

# ***Evaluation of Toledo Public School District Peer Assistance and Review Plan***

Prime Contractor:

John F. Kennedy School of Government  
Harvard University 79 John F. Kennedy Street  
Cambridge, MA 02138

*Dr. Linda Kaboolian*

*Mr. Paul Sutherland, Esq.*

*With the assistance of Allyne Beach*

Subcontractor:

American Institutes for Research

*Andrew J. Wayne*

*Toks Fashola*

*Steve Fleischman*

*With the assistance of Sandra Tang*

***April, 2005***

## Introduction

The finding that teacher quality is one of the most important factors influencing student achievement has led school boards and educators around the country to closely examine such issues as how teachers are prepared and how teachers are supported and evaluated. How teachers are supported and evaluated is major focus in Toledo. In 1981, Toledo adopted a system in which teachers are supported and evaluated by their peers—a system called the Toledo Plan.

Since being established, the Toledo Plan has been replicated in a handful of district around the country. Systems that use teachers to evaluate and support other teachers are called Peer Assistance and Review (PAR) programs.<sup>1</sup> In the more than 20 years since its inception, the Toledo Plan has evolved but has never been systematically evaluated. During the 2003-04 academic year, the Plan was accused by members of the community of discriminating against black interns. This accusation had arisen in the past and was the subject of pending legal action by an intern who was not renewed.

We responded to a Request for Proposals (RFP) issued by the Toledo Public Schools. The RFP provided a detailed series of questions that needed examination. We were to perform a comprehensive evaluation and report to the school board and community. This report is the outcome of that evaluation. It is organized according to the **topic headings** in the State of Work that we received from the Toledo Public Schools.

Task 1: Review and analyze the Toledo Plan's implementation guidelines

Task 2: Compare the Toledo Plan's guidelines and procedures with similar efforts to judge the performance of probationary and veteran teachers.

---

<sup>1</sup> An alternative label used to describe these programs is Peer Assistance and Peer Review (PAPR) programs, but the Toledo Plan documents and most writers refer to them as Peer Assistance and Review (PAR) programs.

Task 3. Compare and contrast both from teacher quality and a cost per non-renewal perspective, the effectiveness and efficiency of the TP to teacher evaluation and non-renewal systems utilized by other large urban districts.

Task 4. Interview school district and union leadership regarding their perceptions of the effectiveness and efficiency of the TP.

Task 5 Interview community stakeholders regarding their perception of the effectiveness and efficiency of the Toledo Plan.

Task 6: Provide recommendations regarding the Toledo Plan

We urge readers to examine the detailed statement of work and the research design, both of which are files on the compact disk.

## **Task 1: Review and analyze the Toledo Plan's implementation guidelines.**

This portion of the review of the Toledo Plan focuses on the effectiveness and efficiency of the system in helping teachers improve their instruction and in retaining entry level and veteran teachers who are meeting or exceeding the classroom performance standards. The Request for Proposal (RFP) from Toledo Public Schools defined effectiveness and efficiency as follows:

- *Effectiveness* is defined as the Toledo Plan's ability to accurately identify those individual who cannot, or will not, met the professional performance standards set by the Toledo Public Schools and the Toledo Federation of Teachers.
- *Efficiency* is defined as achieving the maximum positive impact on the teaching force and student population with the resources available to the Toledo Plan.

We have examined the Toledo Plan's guidelines as well as the forms, handbooks, and other documents used to implement the guidelines. To gain greater knowledge of how the Toledo Plan is implemented, we conducted four hours of informational interviews in January 2005.<sup>2</sup>

Our review involved examination of existing research studies as well as literature on 'best practices' in PAR, teacher evaluation, and teacher support. We also examined a variety of documents describing practices in other districts that use similar approaches. We draw on these resources in the text of our report in order to compare the Toledo Plan to other programs' practices and to current research findings and provide a complete bibliography of the print sources at its conclusion.

