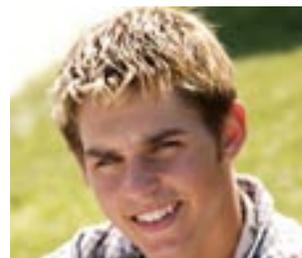




Examining district guidance to schools on teacher evaluation policies in the Midwest Region



Institute of Education Sciences
U.S. Department of Education



Examining district guidance to schools on teacher evaluation policies in the Midwest Region

November 2007

Prepared by

Chris Brandt

Learning Point Associates

Carrie Mathers

Learning Point Associates

Michelle Oliva

Learning Point Associates

Melissa Brown-Sims

Learning Point Associates

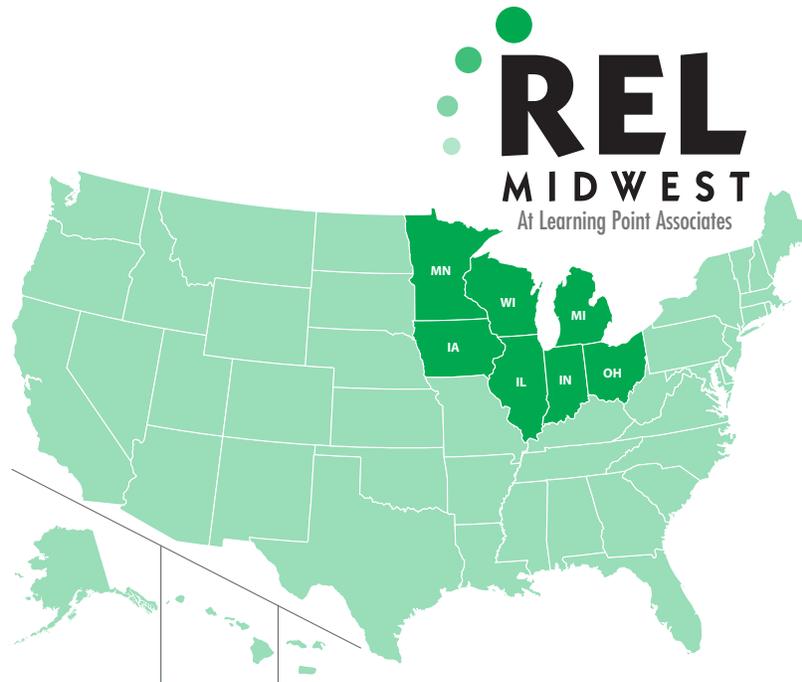
Jean Hess

Learning Point Associates



Institute of Education Sciences

U.S. Department of Education



Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

November 2007

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0019 by Regional Educational Laboratory Midwest administered by Learning Point Associates. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). *Examining district guidance to schools on teacher evaluation policies in the Midwest Region* (Issues & Answers Report, REL 2007–No. 030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the regional educational laboratory web site at <http://ies.ed.gov/ncee/edlabs>.

Summary

Examining district guidance to schools on teacher evaluation policies in the Midwest Region

This descriptive study provides a snapshot of teacher evaluation policies across a demographically diverse sample of districts in the Midwest Region. It aims to lay the groundwork for further research and inform conversations about current policies at the local, district, and state levels.

Effective teaching is a cornerstone of education reform (Whitehurst, 2002) and is critical for student academic achievement. But teachers' abilities to promote student learning vary within and across schools (Aaronson, Barrow, & Sander, 2003; Nye, Konstantopolous, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). Research finds that an important tool for improving teacher effectiveness is the teacher evaluation (Danielson & McGreal, 2000; Howard & McColskey, 2001; Shinkfield & Stufflebean, 1995; Stronge, 1995). Federal highly qualified teacher requirements have led to a surge of state and local education agencies developing new systems to evaluate teachers.

But studies of evaluation policies and their influence on teacher practice are scarce

(Peterson, 2000), and the few that exist are usually descriptive, outdated, and leave many questions unanswered. For example,

- What does the landscape of teacher evaluation policy at the district level look like today?
- What can be learned about the policy process by examining district documents?

This study—which tries to answer these two questions—is the first systematic effort to describe evaluation policies across a demographically diverse sample of districts in the Midwest Region. School district policy for evaluating teachers varies widely across the region—both in the evaluation practices specified in the policy documents and in the details of the policy prescriptions.

This study examines district evaluation policy documents for evidence of 13 common teacher evaluation practices (Ellett & Garland, 1987; Loup, Garland, Ellett, & Rugutt, 1996). In general, district policy documents were more apt to specify the processes involved in teacher

evaluation (who conducts the evaluation, when, and how often) than they were to provide guidance for the content of the evaluation, the standards by which the evaluation would be conducted, or the use of the evaluation results. District policies also varied in how specific they were, though the tendency was to be less, rather than more, specific for the 13 evaluation practices examined. Two-thirds of the district teacher evaluation policy documents provided guidance for fewer than half of the 13 practices. No policies specified more than 10 evaluation practices, and nearly 16 percent reflected *none* of these practices. The most commonly referenced practice was how often evaluations are to be conducted (67 percent), followed by what evaluation tools are to be used (59 percent) and what methods are to be used (49 percent).

The study also finds that Midwest Region districts evaluate teachers primarily to help decide whether to retain or release new teachers. School principals and administrators do most of the evaluations and, at the district's direction, focus on beginning teachers. Beginning teachers are typically evaluated two or more times a year, and experienced teachers just once every two or three years. Several other patterns emerge from the findings:

- Many district policies distinguish between beginning and experienced teachers.
- Few policies spell out consequences for unsatisfactory evaluations.
- Few districts reference using resources or guidance to support evaluations.
- Most evaluations are summative reports used to support decisions about retaining teachers and granting tenure, rather than for professional development.
- Few district policies require evaluators to be trained.
- Vague terminology leaves evaluation policies open to interpretation.
- The specificity of policy and procedures varies across districts.

The report's findings lay the groundwork for additional research, identifying several questions that need further investigation:

- What is the role of state departments of education in the teacher evaluation process?
- How do policy variations affect teacher evaluation at the local level?
- What is the influence of district policy in evaluating beginning teachers, tenured teachers, and unsatisfactory teachers?
- What is the impact of different evaluation models and practices on teacher effectiveness?

TABLE OF CONTENTS

Why this report?	1
The literature on evaluation policies	2
Evaluation policies and purposes	3
Evaluation practices	3
Addressing questions left unanswered by the literature	4
Teacher evaluation policy in the Midwest Region	5
Common patterns in teacher evaluation policies in the Midwest Region	7
Many district policies distinguish between beginning and experienced teachers	7
Fewer than half the policies detail how results are to be used	7
Few districts referenced using resources or guidance to support evaluations	7
Most evaluations are used for summative reports rather than for professional development	8
Few district policies require evaluators to be trained	8
Vague terminology leaves evaluation policies open to interpretation	8
Appendix A Methodology	9
Appendix B Study findings by research question	12
Appendix C Description of teacher evaluation study stratified random sample	18
Appendix D Summative teacher evaluation form	21
Appendix E Formative teacher evaluation form	25
References	26
Notes	29
Boxes	
1 Data and methodology	5
A1 Research questions	10
Tables	
1 Percentage of districts specifying evaluation practices within teacher evaluation policy documents	6
A1 Sample demographic characteristics and response rate	9
A2 Sample district participation rates by state	10
C1 Number of school districts in the region and in the sample by state	18
C2 Free or reduced-price lunch eligibility by region and by sample	18
C3 Minority student population by region and by sample	18
C4 District locale by state, regional data	19
C5 District locale by state, sample data	19

C6	Free or reduced-price lunch eligibility, regional data	19
C7	Free or reduced-price lunch eligibility, sample data	20
C8	Minority student population, regional data	20
C9	Minority student population, sample data	20

This descriptive study provides a snapshot of teacher evaluation policies across a demographically diverse sample of districts in the Midwest Region. It aims to lay the groundwork for further research and inform conversations about current policies at the local, district, and state levels.

WHY THIS REPORT?

Effective teaching is a cornerstone of education reform (Whitehurst, 2002) and is critical for student academic achievement. But teachers' abilities to promote student learning vary both

within and across schools (Aaronson, Barrow, & Sander, 2003; Nye, Konstantopolous, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). Research finds that an important element for improving teacher effectiveness is the teacher evaluation (Danielson & McGreal, 2000; Howard & McColskey, 2001; Shinkfield & Stufflebean, 1995; Stronge, 1995). Federal highly qualified teacher requirements have led to a surge of states and local education agencies developing new systems to evaluate teachers.

