

## EXECUTIVE SUMMARY

# West Virginia Revised Educator Evaluation System for Teachers 2011-2012: First Year Pilot Report

by Anduamlak Meharie and Nate Hixson, June 2013

During the 2011–2012 school year, teachers in 25 West Virginia schools from 12 counties participated in the pilot test of the new educator evaluation system. Twenty of the 25 pilot schools were participants in the federal School Improvement Grant (SIG) program.

Administrators and leadership teams attended professional development on the system in July 2011; teachers had a separate training during the following August. At the beginning of the 2011–2012 school year, administrators assigned all teachers (including specialists), to one of three progression levels based on years of teaching experience: *initial*—3 years or less (170 teachers), *intermediate*—4 or 5 years (81 teachers), and *advanced*—6 years or more (445

teachers). During the school year, depending on their progression level, teachers engaged in one or more of three main component activities as shown in Table 1.

The system has four performance levels: *distinguished*, *accomplished*, *emerging*, and *unsatisfactory*. It is based on five teaching standards and two performance standards (see Table 2). For each of the standards, critical standard elements define what educators must know and do. A predetermined weight was given to each standard and a summative rating was calculated for each participating teacher (Table 2). At the end of the pilot year, 696 teachers received summative ratings.<sup>1</sup>

### Purpose of the Study

Purposes were to (a) determine the extent to which participants adhered to the proposed evaluation model

| Component  | Number required for progression level |               |          |
|--|---------------------------------------|---------------|----------|
|  | Initial                               | Inter-mediate | Advanced |
| <b>Self-reflection</b> (14 critical elements with 4-point scale of performance levels)               |                                       |               | 1        |
| <b>Student learning goals</b> (includes two data points, rigor, and comparability across classrooms) | 2                                     | 2             | 2        |
| <b>Evidence</b>  |                                       |               |          |
| • Scheduled classroom observation (30 minutes)   | 2                                     | 1             |          |
| • Unscheduled classroom observation (30 minutes)   | 2                                     | 1             |          |
| • Supporting evidence  | Optional                              | Optional      | Optional |
| • Conference with evaluator (within 10 days of each observation)                                     | 4                                     | 2             |          |
| • End-of-year conference   | 1                                     | 1             | 1        |

| Standard   | Weight   |     |
|--|----------|-----|
| Total  | 100.00   | 100 |
| Teaching standards   |          |     |
| Standard 1: Curriculum and planning                                | 17.14    | 80  |
| Standard 2: The Learner and the Learning Environment               | 17.14    |     |
| Standard 3: Teaching   | 17.14    |     |
| Standard 4: Professional Responsibilities for Self-Renewal         | 11.14    |     |
| Standard 5: Professional Responsibilities for School and Community | 17.14    |     |
| Performance standards  |          |     |
| Standard 6: Student Learning                                       |          | 20  |
| • Student Learning Goal 1  | 7.50     |     |
| • Student Learning Goal 2  | 7.50     |     |
| • Standardized School Growth Scores                                |          |     |
| • Reading  | 2.50     |     |
| • Math   | 2.50     |     |
| Standard 7: Professional Conduct                                   | Required |     |

(implementation fidelity), (b) assess the potential contribution of the system to the professional growth of educators, (c) determine the relationship among the six professional standards, (d) identify facilitators and barriers to implementation, (e) determine if the training and support provided was sufficient to support implementation, and (f) provide a preliminary analysis of the system's ability to differentiate teacher performance. All data collected and analyzed in this study were shared during the pilot with project leaders. This report is the first summary and cross analysis.

### Methods

We conducted four surveys, including two about professional development provided to teachers and administrators; one about the overall operation of the system, deployed midway through the pilot year; and another about the overall system after educators had received their final summative ratings.

Focus group interviews were conducted to gather feedback on three broad topics: (a) initial reactions of educators to the implementation of the pilot project and its various components, (b) views about the effect of the pilot upon participants' knowledge, beliefs, and behavior, and (c) specific issues and concerns regarding the implementation of the pilot project.

We analyzed data from several electronic documents submitted by educators as required components of the system. All data were extracted from West Virginia Education Information System (WVEIS) on the Web (WOW), including (a) student learning goals worksheets, (b) classroom observation records, (c) evidence submissions, and (d) complete evaluation records ( $n = 696$ ).