### **How PAR programs work**

Exhibit 1 portrays the operation of a PAR program. The Toledo Plan and other PAR programs follow the same basic model, involving a

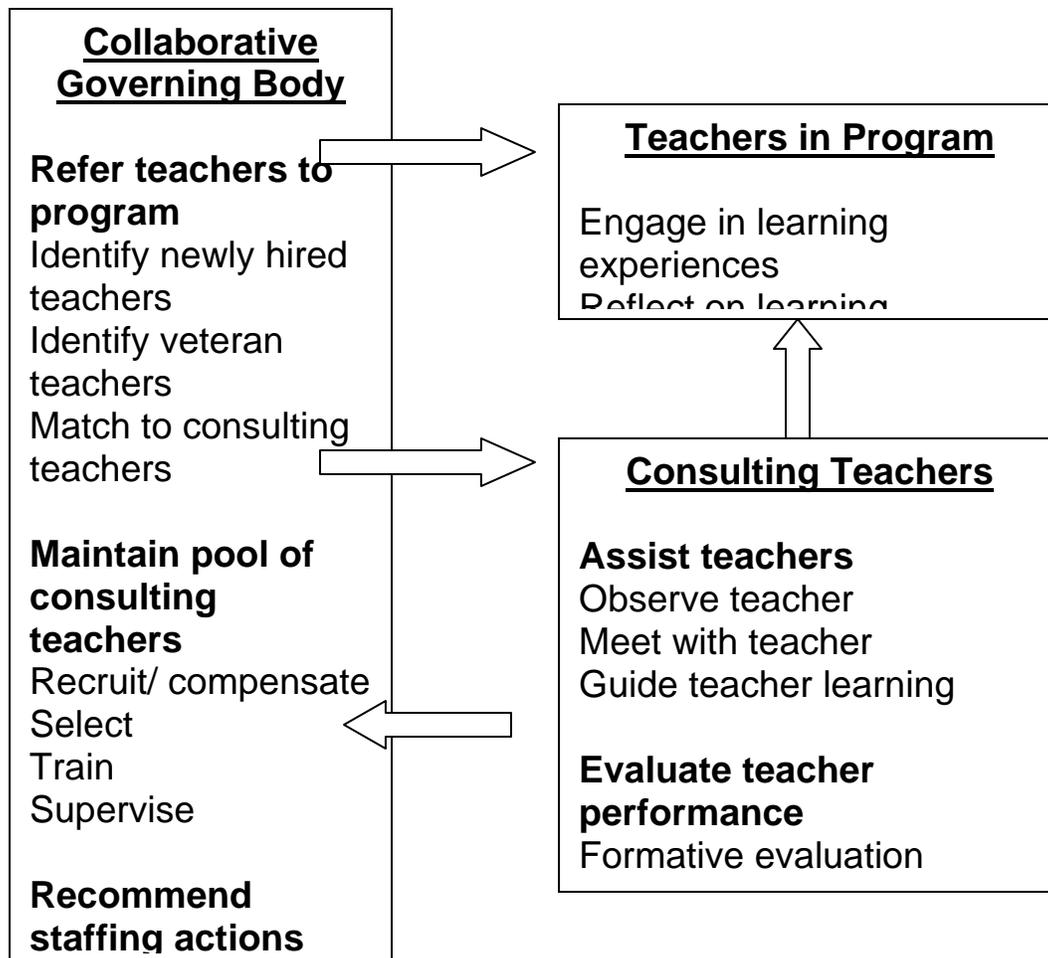
---

<sup>2</sup> These interviews include the following individuals: Dr. Sheila Austin, Chief of Staff, Toledo Public Schools (TPS); Mr. Craig Cotner, Chief Academic Officer, TPS; Ms. Francine Lawrence, President, Toledo Federation of Teachers; one Toledo Plan consulting teacher (interviewed on an anonymous basis); and one teacher who recently participated in the program (interviewed on an anonymous basis).

collaborative governing body, consulting teachers (CTs), and the teachers who are participating in the program.

The governing body is typically a board with approximately equal representation from the union and from district administration. That body's primary functions are to maintain a pool of CTs and to recommend staffing actions, on the basis of data from the consulting teachers.

### Exhibit 1: Peer assistance/ peer review program activities



The CTs evaluate teacher performance but also assist teachers so that their performance will improve. CTs are responsible for reporting the results of their evaluation activities to the governing body.

Generally, teachers who participate in the program consist of new teachers, as most PAR programs require all new teachers to participate. Veteran teachers also participate. The governing body refers veteran teachers to the program after a process involving a nomination from someone in the teacher's building, observation by CTs, and subsequent deliberations of the governing body.