But while new evaluation systems emerge, studies of evaluation policies and their influence on teacher practice remain scarce (Peterson, 2000), and the few that exist are usually descriptive, outdated, and leave many questions unanswered. For example,

- What does the landscape of teacher evaluation policy look like today?
- What can be learned about the policy process by examining district documents?

This study—which tries to answer these two questions—is the first systematic effort to describe evaluation policies across a demographically diverse sample of districts in the Midwest Region. School district policy for evaluating teachers varies widely across the region—both in the evaluation practices specified in the policy documents and in the details of the policy prescriptions.

This study examines district evaluation policy documents for evidence of 13 common teacher evaluation practices (Ellett & Garland, 1987; Loup, Garland, Ellett, & Rugutt, 1996). In general district policy documents were more apt to specify the processes involved in teacher evaluation (who conducts the evaluation, when, and how often) than they were to provide guidance for the content of the evaluation, the standards by which the evaluation would be conducted, or the use of the evaluation results. District policies also varied in how specific they were, though the tendency was to be less, rather than more, specific for the 13 evaluation practices examined. Fully two-thirds of the

The study finds that Midwestern districts evaluate teachers primarily to help decide whether to retain or release new teachers

district teacher evaluation policy documents provided guidance for fewer than half of the 13 practices. No policies specified more than 10 evaluation practices, and nearly 16 percent of the districts' evaluation policies reflected *none* of these practices. The most commonly

referenced practice was how often evaluations are to be conducted (67 percent), followed by what evaluation tools are to be used (59 percent) and what methods are to be used (49 percent).

The study also finds that Midwestern districts evaluate teachers primarily to help decide whether to retain or release new teachers. School principals and administrators do most of the evaluation and, at the district's direction, focus on beginning teachers. Beginning teachers are typically evaluated two or more times a year, while experienced teachers just once every two or three years. Several other patterns emerge from the findings:

- Many district policies distinguish between beginning and experienced teachers.
- Few policies spell out consequences for unsatisfactory evaluations.
- Few districts reference using resources or guidance to support evaluations.
- Most evaluations are summative reports used to support decisions about retaining teachers and granting tenure, rather than for professional development.
- Few district policies specify training for evaluators.
- Vague terminology leaves evaluation policies open to interpretation.
- The specificity of policy and procedures varies across districts.

These findings lay the groundwork for additional research, identifying several questions that need further investigation:

- What is the role of state departments of education in the teacher evaluation process?
- How do policy variations affect teacher evaluation at the local level?

- What is the influence of district policy in evaluating beginning teachers, tenured teachers, and unsatisfactory teachers?
- What is the impact of different evaluation models and practices on teacher effectiveness?

THE LITERATURE ON EVALUATION POLICIES

Only two studies have examined teacher evaluation policies on a large scale (Ellett & Garland, 1987; Loup et al., 1996). Both used the Teacher Evaluation Practices Survey (TEPS) (Ellett & Garland, 1987) of superintendents to collect information about teacher evaluation policies and procedures in the nation's 100 largest school districts. The TEPS is divided into three sections: teacher evaluation policies, purposes, and practices. The policy section asks how teachers are informed of the policy, who will be evaluated, how often, and what are the standards for acceptable teaching. The purpose section asks about the primary purposes for conducting evaluations and how districts use evaluation results, whether for personnel decisions or professional development. The practices section examines who conducts the evaluation, what evaluation methods are used, and what training is required for evaluators.

Ellett and Garland's survey (1987) of superintendents and their analysis of district policy documents suggested that teacher evaluations were most often used for dismissal or remediation rather than for professional development, district policies rarely established performance standards or evaluator training, few districts permitted external or peer evaluations, and superintendents tended to present their district policies more favorably than did independent reviewers.

A decade later Loup et al. (1996) conducted a follow-up study, adapting Ellett and Garland's survey to measure superintendents' opinions about the effectiveness of their teacher evaluation systems. The results mirrored those of Ellett and Garland—teacher evaluation policies in large districts had changed little in 10 years. But superintendents

were no longer satisfied with the status quo, and many reported a need to revise their existing evaluation tools and procedures.

The following sections use the TEPS framework to review other relevant research and professional guidance on teacher evaluation.

Evaluation policies and purposes

According to Loup et al. (1996), most policies give supervisors and assistant superintendents the district-level responsibility for evaluating teachers. Principals and assistant principals are most often responsible at the building level. The conflicting roles of principals who must serve both as instructional leaders and as evaluators are the biggest problem with teacher evaluation according to Wise, Darling-Hammond, McLaughlin, and Bernstein (1984). Many principals, wanting to maintain collegial relations with staff members, are reluctant to criticize them. It is not surprising, then, that Johnson (1990) reports that many teachers believe that administrators are not competent evaluators. Several studies “depict principals as inaccurate raters both of individual teacher performance behaviors and of overall teacher merit” (Peterson, 1995, p.15). According to Dwyer and Stufflebeam (1996), principals usually rate all their teachers as performing at acceptable levels—a welcome, but unlikely, result.

How often teachers are evaluated usually depends on their years of experience. Sweeney and Manatt (1986) report that most districts observe nontenured teachers more frequently than tenured teachers. Guidance on teacher evaluation, however, recommends supervising beginning teachers rather than evaluating them (Nolan & Hoover, 2005).

The frequency of evaluation may also be related to what assessment tool a district uses. Some districts may require observations more frequently than portfolio assessments. Others may use alternative tools, such as the interview protocol and scoring rubric developed by Flowers and Hancock (2003). This tool is designed to accurately and quickly

assess a teacher’s ability to assess and modify instruction to improve student achievement.

Researchers have encouraged districts to modify their practices in ways that increase the likelihood of evaluation informing teachers’ improvement efforts, but there has not been strong evidence that this advice has had its intended effect. Researchers and teachers alike often assume that the intended effect is to inform professional development and recertification (Clark, 1993).

But although districts with performance-based compensation programs use evaluations to determine salary increases (Schacter, Thum, Reifsneider, & Schiff, 2004), few use them to improve teacher practices. Researchers have

suggested that successful evaluation depends on clear communication between administrators and teachers (Darling-Hammond, Wise, & Pease, 1983; Stronge, 1997). To facilitate this, researchers have advocated involving teachers in the design and implementation of the evaluation process (Kyriakides, Demetriou, & Charalambous, 2006). But case studies show that policies and procedures are not regularly communicated to teachers; teachers are thus more likely to see the evaluation as a summative report generated to meet district or state requirements (Zepeda, 2002) than as an opportunity to gauge and improve their teaching capacity.

Although districts with performance-based compensation programs use evaluations to determine salary increases, few use them to improve teacher practices

Evaluation practices

Stronge (2002) holds that measures of teacher planning—such as unit and lesson plans, student work samples, and analyses of student work—are related to student success and should be evaluated. Quality lesson plans should link learning objectives with activities, keep students’ attention, align objectives with the district and state standards, and accommodate students with special needs (Stronge, 2002). And for the lesson

to be effectively delivered, teachers must be able to manage the classroom. One suggested way to manage the classroom learning environment is to establish routines that support quality interactions between teachers and students and between students and their peers (Ellett & Garland, 1987; Loup et al., 1996).

Teachers often have more than one role, which can require “technological, administrative, and social skills in addition to those needed for routine teaching roles” (Rosenblatt, 2001, p. 685). Some districts may thus also evaluate teacher contributions to committees, commitment to professional development, and communication with colleagues and parents (Ellett & Garland, 1987; Loup et al., 1996).

Studies find that districts use a variety of methods to evaluate teachers: observations (Muji, 2006), teacher work samples (Schalock, 1998), interview protocols (Flowers & Hancock, 2003), and portfolio assessments (Gellman, 1992) are just a few. The portfolio assessment has garnered particular attention from administrators, teachers, and researchers. Teachers and administrators find portfolios useful tools for helping teachers self-reflect, identify strengths and weaknesses, and grow professionally (Tucker, Stronge, & Gareis, 2002). Researchers are concerned, however, whether portfolios accurately reflect what occurs in classrooms (Tucker et al., 2002).

Whether the evaluation method should depend on a teacher’s years of experience is debated. Some researchers advocate using one evaluation system for all teachers (Danielson & McGreal, 2000; McLaughlin & Pfeifer, 1988; Stronge, 1997), while others argue that using different evaluation systems for beginning and experienced teachers is more valuable (Millman & Schalock, 1997;

Stiggans & Duke, 1988; Beerens, 2000; Danielson & McGreal, 2000). Peterson (2000), for example, argues that new teacher evaluations should provide opportunities for professional

development in areas the teacher’s preparation program did not address.

