

## MEMORANDUM

TO: Rob Love, Policy and Research Specialist, Indiana Department of Education

FROM: Aleka Calsoyas, Partner; Chris Henderson, Project Director; Christine Rhyner, Project Director; Megan Kinninger, Analyst

DATE: May 7, 2013

RE: Indiana Evaluation Pilot Summative Findings

---

### Introduction

During the 2011-2012 school year, the Indiana Department of Education (IDOE) piloted RISE, a department and stakeholder-developed teacher evaluation system, in three districts. In addition, the department studied the implementation of alternative, locally-developed or purchased teacher evaluation systems in three additional districts.<sup>i</sup> The results of this effort were released in two reports to support the implementation of teacher evaluation systems in districts across the state.

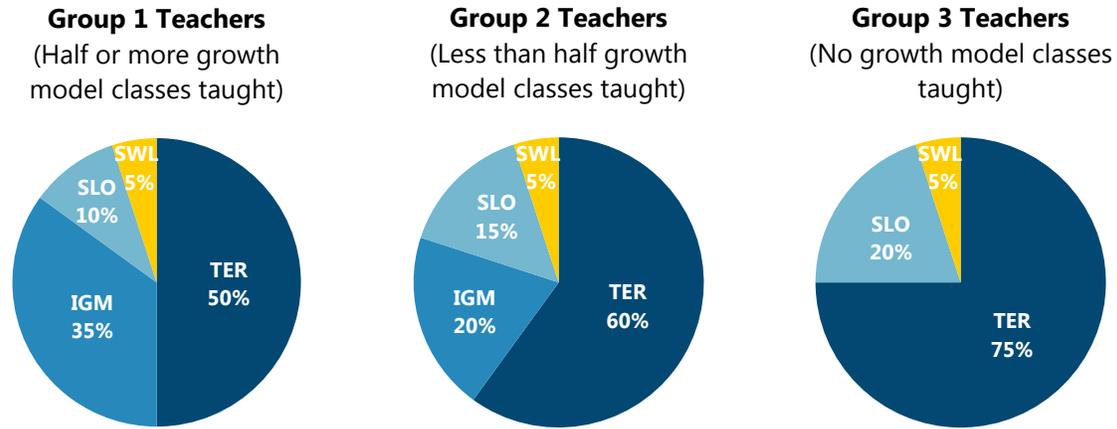
At the time of the year-end report release, ratings for measures of student learning dependent upon state test scores were unavailable. This brief memo focuses on the results of RISE pilot summative evaluation ratings and the relationship between ratings of the RISE systems' multiple measures.<sup>ii</sup>

### Key Findings

1. RISE measures work as intended relative to one another - each measure tells a similar, but not identical, story about teacher performance.
2. The Individual Growth Model had the most meaningful distribution, with only 69 percent of rated Effective or Highly Effective. Ratings for this measure were consistent across all pilot corporations.
3. Classroom observation ratings had the least meaningful distribution, with over 90 percent of pilot teachers in most corporations rated Effective or Highly Effective on this measure. In addition, nearly half of all teachers received Highly Effective ratings on the Student Learning Objective measure.
4. Of all RISE measures, Individual Growth Model measure and the Student Learning Objective measure were the most strongly correlated.
5. 87 percent of RISE teachers received a summative evaluation rating of Effective or Highly Effective. While this level of differentiation represents a step in the right direction compared to previous evaluation systems, progress is needed to improve accuracy of ratings and to provide teachers with better feedback about their performance.

## RISE Analysis Context

RISE was designed to provide teachers feedback on their classroom performance and an accurate, fair evaluation by combining multiple measures of professional practice and student learning. Up to four measures contribute to a teacher’s summative evaluation rating across four performance levels.<sup>iii</sup> The composition and weight of measures comprising a teacher’s summative rating vary to reflect a teacher’s mix of classes taught. More weight is given to the measures most likely to provide an accurate rating. Due to its statistical reliability, when growth model data is available it is included.