The first section of this report provides an analysis of traditional teacher evaluation practices as a context in which to examine the Toledo plan. Traditional evaluation practices have important limitations in helping teachers improve their instruction and in accurately measuring teacher performance.

The remaining sections of the report focus on individual elements of Toledo's PAR program, and compares those elements to research-based practices and to practices from other PAR programs. We focus on the following elements because they affect either (a) how accurately the program measures teacher performance or (b) the degree to which the program helps teachers improve their teaching:

- The Toledo Plan standards
- Gathering and using data on teacher performance
- Mentoring/assistance
- Consulting teacher selection, training, and supervision

## **Traditional Teacher Evaluation Practices**

A good system for evaluating classroom teaching should be effective at identifying underperforming teachers, helping them improve their teaching, and—if necessary—efficiently terminating their employment. Traditional teacher evaluation practices have achieved these goals only inconsistently, for a variety of reasons. Below we describe how teacher evaluation is typically conducted and then

review the challenges that traditional evaluation systems have had at accurately identifying teachers, assisting them, and facilitating termination of employment.

## **How teacher evaluation is typically conducted**

In the traditional approach, classroom teaching performance is evaluated by a principal, assistant principal, or other administrator in the building.<sup>3</sup> Principals and assistant principals typically receive training in supervision and evaluation as part of their credentialing programs and typically have experience as classroom teachers. In middle and high schools, department heads sometimes play a significant role in the observation of teaching in each department.

The designated administrator observes the teacher and uses a form containing guidance for ratings on specific domains of teaching (e.g., classroom management). After the observation, the administrator typically communicates with the teacher about the ratings (Porter, Youngs, and Odden, 2001). Administrators may undertake this process two or more times per year for each teacher. Typically a teacher is observed no more than three times in one year (Danielson and McGreal, 2000; Porter, Youngs, and Odden, 2001).

If an administrator concludes that a teacher's performance falls short of the standards, the administrator must respond with a specific set of steps. These steps vary by district, but generally the process involves conducting additional observations, creating an improvement plan, and participating in training or other assistance. Depending on the teacher's performance changes, the building administrator may recommend termination. Administrators in the district office review this recommendation and, if they concur, move to terminate the teacher's employment.

The system for evaluating a new teacher's teaching is usually similar to that used for a veteran teacher. Some states and school systems mandate that new teachers receive a shorter contract (e.g., 1 year)

---

<sup>3</sup> Nationally, 80 percent of central city principals say that they have "a great deal of influence" over the evaluation of teachers (Schools and Staffing Survey, 1999-2000).

and that they be observed more frequently. School districts can then choose not to renew contracts of underperforming teachers, which involves less administrative burden than termination. Also, many states issue new teachers licenses that expire. These teachers often must be recommended for licensure by their employer to attain permanent licensure. This arrangement gives school districts an alternative process for removing an employee who is not performing well.

### **The challenge of accurately identifying ineffective teachers**

One challenge in the traditional approach to evaluating classroom teaching is the limited time available to observe the classroom. Principals spend significant amounts of their time on supervision and evaluation of teachers. In fact, in the most recent national survey data, 86 percent of central city principals reported that they spent time on faculty supervision and evaluation once per week or more (Schools and Staffing Survey, 1999-2000). Because each principal must supervise so many faculty individually, formal observations of whole class periods tend to occur no more than two or three times per year, as noted earlier.<sup>4</sup> Additional observations, which may be necessary in order to decide whether a marginally skilled teacher is making the grade, present a significant time burden for school administrators, so often the number of observations made is insufficient to accurately identify ineffective teachers (Consortium for Policy Research in Education, 1998).

An additional impediment to obtaining a representative view of each teacher's normal teaching patterns is that teachers may 'game' the system by altering their teaching during the observation, the time of which is often announced in advance. Lacking confidence in the evaluation process, the evaluation criteria, or the administrator's expertise, teachers sometimes put in more effort than usual during the evaluation periods. Teachers may also alter their regular teaching style to fit the evaluation criteria, which are typically shared

---

<sup>4</sup> If the process of observing the teacher once, holding a conference with the teacher, and creating the appropriate documentation requires just two hours, an administrator trying to conduct a round of observations with 20 teachers needs an entire 40 hour week.

with the teacher in advance (Consortium for Policy Research in Education, 1998).