Several case studies suggest that evaluators need to be trained to properly assess teacher behaviors and characteristics (Wise et al., 1984; Stiggans & Duke, 1988). Large-scale studies on teacher evaluation policies have, however, found little evidence of comprehensive evaluation training programs (Loup et al., 1996).

ADDRESSING QUESTIONS LEFT UNANSWERED BY THE LITERATURE

As the literature review makes evident, research on teacher evaluation policies is usually descriptive and often outdated. Professional guidance may be more abundant, but rarely references its research base. Many questions about district policies are left unanswered. For example,

- What does the landscape of teacher evaluation policy look like today?
- What can be learned about the policy process by examining district documents?

This study begins to answer these questions—and to identify where future research may be warranted—by looking at Midwest Region district policies on teacher evaluation. The findings will deepen the field’s understanding of evaluation policies and inform administrators’ conversations about the state of current policies.

The study selected a sample of 216 districts demographically representative of districts across the Midwest Region; 140 participated in the study and provided researchers with their teacher evaluation policies and procedures. The policies and procedures were coded according to 13 research questions adapted from the Teacher Evaluation Practices Survey (Ellett & Garland, 1987; Loup et al., 1996). (Box 1 summarizes the data sources and methodology, appendix A further discusses sample selection, and appendix B provides detailed findings.)

Research on teacher evaluation policies is usually descriptive and often outdated

BOX 1

Data and methodology

This study used a stratified sample design to select school districts that are demographically representative of the majority of districts across the Midwest Region. A data file containing 3,126 public school districts in Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin was generated using the National Center for Education Statistics Common Core of Data (U.S. Department of Education, 2004/05).

Sample districts were selected according to locale (urban, suburban, rural), the percentage of students in the district eligible for free or reduced-price lunch (less than or more than 40 percent of eligible students), and the percentage of minority student enrollment in the district (less than or more than 40 percent non-Caucasian minority students).¹ In all, 216 districts were selected.

Methodology

Sample districts were asked to provide their teacher evaluation policies and procedures. These data were first coded according 13 research questions

adapted from the Teacher Evaluation Practices Survey (TEPS) (Ellett & Garland, 1987; Loup et al., 1996) and then according to the common and unique policies and procedures reported by each district. Summary reports were created for each research question, using counts of codes and representative quotations from the documents to clarify the findings and identify areas for further inquiry.

The data were analyzed using SPSS Complex Samples add-on module, which calculates common statistics, such as the standard errors reported in this paper, based on the parameters of the plan used to select the representative sample. Because different groups had different sampling probabilities in the sample selection process, these weights were also incorporated into the data analysis by the Complex Samples add-on.

Research questions**Evaluation standards and criteria**

- What specific criteria are to be evaluated?
- What external resources were used to inform evaluations?*
- What training is required of evaluators?

- Do evaluation policies differ by content area and/or special populations?*

Evaluation processes

- How often are evaluations to be conducted?
- What evaluation tools are to be used?
- What methods of evaluation are suggested or required?
- Who has responsibility for conducting the teacher evaluation?
- When are evaluations conducted?*
- How is the teacher evaluation policy to be communicated to teachers?
- What formal grievance procedures are to be in place for teachers?

Evaluation results

- How are teacher evaluation results to be used?
- How are teacher evaluation results to be reported?

Notes

* These questions were added to the TEPS categories to capture additional details about evaluation requirements.

1. District size was initially included in the selection strata, but was dropped because size was strongly and positively correlated with district locale.

TEACHER EVALUATION POLICY IN THE MIDWEST REGION

District policy for evaluating teachers varies widely across the region—both in the evaluation practices specified in the policy documents and in the details of the policy prescriptions. Still, patterns in the data reflect some commonalities across the region. The 13 evaluation practices were grouped into three categories: evaluation standards and criteria, evaluation processes, and

evaluation results (table 1). It is clear that district policy documents frequently referenced evaluation processes (such as who conducts the evaluation, when, and how often). Of the 13 evaluation practices examined, the three most commonly referenced all pertained to process: how often evaluations are to be conducted (67 percent), which tools (checklist, rating scale) are to be used (59 percent), and what methods (observation, portfolio, professional development plans) are to be used (49 percent).

Practices pertaining to the standards and criteria that inform the evaluations are less frequently referenced. Fewer than two in five districts (39 percent) provided details in their evaluation policy about the actual criteria used to rate teacher practice. One in five districts (21 percent) indicated that they used external resources, such as evaluation models or standards frameworks, in the design of their system. Only 8 percent of all

district policies referenced any form of training or certification criteria for raters, and a mere 5 percent of all districts recommended different policies based on differences in subject area or student population.

Between these extremes—the more frequent reference to evaluation processes and the less frequent reference to evaluation standards and

TABLE 1

Percentage of districts specifying evaluation practices within teacher evaluation policy documents

Teacher evaluation practices	Total n=140	Locale		
		Rural n=61	Suburban n=67	Urban n=12
<i>Evaluation standards and criteria</i>				
What specific criteria are to be evaluated?	39 (6.1)	33 (8.4)	37 (9.3)	45 (21.2)
What external resources were used to inform evaluations?	21 (3.1)	21 (4.3)	18 (4.5)	23 (12.6)
What training is required of evaluators?	8 (2.1)	10 (2.7)	6 (3.5)	6 (6.6)
Do evaluation policies differ by content area and/or special populations?	5 (4.2)	0* (0)	8 (3.9)	41 (24.6)
<i>Evaluation processes</i>				
How often are evaluations to be conducted?	67 (4.3)	68 (5.6)	61 (6.5)	78 (12.1)
What evaluation tools are to be used?	59 (4.5)	60 (7.1)	54 (5.4)	81 (13.3)
What methods of evaluation are suggested or required?	49 (5.5)	50 (8.2)	42 (7.6)	74 (10.4)
Who has responsibility for conducting the teacher evaluation?	41 (5.6)	40 (10.6)	37 (4.8)	68 (14.6)
When are evaluations conducted?	36 (5.1)	28* (6.6)	38 (7.3)	75 (15.4)
How is the teacher evaluation policy to be communicated to teachers?	32 (5.7)	31 (8.5)	27 (8.1)	39 (19.9)
What formal grievance procedures are to be in place for teachers?	31 (4.4)	34 (7.4)	22 (4.3)	39 (17.8)
<i>Evaluation results</i>				
How are teacher evaluation results to be used?	48 (6.6)	52 (8.4)	39 (10.7)	43 (22.2)
How are teacher evaluation results to be reported?	31 (6.9)	27 (11.4)	31 (10.3)	30 (14)

* indicates a difference between rural and urban or between suburban and urban that was statistically significant using a two-tailed test at 95 percent.

Note: Numbers in parentheses are standard errors. Independent samples chi-square tests were performed to determine whether urban districts were more likely than rural or suburban districts to address any one of the 13 evaluation practices.

Source: Authors' analysis based on data described in box 1 and appendix A.

criteria—was a modest mention of how to use evaluation results. Nearly half the district policies detailed how the evaluation results were to be used (48 percent), although fewer than one-third specified how these results were to be reported to teachers (31 percent).

The extent to which the 13 evaluation practices were addressed in district policy varied. Two-thirds of district policies provided guidance for fewer than half of the 13 practices. No district policies specified more than 10 evaluation practices, and nearly 16 percent of them reflected *none* of these practices.

Independent sample chi-square tests were performed to determine whether urban districts were more likely than rural or suburban districts to address any one of the 13 evaluation practices. Compared with rural districts, urban district policies were significantly more likely to differentiate evaluation practice for special population or content area teachers ($z = -3.211, p < .05$, two-tailed) and to specify when the evaluations should be conducted ($z = -2.44, p < .05$, two-tailed). No statistically significant differences were found between the urban and suburban districts.

COMMON PATTERNS IN TEACHER EVALUATION POLICIES IN THE MIDWEST REGION

The collected district policies present a somewhat uneven picture of teacher evaluation in the sampled Midwest Region districts, but they contain many interesting leads for further research. The findings in this section pertain only to the sample and not to the region as a whole, however, because they are based on small numbers of districts that referenced a given practice within their policy documents. Districts in the sample view teacher evaluation mainly as a tool for making decisions about retaining or releasing new teachers. School principals and administrators do most of the evaluations and, at the district's direction, focus on beginning teachers. Beginning teachers are typically evaluated at least two times a year;

experienced teachers are usually evaluated once every two or three years. Several other patterns emerge and are explored in the following sections.