### Findings

*Implementation fidelity.* Overall, analyses suggest that some components of the system were implemented well, while some will require continued monitoring. Collaboration (required for a *distinguished* rating), was included in more than three fourths of sampled student learning goals. Likewise, almost all teachers who had one or two post-observation conferences indicated they took place within 10 days of the observation, as required. Also, at least 89% of all sampled student learning goals were rated as having met the requirements for *rigor* and *comparability* as defined by the revised evaluation system.

On the other hand, about 12% of teachers in the

intermediate and initial progression levels had fewer than the required number of classroom observations, yet they received summative evaluations at the end of pilot year. Over 90% of those were in the initial progression level, which requires four observations. Also, only slightly more than half of sampled student learning goal worksheets were finalized by the deadline. Approximately three fourths of teachers in the initial and intermediate progressions indicated that they didn't have a postobservation conference in the fall semester of the pilot year. Moreover, the third criterion of student learning goals—that they employ two data points in time—appears to be the weakest aspect of the student-learning goal-setting process for teachers, indicating the need for further training. Lastly, given that administrators had the opportunity to review and approve these goals, the results suggest that they, too, need more in-depth training.

*Contributions to professional growth.* Overall, survey responses to the end-of-cycle survey indicate that the revised system contributed positively to professional growth among teachers. At least three quarters of teachers indicated various components of the system had a positive impact on them, at a moderate level or higher. Two components of the revised system—self-reflection and student learning goals—were rated very high in terms of their positive contribution to educators. Their responses suggest that the revised system has resulted, not only in a greater understanding of the WV professional teaching standards, the process of setting student learning goals, and identifying ways to achieve them, but also in increasing the frequency with which teachers practice elements of effective instructional strategies.

*Relationship among the six professional standards.* Preliminary evidence leads us to believe that at least two factors are being measured by the new evaluation system. These factors could conceivably be conceptualized as *inputs* (items related to Standards 1–5) and *outputs* (the student learning goals portion of Standard 6). Moreover, correlation data indicate that the input measures (Standards 1–5) are clearly and strongly related to one another and to a lesser extent to *some* of the output measures (i.e., student learning goals). Due to technical limitations in our operationalization of student growth—that is, the use of school-wide rather than classroom level growth data—we still have a limited understanding of how student learning is related to the five professional teaching standards.

One last note: We observed much stronger correlations among the standards for educators in our small (and unrepresentative) sample of non-SIG schools that volunteered for the pilot. This finding could be important if it holds with a larger sample, because it could mean that these variables function differently in different types of schools. This could be especially problematic in a high-stakes scenario where personnel decisions are being made based upon these outcomes. Yet, it would be unwise to attribute much meaning to these differences until we have more data.

*Facilitators and barriers.* Educators believed the revised system required too much time and added responsibilities that competed with their daily responsibilities centered on classroom instruction and school improvement efforts. This perception suggests that some educators have not fully accepted or integrated the revised system as a mechanism to improve student achievement. It is worth repeating here that 80% of pilot schools were under a plan of school improvement and were undergoing considerable change as recipients of the SIG grant. Moreover, widely reported technology-related issues negatively impacted educators' perceptions of the revised evaluation system. While the majority of the issues with the online system have been addressed, educators also expressed their desire for access to the system from home. They indicated they have little free time or privacy in the school building to complete tasks related to the revised evaluation system and their local internet access at school was not always reliable.

Notably, although the majority of teachers indicated that various components of the revised system had at least a moderate positive impact on them, a considerably smaller proportion indicated the evaluation system overall made a positive impact on them as educators. This suggests that a large proportion of teachers see value in, say, the process of setting student learning goals or self-reflection, but hold in less regard the overall benefit of the revised system for their professional growth.

*Sufficiency of training and support.* Teachers who attended training sessions gave high marks to the quality of training. However, one third of teachers in pilot schools did not attend the August 2011 training. Perhaps more important is the small proportion of teachers (less than two-thirds) who indicated that they received beneficial feedback from administrators and that the revised system has been implemented well in their schools. The revised system is heavily reliant on the ability of administrators to manage the

implementation of the system, to objectively and consistently evaluate teachers on six teaching standards utilizing various tools (e.g., observations, student learning goals, review of evidence), and to provide valuable feedback that should lead to improved effective teaching practices.

*Ability to differentiate teacher performance.* Overall, 14.5% of teachers were rated at the *emerging* level, 76.1% at the *accomplished* level, and 9.3% at the *distinguished* level. Teachers could not be rated at the *unsatisfactory* level during the pilot. A significantly greater proportion of teachers in the advanced progression received a performance rating of *distinguished* compared to teachers in the intermediate and initial progressions. The proportion of teachers rated as *distinguished* in elementary schools was approximately three times larger compared to middle and high schools. On the other hand, high schools had a comparatively larger proportion of teachers rated as *emerging* compared to middle and elementary schools. Middle schools had the largest proportion of *accomplished* teachers compared to elementary and high schools. Results of the range-of-effectiveness ratings by progression and programmatic levels were statistically significant.

