentages of student growth in the system will be more likely to use all four ratings more frequently, as the student growth component provides what many perceive to be a more objective measure in the evaluation system.[6]

Figure 9: Distribution of Effectiveness Ratings by Percent of Evaluation Based on Student Growth
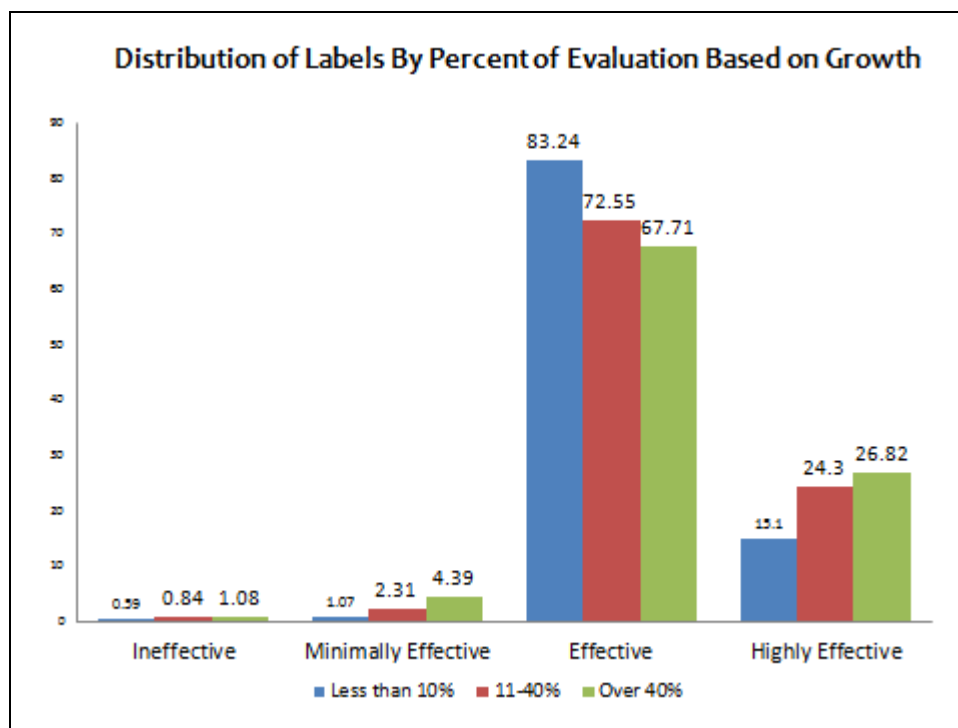
In Figure 9, this appears to be the case. As the percentage of the evaluation based on student growth increases, the number of teachers reported as "effective" (the most common category) decreases, and the number of teachers reported as "highly effective," "minimally effective," and "ineffective" increases. In short, the presense of more student growth data seems to be related to a more frequent utilization of all four evaluation rating categories. This suggests that as the statewide educator evaluation system is developed and implemented, with its prescribed percentage of student growth, that districts will be able to better differentiate and more fully utilize all categories of ratings more frequently.

---

[6] Note: MDE does not agree that student growth is "objective" while other factors (such as observations) are "subjective." We do, however, recognize that student growth data is often perceived to be more concrete and defensible when explaining the results of an evaluation decision. We also feel that as more principals are trained on evaluation systems and how to use observation rubrics/tools, the perception of these as "subjective" measures will greatly decrease.

Understanding Demographic Characteristics and Evaluation Ratings

In order to understand some of the relationships between various characteristics of teachers and their final evaluation rating, MDE performed an analysis that allows us to predict the likelihood of a teacher appearing in the "highly effective" category relative to the other three categories, while holding constant other factors. It is important to note that this analysis only establishes *relationships*; it does not provide information on what *causes* a teacher to be effective or highly effective. Additionally, this analysis is limited by the data available in the state systems; we only collect a limited number of variables on teachers in Michigan. Finally, this analysis does not take into account school and district characteristics that are likely important in terms of the variation in teacher effectiveness ratings. In summary, this analysis is useful in generating a basic picture of highly effective teachers in Michigan, but does not unpack those relationships.[7]

Who is more and less likely to be rated as highly effective in Michigan (based on local evaluation systems and 2011-2012 data)?

Based on the analysis described above, we see that female teachers are more likely to be reported as highly effective, as are teachers with more time in the same district. Teachers with a professional certificate (as opposed to a provisional certificate or other types of certificates) are more likely to be reported as highly effective. Teachers with a master's degree or higher are also more likely to be rated as highly effective, as are those in districts with student growth consisting of 40% or more of their evaluation system.

Conversely, teachers who are less likely to be rated as highly effective are older teachers[8] (although teachers with more time in the same district are MORE likely to be rated as highly effective) and new teachers (those in their first three years of teaching). Teachers in systems where growth is less than 10% of the evaluation system are also less likely to be rated as highly effective. Finally, teachers of mathematics, science, social science, special education and world language are all less likely to be rated as highly effective (as compared with elementary teachers).