Many district policies distinguish between beginning and experienced teachers

Districts that articulated when and how often teacher evaluations should be conducted frequently distinguished between teachers with varying experience. Of the 94 districts that included policy or procedural statements to guide the frequency of evaluations, 88 percent (standard error = 3.2 percent) specified how often beginning teachers should be evaluated. Specifically, 63 percent (standard error = 5.9 percent) of districts required beginning teachers be evaluated at least two to three times a year, usually by a combination of informal observations, formative evaluations, and formal summative evaluations.

Fewer than half the policies detail how results are to be used

Fewer than half the districts spelled out how the teacher evaluation results would be used (48 percent, standard error = 6.6 percent) or outlined procedures for filing official grievances (31 percent, standard error = 4.4 percent). When details were provided, policy documents typically noted that evaluations would be used to make decisions about retaining probationary staff or dismissing nonprobationary staff.

Few districts referenced using resources or guidance to support evaluations

Only 29 of 140 districts (21 percent, standard error = 3.1 percent) provided documentation about the resources and guidance they used to inform the teacher evaluation process. While this seems a surprisingly low number, participating districts may leave it up to administrators and supervisors

The collected district policies present a somewhat uneven picture of teacher evaluation in the sampled Midwest Region districts, but they contain many interesting leads for further research

to find and use supporting resources. Districts may also provide administrators with resources and guidance that are not mentioned in the procedures submitted to the researchers. A follow-up study could further investigate—through interviews or surveys of district and union personnel—what resources and guidance are most widely used. Research is also needed to determine how different models of teacher evaluation affect teacher effectiveness and student achievement.

Most evaluations are used for summative reports rather than for professional development

Of the 43 districts whose policies specified how evaluation results should be reported, 34 (79 percent, standard error = 6.2 percent) specified that results should be reported in summative form. Milanowski's (2005) quasi-experimental study suggests that the type of evaluation (formative or summative) is less important to professional development than the quality of developmental support, the credibility and accessibility of a mentor, and the personal compatibility of the evaluator with the person being evaluated—especially for beginning teachers. Only 45 percent (standard error = 6.4 percent) of districts whose policies provided for how evaluation results would be used, however, stated that the results would inform professional development, intensive assistance, or remediation for teachers receiving unsatisfactory evaluations.

Few district policies require evaluators to be trained

Only 8 percent (standard error = 2.1 percent) of district policies included information about required training for evaluators. Two percent (standard error = 1.6 percent) specified only that evaluators must have obtained appropriate certification or licensure to supervise instruction. The remaining 6 percent (standard error = 2 percent) specified that, in addition to administrative certification, evaluators must participate in state-sponsored or other training to evaluate teachers. More

research is needed to examine the link between effective training and evaluation. A systematic descriptive study examining the components of current evaluator training requirements may inform future experimental research.

Vague terminology leaves evaluation policies open to interpretation

The definition of evaluation varied across districts. Even when districts used more targeted terms—formative evaluation, summative evaluation, and formal and informal evaluation, to name a few—it was often difficult to understand from the documents what the districts actually intended. The term “observation” further complicates understanding the evaluation process. Some districts described pre-observation and post-observation conferences. Others simply mentioned that observations were to take place, without describing what those observations might include. A few provided a detailed observation and evaluation process that left no question about how the process was to be conducted. Ellett and Garland (1987) and Loup et al. (1996) did not report this problem of varied definitions because their primary data source was a survey. But looking at the policies revealed an interesting pattern of vague terminology. Further study about using precise language in evaluation policies could be useful. For instance, do certain types of districts intentionally use vague language intended for broad interpretation? When, and under what conditions?

Further research is also needed to determine the extent to which districts and schools provide other opportunities for evaluation that are not formally stated in policy but that are expected or encouraged. Examples might include peer reviews, critical friends, action research, and self-evaluation. A standard survey with a full round of follow-up promptings could enable a comparison of what is contained in districts' written policy and procedures and what actually occurs in the district. A survey of principals and teachers could provide a better understanding of how effectively the formally stated district policies are implemented.

Further research is needed to determine the extent to which districts provide other opportunities for evaluation that are not formally stated in policy

APPENDIX A METHODOLOGY

Sampling procedure

This study used a stratified sample design to select school districts that demographically represented the majority of districts across the Midwest Region. A data file containing 3,126 public school districts in Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin was generated using the National Center for Education Statistics Common Core of Data. The SPSS Complex Samples add-on module was used to create the sample and to account for the sample design specifications. Districts were stratified and selected to represent the total number of districts in the region.

Two hundred sixteen sample districts were selected according to locale (urban, suburban, rural), percentage of students in the district eligible for free or reduced-price lunch (less than or more than 40 percent of eligible students), and percentage of minority student enrollment in the district (less than or more than 40 percent non-Caucasian minority students). District size was initially included in the selection strata, but it was dropped

because it was strongly and positively correlated with district locale.

Schools with more than 40 percent of students eligible for free or reduced-price lunch qualify as “school-wide program” schools, which have more flexibility in the use of Title I funds and the delivery of services. For example, staff members must work together to develop curriculum and instruction and are free to work with all students in the building. Midwest state education staff indicated that the flexibility offered to schools through the school-wide program can influence local policy decisions, including those related to teacher evaluation. The 40 percent point for minority enrollment was set after conversations with regional educators who felt that 40 percent represents a critical mass that begins to significantly affect a district’s policies and procedures.

The sample’s demographic characteristics and response rates are provided in table A1. See appendix C for tables comparing the total number of districts within each stratum—both across the region and within each state—with the sample districts.

Of the recruited districts, 76 did not participate. Further follow-up with these districts would be

TABLE A1

Sample demographic characteristics and response rate

District characteristics	Districts recruited	Districts participating	Response rate (percent)
<i>Locale</i>			
Rural	104	61	59
Suburban	99	67	68
Urban	13	12	92
Total	216	140	65
<i>Free or reduced-price lunch</i>			
More than 40 percent of students eligible	51	33	65
Less than 40 percent of students eligible	165	107	65
Total	216	140	65
<i>Minority enrollment</i>			
More than 40 percent non-Caucasian minority	14	12	86
Less than 40 percent non-Caucasian minority	202	128	63
Total	216	140	65

Source: Authors’ analysis based on data described in this appendix.

useful to determine whether nonrespondents are less likely to have formalized teacher evaluation policies and practices and thus less likely to respond because they did not understand the importance of the study to their situations.

Data collection

The Midwest Regional Educational Laboratory, with the support of the region's state education agencies, sent letters to the 216 district superintendents to inform them of the study's goals and invite them to participate. To reduce the burden on districts, the study authors downloaded teacher evaluation policies from the district web sites when possible. If the policies were unavailable, the National Opinion Research Center (NORC) used a variation of the Tailored Design Method (Dillman, 2000) to collect the data—first contacting school districts with an advance letter, followed by a reminder post card and a second reminder letter.

NORC used FedEx for the second mailing because many district officials reported not receiving the documents through U.S. Postal Service standard mail. The letters described the study goals, listed the research questions, and requested district policy and procedural documentation for teacher evaluations (see box A1 for the list of research

questions). The respondents were informed that electronic versions were preferred to facilitate analyzing the data with the NVivo software. If districts were unable or unwilling to provide electronic documents, FedEx post-paid envelopes were provided for mailing hard copies of the policy documents. NORC staff prompted districts that did not provide documentation within two weeks of the first mailing, in many cases making multiple calls. The prompters also e-mailed and faxed documentation to districts. The response rate for each of the Midwest Region states is tabulated in table A2.

TABLE A2
Sample district participation rates by state

State	Districts recruited	Districts participating	Response rate (percent)
Iowa	31	24	77
Illinois	58	30	52
Indiana	20	16	80
Michigan	38	26	68
Minnesota	23	13	57
Ohio	16	13	81
Wisconsin	30	18	60
Total	216	140	65

Source: Authors' analysis based on data described in this appendix.

BOX A1

Research questions

The research questions were adapted from the Teacher Evaluation Practices Survey (TEPS) (Ellett & Garland, 1987; Loup et al., 1996).

Evaluation standards and criteria

- What specific criteria are to be evaluated?
- What external resources were used to inform the evaluation?*
- What training is required of evaluators?

- Do evaluation policies differ by content area and/or special populations?*

Evaluation processes

- How often are evaluations to be conducted?
- What evaluation tools are to be used?
- What methods of evaluation are suggested or required?
- Who has responsibility for conducting the teacher evaluation?
- When are evaluations conducted?*

- How is the teacher evaluation policy to be communicated to teachers?
- What formal grievance procedures are to be in place for teachers?