Finally, while administrators' background and training allows them to make a legitimate decision about performance, often administrators have not been in classroom teaching positions for many years. At the secondary level, administrators usually have classroom teaching experience in a subject that differs from that of the teacher being evaluated. Researchers who develop systems for analyzing teacher practice have increasingly argued that judging good practice requires taking account of the subject one is teaching, the background knowledge of the students, and other factors (Danielson and McGreal, 2000; Weiss and Weiss, 1998).

### **The challenge of promoting teacher development and/or improved teaching practice**

Another challenge for school systems using traditional evaluation is promoting teacher development and improvement in practice—which is a stated goal for many teacher evaluation systems. Principals are sometimes able to engage teachers in discussions about their instructional practice. In fact, nationally, 45 percent of central city teachers either 'somewhat agree' or 'strongly agree' that their principals talk with them frequently about their instructional practices (Schools and Staffing Survey, 1999-2000). But the interactions that take place as part of traditional evaluation practices are not often credited with promoting teacher development and improvement in practice. Rather, both administrators and teachers perceive evaluation as an adversarial process (Porter, Youngs, and Odden, 2001).

One related challenge is that evaluation instruments are often designed to determine whether teaching reflects a minimal standard rather than to 'norm' the teacher's performance on a particular domain (Weiss and Weiss, 1998). It is very common for principals to 'pass' most teachers on all domains (Danielson and McGreal, 2000; Porter, Youngs, and Odden, 2001). In fact, nationally, central city principals report that an average of 79 percent of their faculty is presently teaching to "high academic standards" (Schools and Staffing Survey, 1999-2000). There may be ways to conduct

traditional evaluation and promote discussion of instructional issues, but these practices are likely rare (Danielson and McGreal, 2000).

### **The challenge of facilitating the termination process**

A final challenge for traditional evaluation systems has been the termination process that must occur after a teacher has been identified as ineffective and has not improved. Nationally, in a survey of practices in the 1999-2000 school year, 76 percent of central city school districts reported dismissing at least some teachers using termination, as opposed to either nonrenewing or 'counseling out' teachers. The remaining districts—almost a quarter of all central city districts—did not report using termination for any dismissals (Schools and Staffing Survey, 1999-2000).

One key link in the termination process has been the role of the teachers' associations, which typically advocate for teachers in order to protect against arbitrary terminations. Central city principals reporting on the termination process at their school have cited the role of teacher associations among other factors in Exhibit 3 as barriers to the dismissal of poor or incompetent teachers. In general the administrative burden of the termination process appears to sometimes prevent the process from being completed (Weiss and Weiss, 1998).

Exhibit 3: Percentage of central city principals indicating that various considerations are barriers to the dismissal of poor or incompetent teachers at their school.

Barrier to Dismissal	Percentage
Personnel policies	60.6
Termination decisions not upheld by third party adjudicators	41.3
Inadequate teacher assessment documentation	36.2
Tenure	68.4
Teacher associations and organizations	71.0
Dismissal is too stressful and uncomfortable for those involved	42.3

Source: Schools and Staffing Survey, 1999-2000.

## PAR and the Toledo Plan

The challenges associated with traditional teacher evaluation led communities to consider PAR programs. Toledo was the first of several communities that shifted responsibility for teacher evaluation to a partnership between the school system and the teachers' association. At the school level, the responsibilities that had fallen on building administrators shifted to consulting teachers (CTs), who were expected to have specialized expertise and more time for evaluation—potentially making them more effective at identifying ineffective teachers, assisting them, and facilitating their removal.

This section on traditional teacher evaluation thus provides a context in which to examine the Toledo plan. Like a traditional evaluation system, the Toledo Plan has specific elements that determine its effectiveness and efficiency. In the remaining sections, we focus on specific elements of the Toledo Plan, beginning with the Toledo Plan standards.

## The Toledo Plan Standards

A key element of every teacher assessment system is its standards—its set of criteria for distinguishing good teaching from bad teaching. Under the Toledo Plan, these criteria appear in the handbook given to teachers participating in the Toledo Plan and to the consulting teachers (CTs), who are responsible for providing evaluation and assistance. The criteria also appear on the form used by CTs to summarize teacher performance (see Appendix A). The criteria fall into four categories:

1. Teaching procedures
2. Classroom management
3. Knowledge of subject – academic preparation
4. Personal characteristics and professional responsibility

These categories include 23 criteria in total, and CTs must rate performance on each criterion as outstanding, satisfactory, or unsatisfactory.<sup>5</sup> CTs must also rate overall performance for each of the four categories.