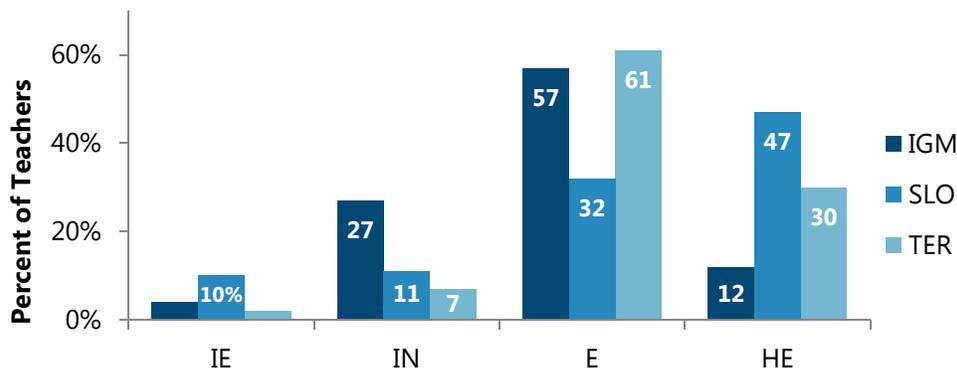


Teachers were grouped by administrators in the pilot. Summative evaluation ratings were calculated based on a teacher’s assigned group. For purposes of this report, summative evaluation ratings were only calculated when data for all measures was available. In total, summative ratings were calculated for only 44 percent of RISE pilot teachers, largely due to missing or invalid SLO data.<sup>iv</sup>

## Relationship of RISE’s Measures

Of the four RISE measures, IGM model ratings were the most evenly distributed across the four performance levels. SLO and TER ratings were skewed, with the overwhelming majority of teachers rated Effective or Highly Effective.

Figure 1: Comparison of distribution by performance level for RISE measures



Overall, RISE’s multiple measures provide similar ratings for a teacher’s effectiveness. All individual teacher measures - TER, IGM, and SLO - are positively correlated. The strength of relationship between measures is similar to other evaluation systems.<sup>v</sup> RISE measures work as intended relative to one another - each measure tells a similar, but not identical, story about teacher performance meaning teachers receive accurate, differentiated feedback on their performance.

Figure 2: Correlation of RISE Measures

	TER	SLO	IGM
Teacher Effectiveness Rubric (TER)	1.00	.138**	.196**
Student Learning Objectives (SLO)		1.00	.344**
Individual Growth Model (IGM)			1.00

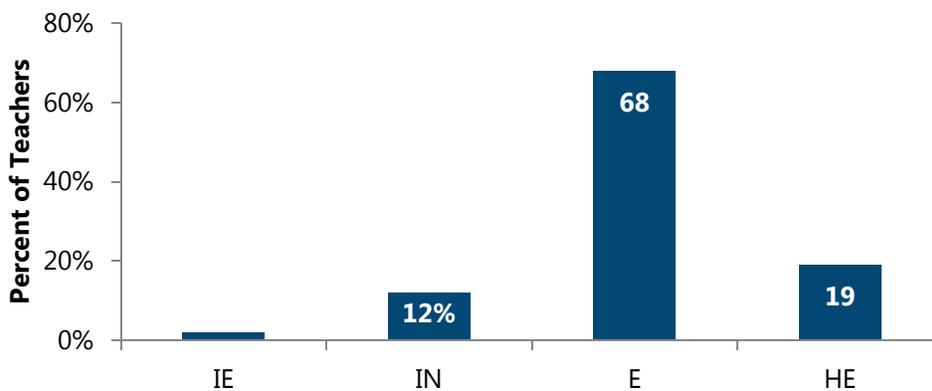
\*\* p<0.01

While all three measures are positively correlated, IGM and SLO ratings are the most highly correlated. SLOs are designed to measure student learning, so we want to see a strong correlation with IGM, which serves a similar function. A correlation of .344 indicates SLOs are a valid measure of student learning, especially for teachers without IGM data. While positively correlated with other measures, there is still room to improve the accuracy and instructional utility of SLO implementation as teachers and administrators gain fluency in the process.