Evaluation results

- How are teacher evaluation results to be used?
- How are teacher evaluation results to be reported?

Note

* These questions were added to the TEPS categories to capture additional details about evaluation requirements.

Eight districts explicitly declined to participate (three Illinois districts, two Michigan districts, one Minnesota district, and two Wisconsin districts). Five of the eight did not give a reason. Two indicated that they were understaffed, and one that it did not have a teacher evaluation policy or documentation available. Sixty-five districts simply did not respond to the recruitment letter.

Data analysis

Districts submitted formal and informal policy documents. Formal documents included official district documents, such as contracts, employee handbooks, district and school web sites, and district evaluation forms. Informal documents included any documents created for this study, such as e-mailed and handwritten responses to analysis questions. The informal and formal documents were coded for this study. Twenty-five districts provided both formal and informal documents and nine provided only informal documents.

The data were first coded by research question and then by the common and unique policies and procedures reported by each district. The codes were then grouped according to common themes. Summary reports were created for each research question, using counts of codes and representative quotations from the documents to clarify the findings and identify areas for further inquiry.

The data were analyzed using SPSS Complex Samples add-on module, which calculates common statistics, such as the standard errors reported in this paper, based on the parameters of the plan used to select the representative sample. Because different groups had different sampling probabilities in the sample selection process, these weights were also incorporated into the data analysis by the Complex Samples add-on.

Limitations

Given the lack of information about teacher evaluations and the nature of the policy documents

collected, this study is mainly descriptive and has several limitations.

Policy and procedural documents may not tell the whole story. All the district documents may not have been provided. Thus the study may not be an exhaustive profile of the teacher evaluation process in all the districts. And policy documents may not always reflect actual evaluation practices and processes. One district policy document stated that “a teacher’s work will be evaluated at least once every two years, and a written report shall be made on each teacher by the principal, curriculum coordinator, or supervisor.” An administrator from this district specified, however, that “new teachers are evaluated (two) times/year,” while teachers who have two years of experience are evaluated “once every two years.” Several districts also referenced formal documents that were not provided.

Only district-level policies and procedures were examined. Individual schools may have substantial autonomy in their evaluation policy and practice and may have formal documentation developed independent from the district central office. School-level policies were not included in this study, however. This study also did not look at how closely schools adhere to the district evaluation policies and procedures.

Sample sizes in some demographic categories are small. District responses are broken down by locale (urban, rural, and suburban). In some cases, the samples for these demographic categories are small and should thus be interpreted with caution. Reported categorical differences may not always reflect the region as a whole.

Teacher contracts were rarely included in the analyses. Teacher contracts likely contain information about teacher evaluations, but the Midwest Regional Educational Laboratory received very few teacher contracts from participating districts. This missing data source makes it likely that the policies used to address the research questions do not fully represent policies in the Midwest Region.

APPENDIX B

STUDY FINDINGS BY RESEARCH QUESTION

This summary of findings includes the percent of districts with policy documents addressing the 13 research questions and identifies patterns in the policy prescriptions. Standard errors are included for population estimates of districts with policy addressing each research question. The standard errors represent the variability of the estimates. The larger the reported standard error, the less precise the estimate. For estimates disaggregated by locale, see table 1 in the main body of the report. The disaggregated results by minority and free or reduced-price lunch are not discussed because differences between the groups showed little variance.

The patterns of detail described in the policy descriptions *within* each research question should be interpreted with caution. The comparisons are based on small numbers of districts that referenced a given practice within their policy documents. These findings pertain only to the sample and not to the region as a whole.

1. What specific criteria are to be evaluated?

Fifty-four (39 percent, standard error = 6.1 percent) districts identified the specific evaluation criteria. Forty-nine reported that more than one criterion should be used to evaluate teachers, 25 identified five to six criteria, and 24 fewer than four. The following sections group the district criteria into five categories used for coding.

Knowledge and instruction. The category “knowledge and instruction” refers to teacher content and pedagogical knowledge. If a district reported that a teacher’s content knowledge or pedagogical strategies should be evaluated, then the district was counted under the “knowledge and instruction” column. One rural district measured teachers’ knowledge and pedagogical practices by observing and critiquing teacher “strategies to deliver instruction that meets the multiple learning needs of students.” Of the 54 districts that detailed evaluation criteria, 81 percent included “knowledge and instruction” as a criterion.

Monitoring. “Monitoring” refers to teachers’ monitoring of student learning, assessing student performance, and reflecting on their own teaching performance. Half the districts with policies detailing the areas of evaluation included “monitoring.” Twenty-two districts dictated that the teacher evaluations should look at whether teachers examine their students’ performance through measures such as assessment data. Nine districts required that teachers be evaluated on two or more of the monitoring subcomponents. Six districts required evaluations to determine how teachers use self-reflection to respond to student needs. Eight districts required teachers to provide demonstrated knowledge of their students’ background and skills.

Professional responsibilities. Forty-three of the 54 districts (80 percent) use the “professional responsibilities” criterion, which evaluates teachers’ professional development, fulfillment of responsibilities, and communication with colleagues, students, and families as part of their evaluation. Eighteen districts required evaluations to contain information on teacher communication and feedback, and 15 district policies stated that teachers should be evaluated on how they fulfill their professional responsibilities (involvement in school and district activities, adherence to school and district policies, cultivation of professional and student relationships). Seven districts required teachers to be evaluated on their general “professional development.”

Classroom management. The “classroom management” criterion refers to a teacher’s ability to engage students and to maintain a positive learning environment. Forty-seven districts (87 percent) included classroom management as part of their evaluation, 27 required a general focus on classroom management, and 25 required an evaluation of teachers’ abilities to maintain positive learning environments (an atmosphere of respect and risk taking). Nine district documents dictated that student engagement should be part of the evaluation.

Planning and preparation. The “planning and preparation” criterion refers to teacher use of

goals, lesson plans, and school resources to enhance instruction. Thirty-six district policies focused on teachers' planning and preparation. Twenty districts used the phrase "planning and preparation" as a key component in evaluation, and 10 districts required that teachers be evaluated on use of school resources—such as technology or computers—to enhance student learning. Eight districts required an evaluation of teachers' lesson plans, and five required teachers to be evaluated on how they include instruction and student achievement goals. Seven districts required that teachers be evaluated in two or more areas related to "planning and preparation."

Other criteria. The evaluation requirements of two districts did not fit in the evaluation categories and were coded as "other." One district stated that "evaluation criteria are established by the board of education and are subject to change," while the other evaluated teachers according to "a breakdown of the standards to which each teacher is held and on which each teacher is evaluated."

2. What external resources were used to inform evaluations?

Only 29 districts (21 percent; standard error = 3.1 percent) identified resources—such as evaluation models and standards frameworks that informed their evaluation policies and procedures.

Specific evaluation models and guidance used to guide teacher evaluation practice. Of the 29 districts that provided documentation about guidance and resources used, 10 (34 percent) referenced the use of Danielson's Framework for teacher evaluation, while 4 (14 percent) noted the National Board for Professional Teaching Standards' five core propositions of accomplished teaching, Schlechy's (1998) Standard Bearer Quality Work Framework, the Mayerson Academy's Essential Principles of High Performing Learning and Leading, or Manatt's Teacher Performance Evaluation (1988).

State standards used to guide teacher evaluation practice. A similar number of district policies

identified state standards as informing teacher evaluation—two used the Illinois Teaching Standards and seven the Iowa Teaching Standards.

Other models. Five of the 29 district policies (17 percent) stated evaluators were provided with specific evaluation resources but did not identify the resources.

3. What training is required of evaluators?

Only 11 districts (8 percent, standard error = 2.1 percent) had written documentation detailing any form of training requirements for evaluators. In several cases, districts did not provide any details about the training—other than that it was about teacher evaluation—or about the length of the training. A few districts stated that trainings were one or two days long. Documentation in one district suggested that new administrators be paired with experienced administrators in the district. This program appeared to be a mentoring program for new administrators, although follow-up is needed to confirm this. The following sections describe the training categories identified in district policies and procedural documents.

State certification and licensure requirements. Five districts indicated that evaluators must obtain state school administration certification and licensure (such as Type 75 administrator's certification). These districts may thus not include peer or community member evaluations as a formal part of the evaluation process. Two districts required administrators to obtain additional certification in teacher evaluation.