*Limitations.* The findings in this report are in no way generalizable outside of the pilot schools for four main reasons (a) the sample is comprised almost entirely of educators from historically low performing SIG schools; (b) only a small number of pilot participants from non-SIG schools (N = 5 schools) volunteered to take part in the pilot, which makes for a strong probability of selection bias; (c) *unsatisfactory* ratings were not included during the pilot year, so results could change when the full breadth of ratings is included; and (d) we do not have data about quality of implementation at individual schools. Therefore, at this time we must recommend that no summative judgments be made based upon these results.

## Recommendations

The recommendations provided below are based on one year of implementation and data collection. We caution readers to keep this limitation in mind.

### 1. Provide ongoing training and support

*Provide extensive training and ongoing support for administrators in all aspects of the revised system so they can provide guidance to teachers in their schools.* The overwhelming portion of the summative evaluation (95%) is dependent upon the ability of each school administrator to carry out his or her

responsibilities effectively. Training on the new system should be incorporated into the new administrator induction process.

*Provide similar support for educators by incorporating training on the new evaluation system as a requirement for all teachers and as part of the new teacher induction process within each county.* Making this a requirement will guarantee all teachers receive a minimum standard of education related to the new system before they are subject to evaluation.

*Consider providing more rigorous and extensive training on the process of student goal setting.* This component was identified by all respondents as the most beneficial component of the system. Yet, feedback from respondents indicates that participants still consider this process to be the most challenging part of the pilot project. Training should be at least a full day with follow-up support. Examples of compliant student learning goals should be provided online that are applicable to various types of educators.

*Consider providing extensive training on the online system specifically, to individuals either at the RESA or district level who can serve as contact persons for their schools.*

*Consider making West Virginia Education Information System (WVEIS) on the Web (WOW) accessible to all educators outside of the school building.* This may allow teachers to devote more time to various components of the system outside of the school day and avoid some technical difficulties attributed to limited bandwidth at some schools. This solution can potentially result in better overall quality of implementation. Embed a mechanism into the system that allows for follow-up (e.g., collect e-mail addresses).

## 2. Establish comprehensive monitoring

*We recommend that the WVDE or counties implement a continuous monitoring process to—*

- Devote adequate resources, especially at the state level, to closely monitor the implementation of the revised system to ensure various evaluation tasks are completed on time;
- Continually assess the quality of implementation at individual schools at regular intervals; and
- Identify supplemental training needs for schools on an ongoing basis and put in place a mechanism to provide it as needs arise.

*Continue monitoring the relationships among professional teaching standards and differences*

*observed among groups of schools as the pilot is expanded.* It will be absolutely critical to re-examine all of these relationships using a representative sample of educators.

*Continue monitoring the range-of-effectiveness ratings and differences among teachers by progression level and schools and counties.*

## 3. Develop classroom-level measure of student growth.

*Develop a method to measure student growth at the classroom level and after establishing its validity and reliability, explore its inclusion in the evaluation process.* Reassess the relationship among the performance standards and student growth once a classroom level measure is established.

- This will require the development of a unique teacher identification number in WVEIS and a multistep student roster verification process.
- The roster verification process should allow educators and administrators to modify and verify each educator's roster. This will allow for adaptability for co-teaching and other unique circumstances.

## 4. Other recommendations

*Convene a technical advisory committee (TAC) charged with reviewing the revised evaluation system and providing high-level expert advice* to ensure the system meets technical rigor and is defensible in high-stakes decision-making scenarios. Initial discussions have taken place to establish this committee in advance of the 2013–2014 school year.

*Consider making revisions to the Evidence Form based upon the most commonly reported types of evidence* submitted by educators for each Critical Element during the pilot study. Removing unused categories of evidence will result in a streamlined form, which may contribute to a less cumbersome reporting experience for educators.

*Establish a protocol for managing the revision of student learning goals.* Such a protocol should be flexible enough to allow educators to revise their goals in response to legitimate contextual changes that occur throughout the year, but prescriptive enough to prevent gaming.

---

<sup>1</sup> It should be noted that educators in participating schools reserved the right not to use the pilot evaluation as the evaluation of record.