## RISE Pilot Summative Evaluation Ratings

RISE summative evaluation ratings show that most teachers were rated Effective or Highly Effective.

Figure 3: Distribution by performance level of RISE pilot summative scores



A high number of teachers, 87 percent, received a summative evaluation rating of Effective or Highly Effective. Summative ratings were most largely influenced by TER data, the measure weighted most heavily in summative calculations and whose distribution included a large percentage of Effective and Highly Effective ratings. In most RISE pilot corporations, over 90 percent of teachers were rated Effective or Highly Effective on classroom observations. As covered in the end-year report, inflated, undifferentiated ratings might indicate that teachers are not receiving the feedback they deserve to improve. Work to support school leaders on classroom observation skills, so that they may accurately

identify strengths and weaknesses in teacher performance, should continue. One important reflection of progress is that only 19% of teachers received a summative rating of Highly Effective, meaning that a meaningful percentage of high performers received recognition in the pilot districts.

SLO data also contributed to high summative ratings, with 47 percent of teachers receiving a score of Highly Effective on this measure. We recommend teachers and leaders receive continued support on writing strong assessments and measuring SLOs to ensure scores accurately reflect student performance.

## Conclusion

Analysis shows that RISE is working as intended. The system's multiple measures are positively correlated and provide similar, yet not identical, ratings suggesting teachers are receiving reliable feedback on these dimensions and fair summative evaluation ratings. While these results are encouraging, a large number of teachers received Effective or Highly Effective ratings. Continued support should be given to evaluators to ensure teachers receive accurate feedback and development on all measures.

While there is still much work to do, RISE pilot findings suggest progress. Historical data shows that many evaluation systems nationwide previously rated 99 percent of teachers as Effective/Satisfactory<sup>vi</sup> and did not evaluate teachers on an annual basis.<sup>vii</sup> Providing teachers with regular observations and increased feedback based on classroom and student performance leads to improved instruction that will help Indiana schools fulfill their most important mission: preparing students for successful futures.<sup>viii</sup>

---

<sup>i</sup> RISE districts were Bloomfield Schools, Greensburg Schools, and Fort Wayne Community Schools. Alternative Districts were Beech Grove Schools, Bremen Schools, and MSD Warren Township.

<sup>ii</sup> Summative ratings were not calculated for non-RISE corporations due to missing student growth measures in the final data set. For this reason, the majority of this memo focuses on findings from RISE pilot corporations.

<sup>iii</sup> RISE measures are Teacher Effectiveness Rubric (TER), Student Learning Objectives (SLOs), Individual Growth Model (IGM) when available, and School Wide Learning (SWL). The four performance categories for all evaluation systems in Indiana is Ineffective (IE), Improvement Necessary (IN), Effective (E), and Highly Effective (HE).

<sup>iv</sup> One exception: Fort Wayne Community Schools wanted schools to remain anonymous, so we were unable to incorporate school-wide learning measures in the summative rating. Because the weight of the measure was minor (5%), summative ratings were calculated without this measure.

<sup>v</sup> Both the Measures of Effective Teaching Project (MET) study (Bill & Melinda Gates Foundation. (2010). *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project.*) and the recently released TNTP study of the Assessment of Classroom Effectiveness (TNTP. (2012). *Leap Year: Assessing and Supporting Effective First-Year Teachers.*) show positive correlations between multiple measures. The Assessment of Classroom Effectiveness had significant correlations ranging from .32 to .17 for different measures.

<sup>vi</sup> Districts that use binary evaluation ratings categorize more than 99 percent of teachers as Satisfactory (TNTP. (2009). *The Widget Effect*)

<sup>vii</sup> For this pilot, we were unable to collect meaningful historical evaluation data because most of the pilot districts did not evaluate all teachers on an annual basis prior to the implementation of RISE.

<sup>viii</sup> For more data related to the pilot, see TNTP's Summer 2012 Report: Creating a Culture of Excellence in Indiana's Schools.