State-sponsored training. Three districts required evaluators to be trained by the state department of education or an affiliated state-sponsored organization. This formal training generally occurred in the evaluator's first year in the position.

Other training provided. Four districts required other types of training. In one district first-year administrators were required to participate in a mentoring program with an experienced

administrator in the district. Two districts referred to “national” training in their documentation, but did not specify who provided this training or how long it lasted. The fourth district indicated that its administrators attend a two-day workshop on effective teacher evaluation, which was provided by an external education consultant. One of the previous four district policies stated that the state required administrators in certain districts to take state-level training. Further investigation is necessary to determine whether states actually require administrator training for teacher evaluation.

4. Do evaluation policies differ by content area and/or special populations?

District evaluation procedures were examined to determine whether they varied by content area (mathematics, language arts) or special populations (special education, English language learners). Seven of 140 participating districts (5 percent, standard error = 4.2 percent) reported having evaluation policies or procedures that differ by content area or special population.

Evaluation procedures for special education teachers. Of the seven districts, six had evaluation procedures for teachers who work with students with special needs (gifted students, special needs). In some cases, evaluators of special education teachers were required to use a rubric that varied slightly from that used for regular education teachers. One district explicitly held special education teachers to a different standard: “Teachers who are given unusual responsibilities or difficult situations in which to teach . . . will not be expected to meet the same performance standards as other teachers.”

Evaluation instrument for nonclassroom teachers. One suburban district developed a specific tool for a summative evaluation of nonclassroom teachers.

5. How often are evaluations to be conducted?

Ninety-four of the participating districts (67 percent, standard error = 4.3 percent) specified how often evaluations should occur. Eighty-three

districts further specified the frequency for beginning or “probationary” teachers (59 percent), 81 for experienced or “tenured” teachers (58 percent), and 7 for unsatisfactory teachers (5 percent).

Of the 83 districts with policies directing the frequency of evaluation for beginning teachers, 59 specified that beginning teachers must be evaluated at least two or three times a year. Fourteen required at least one evaluation a year for beginning teachers, while the remaining 10 included other evaluation instructions.

Of the 81 districts with policy addressing the frequency of evaluation for experienced teachers, 42 required that experienced or “tenured” teachers be observed once every three years. Twenty-two required at least one evaluation a year for experienced teachers, while six districts required more than one a year. Three districts required an evaluation only once every four or five years. One district policy indicated that a formal evaluation be conducted every three years for tenured teachers, with several informal evaluations occurring between cycles.

Surprisingly, district policies rarely specified procedures for unsatisfactory teachers. Only seven of the 140 districts participating in the study (5 percent) contained statements about unsatisfactory teachers. Five required that unsatisfactory teachers be evaluated several times a year. One district did not indicate a specific number of evaluations, but required that an intensive improvement plan be implemented: “If the summative evaluation is below district standards, the evaluator shall set forth in writing steps that shall be taken to improve the performance to meet the district standards.”

6. What evaluation tools are to be used?

Policies were analyzed to determine what evaluation tools districts use in the evaluation process. Of the 140 participating districts, 83 (59 percent, standard error = 4.5 percent) specified tools for schools to use in evaluating teachers. Most tools were used for summative evaluations and included

quantitative rating scales to assess teacher performance. In some cases, these quantitative forms included a limited number of open-ended questions, allowing the rater to support his or her ratings with descriptive information. A few districts also used classroom observation tools that allowed evaluators to describe the classroom and teacher and student behaviors. In a few cases policies coupled evaluation tools with teacher self-assessment tools. Examples of typical summative and formative tools are provided in appendixes D and E.

7. What methods of evaluation are suggested or required?

Of the 140 participating districts, 68 (49 percent, standard error = 5.5 percent) reported that their written policy either suggested or required specific evaluation methods. The three methods specifically noted were classroom observations, portfolio assessment, and professional development plans.

Classroom observations. Of the 68 districts, 41 (60 percent) suggested or required formal observations, including scheduled observations. The definition of “formal observation” was not always clear in district documentation, but the language suggested that it often included a “pre-conference, formal observation, and post-conference.” Teachers were aware ahead of time that they would be observed. An evaluation instrument was often used to document feedback, which was later shared with the teacher and placed in a permanent file. Twenty-four of these 41 districts (35 percent of the total 68 districts), in addition to requiring formal observations, also suggested or required informal observations, including unscheduled or unannounced observations. Districts requiring formal observations also suggested the use of informal observation. Informal observations were often referred to as classroom “walkthroughs” or “visitations.” Furthermore, 21 of the 68 districts (31 percent) articulated specific evaluation methods for new, nontenured, and probationary teachers, and 10 (15 percent) had distinct evaluation methods for tenured teachers.

Portfolio, professional development and growth plans, and other evaluation methods. Of the 68

districts, 13 (19 percent) required portfolios, 5 (7 percent) required individual professional development or growth plans, and 5 (7 percent) suggested or required other evaluation methods such as daily contact, or “weekly lesson plans reviewed by principal at least six times a year.” Most professional development or growth plans were designed “to improve the quality of instruction through a process of goal setting and collegial sharing.” Such plans typically included requirements to align the professional development plan with district and school goals.

8. Who has responsibility for conducting the teacher evaluation?

Of the 140 districts that provided policy and procedural documentation, 57 (41 percent, standard error = 5.6 percent) identified who is responsible for conducting teacher evaluation. Forty-four identified building administrators—principals, vice principals, or content specialists—as responsible. Seven districts (5 percent) had policies that directed district administrators and supervisors to conduct teacher evaluations. Two suburban districts had rather unique strategies of peer evaluations under certain circumstances. One allowed for “co-teaching” as an evaluation method after the teacher completed the first year with a satisfactory rating. The other required all nontenured teachers to be evaluated at least once a semester by a director, but peers could also provide feedback on lesson plans, exams, and instruction.

9. When are evaluations conducted?

The timely provision of constructive feedback is one of the important aspects of evaluation, especially for new teachers who may find the first year or two of teaching tremendously challenging. Fifty-one districts (36 percent, standard error = 5.1 percent) included language about when to conduct evaluations for beginning (or probationary) and for career (tenured) teachers.

District documentation specified the timing of evaluations for beginning teachers more often than for teachers with more experience. Twenty out of

the 51 district policies mentioned above contained language concerning new teachers (probationary, nonpermanent). Many rural and urban school district policies specified when beginning teachers are to be evaluated—usually in the fall. The remaining 31 district policies did not have different policies for teachers with different levels of experience, but included general language such as “teacher evaluations will take place throughout the school year.”

10. How is the teacher evaluation policy to be communicated to teachers?

Of 140 districts, 45 (32 percent, standard error = 5.7 percent) communicated evaluation policy to teachers: 24 percent through teacher contracts,¹ 36 percent through the teacher handbooks, 36 percent through group orientation, and 33 percent through one-on-one communication with the supervisor. Most of the 45 districts included more than one of these methods in their policy. The fact that only one-third of participating districts included information specifying how district evaluation policies should be communicated to teachers raises questions about the consistency of evaluation practice, both within and across schools, and about the criteria used in making decisions about professional development, tenure, and termination.

11. What formal grievance procedures are to be in place for teachers?

Of the 140 participating districts, 44 (31 percent, standard error = 4.4 percent) submitted policies that included teachers’ rights to file grievances. Most districts authorized teachers to provide an addendum to the evaluation (82 percent), but some gave teachers the right to request another evaluation (7 percent), to request a different evaluator (5 percent), or to allow an arbitrator to review the evaluation (5 percent).

12. How are teacher evaluation results to be used?

The written evaluation policies of 67 of the 140 participating districts (48 percent, standard error = 6.6 percent) stipulated how the evaluation results should be used: to inform personnel

decisions (60 percent), to suggest instructional improvements (39 percent), to inform professional development goals (28 percent), and to use for remediation (12 percent).

To inform personnel decisions. Policies from 40 districts (60 percent) state that evaluation results should be incorporated into teacher employment status and personnel decisions, in particular those of probationary and nontenured teachers.

To suggest improvement. The policies from 26 districts (39 percent) called for evaluations to help teachers improve their practice and in particular their identified areas of weakness. Nineteen (28 percent) district documents stated that the evaluation results would be used to determine professional development needs. The resulting professional development initiatives may be internal to the district or external. Documentation from eight districts (12 percent) was coded as “remediation reevaluation,” which meant that a poor evaluation resulted in “intensive assistance” for the teacher.