---

[7] Technical note: this analysis is based on an ordered logistic regression, which allows for the calculation of the proportional odds of being in the highly effective category relative the other three categories, as well as the proportional odds of being in the effective category relative the other two categories. See Appendix B for detailed output.

[8] Age is a continuous variable; in the model, increases in age are related to a decreased likelihood of being reported as highly effective.

Priority, Focus and Reward Schools: Teacher Effectiveness Ratings

Teacher effectiveness ratings and school-level accountability are not the same thing. For example, it is easy to think of a school that has overall low levels of student proficiency where many teachers within that school are showing growth with students and are effective at their roles. Teacher effectiveness ratings are based on the ability of an educator to move a student forward, to help them show growth. School-level accountability is based on a mixture of achievement (which is strongly related to what a child comes in with), growth, and achievement gaps. Therefore, it is unlikely strong relationships between these two types of metrics.[9]

Given these concerns, MDE did analyze the effectiveness ratings in our three new categories of schools: Priority, Focus and Reward. The results are presented in Table 1 below.

Table 1: Percentage of Teachers in Each Effectiveness Rating by Priority, Focus and Reward Schools

	Ineffective	Minimally Effective	Effective	Highly Effective
Priority	2.5%	6.3%	74%	18%
Focus	0.5%	1.5%	80%	18%
Reward	0.5%	0.85%	74%	24%
Statewide	**0.8%**	**2%**	**75%**	**23%**

This table shows that there are larger numbers of teachers reported as "ineffective" and "minimally effective" in Priority Schools than in the other two types of schools, as well as more in Priority Schools than in all schools statewide. We would expect to see this relationship. This could also be a result of the requirements placed on districts with Priority Schools, which includes implementing a transformation model for school improvement.   Keep in mind, however, that Priority Schools are required to undertake certain staffing changes after being named to the Priority List.

Table 1 also shows that Reward Schools have a greater percentage of teachers in the highly effective category than Focus and Priority Schools, although it's similar to the statewide percentage. Again, we

---

[9] A recent report by an advocacy organization made the assumption that because a district had low levels of proficiency, it was not possible that they have high numbers of effective teachers. This is an erroneous conclusion. Teachers can be effective in showing student growth within a system with low proficiency levels. Additionally, successful districts and schools are systems issues; teachers may be effective in their own classrooms but the system may be flawed at the school and district level. While certainly all of these factors must work together to support student achievement, unfounded leaps of logic to place all of the "blame" on teachers rather than considering how the *system* as a whole succeeds and fails with students should not be made.

would expect some relationship here, given that two of the ways to be named a Reward school are related to a school's ability to show growth with their students.

For the 2011-2012 and 2012-2013 school years, educator evaluations are governed by the following legislation, which is MCL 380.1249.

 (1) Not later than September 1, 2011, and subject to subsection (9), with the involvement of teachers and school administrators, the board of a school district or intermediate school district or board of directors of a public school academy shall adopt and implement for all teachers and school administrators a rigorous, transparent, and fair performance evaluation system that does all of the following:

(a) Evaluates the teacher's or school administrator's job performance at least annually while providing timely and constructive feedback.
(b) Establishes clear approaches to measuring student growth and provides teachers and school administrators with relevant data on student growth.
(c) Evaluates a teacher's or school administrator's job performance, using multiple rating categories that take into account data on student growth as a significant factor. For these purposes, student growth shall be measured by national, state, or local assessments and other objective criteria. If the performance evaluation system implemented by a school district, intermediate school district, or public school academy under this section does not already include the rating of teachers as highly effective, effective, minimally effective, and ineffective, then the school district, intermediate school district, or public school academy shall revise the performance evaluation system within 60 days after the effective date of the amendatory act that added this sentence to ensure that it rates teachers as highly effective, effective, minimally effective, or ineffective.
(d) Uses the evaluations, at a minimum, to inform decisions regarding all of the following:

(i) The effectiveness of teachers and school administrators, ensuring that they are given ample opportunities for improvement.
(ii) Promotion, retention, and development of teachers and school administrators, including providing relevant coaching, instruction support, or professional development.
(iii) Whether to grant tenure or full certification, or both, to teachers and school administrators using rigorous standards and streamlined, transparent, and fair procedures.
(iv) Removing ineffective tenured and untenured teachers and school administrators after they have had ample opportunities to improve, and ensuring that these decisions are made using rigorous standards and streamlined, transparent, and fair procedures.

The manner in which each district chooses to implement and combine these criteria does differ across districts for the 2011-2012 and the 2012-2013 school year.

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.mi.gov/educatorevaluations, and click on "Educator Evaluation Updates."

To do this analysis, we utilized the ordered logistic regression function in the Stata software package. This is a type of logistic regression performed when the outcome variable is ordinal (i.e. has categories that are ordered).