Other uses. Eight district policies (12 percent) included only very broad uses for evaluations. One district indicated that teacher evaluation results were used “to comply with mandates.” Another district used results “to improve the educational program.” Such all-encompassing uses allow for expansive interpretation.

13. How are teacher evaluation results to be reported?

Of the 140 participating districts, 43 (31 percent, standard error = 6.9 percent) had policies stating how evaluation results should be compiled and reported. Several reporting methods emerged, including the use of summative evaluation forms (79 percent), written narratives (30 percent), formative evaluation (19 percent), and teacher conferences (14 percent).

Summative evaluation forms. Of the 43 districts that specified how to report the evaluation results, 34 (79 percent) required that a completed summative evaluation form be signed by the evaluator and the

teacher and then stored in the teacher's personnel file. In some cases, the superintendent also received a copy. The form itself was often the data-gathering tool used to assess the teacher's performance.

Written narratives. Documentation from 13 districts (30 percent) explicitly required a written narrative. One policy stated that "narrative summaries are to include specific teaching assignments, length/service in the district, dates of observations, specific strengths, weaknesses or areas cited for improvement needed, improvement tasks required, specific contract recommendations and the teacher's signature and the evaluator's signature." The narratives were often considered the summative evaluation report.

Multiple formative evaluations. Eight policies use data from formative evaluations conducted throughout the school year to complete a summative evaluation form. The formative evaluation forms were typically required for classroom observations. In some districts the set of behaviors and practices observed in the formative and summative evaluation was the same.

Conferences. Documentation from six districts explicitly stated that the administrator, supervisor, or evaluator was to hold a conference with the teacher to discuss summative evaluation findings. Some districts required that the conference take place before the summative evaluation report was filed.

APPENDIX C

DESCRIPTION OF TEACHER EVALUATION STUDY STRATIFIED RANDOM SAMPLE

TABLE C1

Number of school districts in the region and in the sample by state

State	Districts in region	Districts in sample
Iowa	370	31
Illinois	887	58
Indiana	291	20
Michigan	553	38
Minnesota	348	23
Ohio	240	16
Wisconsin	437	30
Total	3,126	216

Source: Authors' analysis based on data described in appendix A.

TABLE C2

Free or reduced-price lunch eligibility by region and by sample

Free or reduced-price lunch eligibility	Districts in region	Districts in sample
Districts with less than 40 percent of students eligible	2,390	166
Districts with more than 40 percent eligible	736	50

Source: Authors' analysis based on data described in appendix A.

TABLE C3

Minority student population by region and by sample

Minority student population	Districts in region	Districts in sample
Less than 40 percent non-Caucasian minority	2,873	202
40–100 percent non-Caucasian minority	248	14
Missing	5	0

Source: Authors' analysis based on data described in appendix A.

TABLE C4

District locale by state, regional data

State	District locale			Total
	Urban	Suburban	Rural	
Iowa	15	92	262	369
Illinois	34	516	337	887
Indiana	25	143	122	290
Michigan	37	269	247	553
Minnesota	14	116	215	345
Ohio	22	149	69	240
Wisconsin	26	189	222	437
Total	173	1,474	1,474	3,121

Source: Authors' analysis based on data described in appendix A.

TABLE C5

District locale by state, sample data

State	District locale			Total
	Urban	Suburban	Rural	
Iowa	2	6	23	31
Illinois	2	33	23	58
Indiana	2	10	8	20
Michigan	3	18	17	38
Minnesota	1	8	14	23
Ohio	1	10	5	16
Wisconsin	2	13	15	30
Total	12	98	100	216

Source: Authors' analysis based on data described in appendix A.

TABLE C6

Free or reduced-price lunch eligibility, regional data

State	Free or reduced-price lunch eligibility		Total
	Less than 40 percent of students eligible	More than 40 percent of students eligible	
Iowa	310	60	370
Illinois	679	208	887
Indiana	223	68	291
Michigan	362	191	553
Minnesota	246	102	348
Ohio	184	56	240
Wisconsin	386	51	437
Total	2,390	736	3,126

Source: Authors' analysis based on data described in appendix A.

TABLE C7

Free or reduced-price lunch eligibility, sample data

State	Free or reduced-price lunch eligibility		Total
	Less than 40 percent of students eligible	More than 40 percent of students eligible	
Iowa	25	6	31
Illinois	38	20	58
Indiana	13	7	20
Michigan	28	10	38
Minnesota	19	4	23
Ohio	13	3	16
Wisconsin	30	0	30
Total	162	48	216

Source: Authors' analysis based on data described in appendix A.

TABLE C8

Minority student population, regional data

State	Non-Caucasian minority student population		Total
	Less than 40 percent	More than 40 percent	
Iowa	366	3	369
Illinois	758	129	887
Indiana	279	11	290
Michigan	507	46	553
Minnesota	330	15	345
Ohio	217	23	240
Wisconsin	416	21	437
Total	2,873	248	3,121

Source: Authors' analysis based on data described in appendix A.

TABLE C9

Minority student population, sample data

State	Non-Caucasian minority student population		Total
	Less than 40 percent	More than 40 percent	
Iowa	31	0	31
Illinois	48	10	58
Indiana	20	0	20
Michigan	35	3	38
Minnesota	23	0	23
Ohio	16	0	16
Wisconsin	29	1	30
Total	196	14	216

Source: Authors' analysis based on data described in appendix A.

**APPENDIX D
SUMMATIVE TEACHER EVALUATION FORM**

Teacher _____ Administrator _____

Building _____ Subject/Grade Level(s) _____ Date _____

Dates of Observations: _____

Date(s) of I.D.P. Planning/Goal Setting _____

Tenure: _____ Yes _____ No

Probationary: _____ 1st year _____ 2nd year _____ 3rd year _____ 4th year Mentor: _____

Is a Professional/Individual Development Plan part of this Evaluation? _____ Yes _____ No

I. Instruction

S = Satisfactory NI = Needs improvement UN = Unsatisfactory NA = Not applicable/not observed

Overall rating	S	NI	UN	NA
1. Prepares for assigned classes and responsibilities (shows evidence of adequate preparation)	S	NI	UN	NA
2. Demonstrates clear purpose and objectives	S	NI	UN	NA
3. Provides instruction at the appropriate level of difficulty for each learner	S	NI	UN	NA
4. Responds to the efforts of the learners and adjusts instruction to maximize learning by using a variety of methods and materials	S	NI	UN	NA
5. Provides opportunities for active involvement of the learner	S	NI	UN	NA
6. Monitors learning interactions and checks learners for understanding	S	NI	UN	NA
7. Implements District approved curriculum	S	NI	UN	NA
8. Demonstrates competency in subject matter	S	NI	UN	NA
9. Appropriately assesses and records learner performance	S	NI	UN	NA
10. Demonstrates productive use of time on task	S	NI	UN	NA

11. Appropriately utilizes available technological resources	S	NI	UN	NA
12. Organizes instruction and monitors achievement toward mastery learning for all students	S	NI	UN	NA
13. Utilizes current research-based instructional strategies to enhance learning	S	NI	UN	NA
14. Monitors and adjusts to accommodate learning styles	S	NI	UN	NA

Comments/recommendations on instruction: _____

II. Environment

S = Satisfactory NI = Needs improvement UN = Unsatisfactory NA = Not applicable/not observed

Overall rating	S	NI	UN	NA
1. Establishes an environment that focuses on student learning	S	NI	UN	NA
2. Takes all necessary and reasonable precautions to provide a healthy and safe environment	S	NI	UN	NA
3. Utilizes equipment, materials, and facilities appropriately	S	NI	UN	NA
4. Treats individuals within the school community with dignity and respect	S	NI	UN	NA

Comments/recommendations on instruction: _____

III. Communication

S = Satisfactory NI = Needs improvement UN = Unsatisfactory NA = Not applicable/not observed

Overall rating	S	NI	UN	NA
1. Demonstrates active listening skills	S	NI	UN	NA
2. Establishes and maintains open lines of communication	S	NI	UN	NA
3. Demonstrates effective verbal and written communication	S	NI	UN	NA

Comments/recommendations on instruction: _____

IV. Policy and procedures

S = Satisfactory NI = Needs improvement UN = Unsatisfactory NA = Not applicable/not observed

Overall rating	S	NI	UN	NA
1. Maintains records as required by law, district policy, and administrative regulations	S	NI	UN	NA
2. Attends and participates in district, faculty and departmental meetings	S	NI	UN	NA
3. Abides by school district policies, building procedures, master agreement and state and federal law	S	NI	UN	NA

Comments/recommendations on instruction: _____

APPENDIX E FORMATIVE TEACHER EVALUATION FORM

Class taught _____ Grade _____

Date _____ Teacher _____

Pre-observation report	Post-observation report
1. What topic/unit will be taught? Is this new input, practice on objectives, review, or a diagnostic lesson?	
2. What are the objectives for this lesson?	2. Were the objectives observed during the lesson?
3. What procedure will the teacher use to accomplish the objectives?	3. Were the procedures implemented? Were they effective?
4. What activities will the students be doing?	4. Were the student activities implemented as planned? Were they effective?
5. Which particular criterion/criteria do you want monitored?	5. Indicate pertinent data gathered relevant to the criteria.

- A. Demonstrates effective planning skills
- B. Implements the lesson plan to ensure time on task.
- C. Provides positive motivational experiences
- D. Communicates effectively with students.
- E. Provides for effective student evaluation.
- F. Displays knowledge of curriculum and subject matter.
- G. Provides opportunities for individual differences.
- H. Demonstrates skills in classroom management
- I. Sets high standards for student behavior.
- J. Other:

REFERENCES

- Aaronson, D., Barrow, L., & Sander, W. (2003). *Teachers and student achievement in the Chicago public high schools* (Working Paper 2002-28; Rev. ed.). Chicago: Federal Reserve Bank of Chicago. Retrieved January 31, 2007, from <http://www.chicagofed.org/publications/working-papers/papers/wp2002-28.pdf>
- Beerens, D. R. (2000). *Evaluating teachers for professional growth*. Thousands Oak, CA: Corwin Press.
- Clark, D. (1993). *Teacher evaluation: A review of the literature with implications for evaluators*. Long Beach, CA: California State University. (ERIC Document No. ED359174)
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285–328.
- Dey, I. (1993). *Qualitative data analysis: A user friendly guide for social scientists*. London: Routledge and Kegan Paul.
- Dillman, D. (2000). *Mail and internet surveys: The tailored design method*. New York: John Wiley & Sons, Inc.
- Ellett, C. D., & Garland, J. (1987). Teacher evaluation practices in our 100 largest school districts: Are they living up to “state-of-the-art” systems? *Journal of Personnel Evaluation in Education*, 1, 69–92.
- Evaluation of Certified Employees, Illinois Compiled Statutes, §§105-5- 24a. Retrieved May 1, 2007, from <http://www.ilga.gov/legislation/ilcs/ilcs4.asp?DocName=010500050HArt%2E+24A&ActID=1005&ChapAct=105%26nbsp%3BILCS%26nbsp%3B5%2F&ChapterID=17&ChapterName=SCHOOLS&SectionID=49335&SeqStart=140500000&SeqEnd=141500000&ActName=School+Code%2E>
- Flowers, C. P., & Hancock, D. R. (2003). An interview protocol and scoring rubric for evaluating teacher performance. *Assessment in Education: Principles, Policy & Practice*, 10(2), 161–168.
- Gellman, E. S. (1992). The use of portfolios in assessing teacher competence: Measurement issues. *Action in Teacher Education*, 14(4), 39–44.
- Goldstein, J. (2007, May). Easy to dance to: Solving the problems of teacher evaluation with peer assistance and review. *American Journal of Education*, 113, 479–508.
- Howard, B. B., & McColskey, W. H. (2001). Evaluating experienced teachers. *Educational Leadership*, 58(5), 48–51.
- Johnson, S. M. (1990). *Teachers at work: Achieving success in our schools*. New York: Basic Books.
- Kennedy, M. (1998). Education reform and subject matter knowledge. *Journal of Research in Science Teaching*, 35, 249–263.
- Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research*, 48(1), 1–20.
- LeCompte, M. D. (2000). Analyzing qualitative data. *Theory into Practice*, 39(3), 146–154.
- Loup, K. S., Garland, J. S., Ellett, C. D., & Rugutt, J. K. (1996). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest school districts. *Journal of Personnel Evaluation in Education*, 10(3), 203–226.
- Manatt, R. P. (1988). Teacher performance evaluation: A total systems approach. In S. J. Stanely and W. J. Popham (Eds.), *Teacher Evaluation: Six prescriptions for success* (pp. 79–108). Alexandria, VA: Association of Supervision and Curriculum Development.
- McLaughlin, M. W., & Pfeifer, R. S. (1988). *Teacher evaluation: Improvement, accountability, and effective learning*. New York: Teachers College Press.

- Milanowski, A. (2005). Split roles in performance evaluation—a field study involving new teachers. *Journal of Personnel Evaluation in Education*, 18(3), 153–169.
- Millman, J., & Del Shalock, D. (1997). Beginnings and introductions. In J. Millman, *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.
- Mujis, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, 12(1), 53–74.
- Nolan, J. Jr., & Hoover, L. A. (2005). *Teacher supervision and evaluation: Theory into practice*. New York: John Wiley and Sons.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237–257. Retrieved January 31, 2007, from <http://www.sesp.northwestern.edu/publications/169468047044fcbd1360b55.pdf>
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. Retrieved January 31, 2007, from <http://edpro.stanford.edu/Hanushek/admin/pages/files/uploads/teachers.econometrica.pdf>
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252.
- Rosenblatt, Z. (2001). Teachers' multiple roles and skill flexibility: Effects on work attitudes. *Educational Administration Quarterly*, 37(5), 684–708.
- Schacter, J., Thum, Y. M., Reifsneider, D., & Schiff, T. (2004). *The teacher advancement program report two: Year three results from Arizona and year one results from South Carolina TAP schools*. Santa Monica, CA: Milken Family Foundation.
- Schalock, M. (1998). Accountability, student learning, and the preparation and licensure of teachers: Oregon's Teacher Work Sample Methodology. *Journal of Personnel Evaluation in Education* 12 (3), 269–285.
- Schlechty, P. (1998). *Inventing Better Schools: An Action Plan for Educational Reform*. San Francisco: Jossey-Bass.
- Shinkfield, A. J., & Stufflebeam, D. L. (1995). *Teacher evaluation: Guide to effective practice*. Boston: Springer.
- Stiggins, R. (2001). *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Stiggins, R., & Duke, D. L. (1988). *The case for commitment to teacher growth: Research on teacher evaluation*. Albany, NY: State University of New York Press.
- Stronge, J. H. (1995). Balancing individual and institutional goals in educational personnel evaluation: A conceptual framework. *Studies in Educational Evaluation*, 21(2), 131–151.
- Stronge, J. H. (1997). Improving schools through teacher evaluation. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 1–23). Thousand Oaks, CA: Corwin Press.
- Stronge, J. H. (2002). *Qualities of Effective Teachers*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Sweeney, J., & Manatt, R. (1986). Teacher evaluation. In R. Berk (Ed.), *Performance assessment: methods and applications*. Baltimore: John Hopkins University Press.
- Sykes, G. (1999). Make subject matter count. *Journal of Staff Development*, 20(2), 50–51.
- Teacher's Tenure Act, 4 of 1937 Mich. Comp. Laws. §§38.93. Retrieved May 1, 2007, from, [http://www.legislature.mi.gov/\(S\(lp1cx5asdwszcu2avubdydbfl\)\)/mileg.aspx?page=getObject&objectName=mcl-38-93](http://www.legislature.mi.gov/(S(lp1cx5asdwszcu2avubdydbfl))/mileg.aspx?page=getObject&objectName=mcl-38-93)
- Tucker, P. D., Stronge, J. H., & Gareis, C. R. (2002). *Handbook on teacher portfolios for evaluation and professional development*. Larchmont, NY: Eye on Education.

U.S. Department of Education, National Center for Education Statistics. *Common core of data, 2004/05*. Retrieved from <http://nces.ed.gov/ccd/>

Whitehurst, G. J. (2002, April 24). Teacher recruitment, preparation, and development. (Statement made before the House Subcommittee on Labor/Health and Human Services/Education Appropriations). Washington, DC: U.S. Department of Education. Retrieved January 31, 2007, from <http://www.ed.gov/Speeches/04-2002/20020424d.html>

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Case studies for teacher evaluation: A study of effective practices*. Santa Monica, CA: Rand Corporation.

Zepeda, S. (2002). Linking portfolio development to clinical supervision: A case study. *Journal of Curriculum and Supervision, 18*(25), 83–102.

NOTES

The Midwest Regional Education Laboratory would like to recognize the National Opinion Research Center for its contribution to this report. Its support in collecting district teacher evaluation policies was critical to the completion of this report.

1. Few teacher contracts were collected. The finding that evaluation policy was communicated through teacher contracts was based on references identified in other district documents.