The outcome variable is effectiveness rating, where highly effective is the highest rating possible, and ineffective is the lowest rating possible. Therefore, the ordered logistic model predicts the likelihood of a teacher appearing in each successively higher category of effectiveness. We choose to express these values as odds ratios, which are actually proportional odds ratios, which allow for the comparison of teachers in groups greater than a given level with teachers in groups less than a given level. In other words, we can compare all teachers in the highly effective categories with all teachers in a category less than highly effective; we can also do this with teachers that are in the effective and highly effective category relative to teachers in all other categories.

The predictor variables used are as follows:
- Female: dummy variable compared to male.
- Age: the age of the teacher, expressed as a continuous variable.
- New teacher ("newteach"): indicator variable for whether a teacher is labeled as a new teacher which means a teacher within his/her first three years of teaching.
- Time in district ("timeindistrict"): continuous variable expressing the number of years a teacher has been in the same district the one in which they are currently reported.
- Professional: indicates a professional license; dummy variable comparing professional license to all other types of license (i.e. provisional and others).
- Minority: dummy variable comparing any minority teacher to white teachers.
- Masters Plus: dummy variable, compares a master's degree or greater to less than a master's degree.
- ELA, math, science, social science, Special Education, arts, world language, technology, wellness, support: these are all subject assignment codes grouped into large categories; here, compared to the reference category of elementary teacher.
- Less than 10 and growth over 40: the percentage of the evaluation based on growth in a given district; less than 10 indicates less than 10% based on growth, and growth over 40 indicates greater than 40% of the evaluation based on growth. These are both compared to the reference category of 11-39% growth.
- Yes Major and Major NA: comparing teachers with a major in their subject area (or teachers where a major is not applicable) to those without a major in their subject area.

The model presented below is displayed using proportional odds ratios, which have the interpretation that for a one unit change in the predictor variable, the odds for cases in a group that is greater than k versus less than or equal to k are the proportional odds larger.

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.

Raw output is displayed below. Again, we note that the purpose of this model was to understand relationships between teacher characteristics and effectiveness labels; the model does not take into account school or district effects and should not be interpreted using a causal inference framework.

```
Iteration 0:   log likelihood = -56364.646
Iteration 1:   log likelihood = -55129.748
Iteration 2:   log likelihood = -55107.885
Iteration 3:   log likelihood =  -55107.86
Iteration 4:   log likelihood =  -55107.86

Ordered logistic regression                    Number of obs   =      84479
                                               LR chi2(21)     =    2513.57
                                               Prob > chi2     =     0.0000
Log likelihood =  -55107.86                    Pseudo R2       =     0.0223

------------------------------------------------------------------------------
 effcode4cat | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.273788    .025598    12.04   0.000     1.224592     1.32496
         age |   .9826755   .0010638   -16.14   0.000     .9805927    .9847628
    newteach |   .6066601    .021983   -13.79   0.000     .5650686    .6513128
  timeindist~t |  1.02602   .0014016    18.80   0.000     1.023276    1.028771
professional |   1.074608   .0195755     3.95   0.000     1.036918    1.113668
    minority |    .490518   .0157384   -22.20   0.000     .4606213    .5223552
  mastersplus |  1.099911   .0202741     5.17   0.000     1.060884    1.140374
         ELA |   .9893536   .0280276    -0.38   0.706     .9359178     1.04584
        math |   .8810976   .0285531    -3.91   0.000     .8268748     .938876
     science |    .911875   .0318643    -2.64   0.008     .8515129    .9765161
  socialscie~e |  .8473319   .0311448    -4.51   0.000     .7884362    .9106271
       speced |   .8476577   .0319671    -4.38   0.000     .7872629    .9126856
        arts |   .9534063   .0290824    -1.56   0.118     .8980763    1.012145
  worldlangu~e |  .7887526   .0359121    -5.21   0.000     .7214154    .8623751
        tech |   .8797483   .0465611    -2.42   0.015     .793064    .9759075
     wellness |   .8305708   .0345819    -4.46   0.000     .7654834    .9011925
     support |   1.174738    .054836     3.45   0.001     1.072031    1.287284
    lessthan10 |  .6347098   .0117102   -24.64   0.000     .6121682    .6580813
  growthover40 |  1.101402   .0278448     3.82   0.000     1.048157    1.157351
     YesMajor |   1.028393   .0265894     1.08   0.279     .9775773    1.081851
      MajorNA |   1.009021   .0393887     0.23   0.818     .9346996    1.089251
-------------+----------------------------------------------------------------
       /cut1 |  -5.352867   .0666052                     -5.483411   -5.222324
       /cut2 |  -4.004358   .0577495                     -4.117545   -3.891171
       /cut3 |    .8993624   .0537454                      .7940234    1.004701
------------------------------------------------------------------------------
```

For questions or comments regarding the content of this report, please contact:
Venessa Keesler and/or Carla Howe
Bureau of Assessment and Accountability
Michigan Department of Education
517-373-1342 or mde-accountability@michigan.gov.