The structure of the Toledo Plan standards—categories containing specific criteria—mirrors the structure of most standards, except that some standards include overarching statements of principle that are meant to guide the use of the specific criteria (Ingvarson and Kleinhenz, 2003). For example, one of the five overarching statements in the National Board for Professional Teaching Standards (NBPTS) is “Teachers are committed to students and their learning.”<sup>6</sup>

To discuss the effectiveness and efficiency of the Toledo Plan’s standards, it is first necessary to identify what makes standards useful. Standards can play a key role in (a) supporting the accurate identification of underperforming teachers and (b) facilitating the process of helping teachers improve their teaching. To support these functions, standards should focus on aspects of teacher performance that are truly important, making the evaluation system as a whole

---

<sup>5</sup> The rating outstanding is not considered a valid option on several of the criteria (e.g., Personal appearance).

<sup>6</sup> For an explanation of the role of such statements and statements at other ‘levels’ within a set of standards, see Appendix B.

more valid. Standards should also provide observers with the information needed to ensure systematic ratings and observations, making the evaluation system more reliable. Our investigation identified three aspects of the Toledo Plan's standards that pertain to its validity and reliability: (1) the specificity of the standards, (2) how the standards distinguish different levels of performance, and finally (3) the premises behind the standards.

### **Levels of specificity**

Experts on Peer Assistance and Review (PAR)<sup>7</sup> programs recommend that standards be specific and clear (Koppich, 2000; Kelly, 2000). Standards must be specific enough to allow a CT who observed a teacher to give concrete feedback, and thus to help the teacher improve his or her performance (Danielson and McGreal, 2000). Well specified standards also ensure accurate identification of underperforming teachers, because the standards force evaluators to focus consistently on the same aspects of teaching (Anderson and Pellicer, 2001). Without sufficient specificity, we cannot be assured that two CTs observing the same teacher's performance would give the same ratings.

Consider what would happen if a CT was required to develop only four ratings corresponding to the four categories that were identified earlier (i.e., "Teaching Procedures," "Classroom Management," "Knowledge of Subject," and "Personal Characteristics and Professional Responsibility"). Two CTs observing the same teacher's performance could easily reach different ratings for the category "Teaching Procedures," because they would not be required to focus on the same specific aspects of "Teaching Procedures." One observer might arrive at a rating on the basis of planning procedures while another might arrive at a rating on the basis of assessment and evaluation procedures. As a result, the two observers might reach different conclusions.

---

<sup>7</sup> An alternative label used to describe these programs is Peer Assistance and Peer Review (PAPR) programs, but the Toledo Plan documents and most writers refer to them as Peer Assistance and Review (PAR) programs.

Likewise, the two CTs might focus on different aspects of “Teaching Procedures” when giving assistance. Standards should be well specified so that a CT will be likely to help the teacher in every area needed (Kelly, 2000).

The Toledo Plan’s standards specify 23 criteria. In addition, for each criterion, the *Intern, Intervention, Evaluation* booklet provides three pieces of information. For each criterion, these pieces are:

- (1) a ‘definition’ listing a series of specific indicators of the criterion;
- (2) a series of examples of ‘behavioral performance goals’ that CTs might give to teachers;
- (3) a series of examples of ‘supportive activities’ to be undertaken by others (e.g., the CT) to help the teacher achieve those specific goals.

Exhibit 4 illustrates this information for one criterion under Teaching Procedures called, “Skill in making assignments.”

Exhibit 4: Specific guidance given regarding the criterion “Skill in making assignments”

Definitions Given	Examples Given of Behavioral Performance Goals	Examples Given of Supportive Activities by Other Personnel
<ul style="list-style-type: none"> <li>• Pupils are helped to relate new subject matter to previous learnings.</li> <li>• Presentation follows planned steps for most effective learning and shows adjustment to needs of the group and individuals.</li> <li>• Worthwhile and interesting assignments are presented clearly, and explicit directions are given so that pupils know what is expected of them.</li> </ul>	<ul style="list-style-type: none"> <li>• Assigns lessons suitable in lengths and difficulty to the students’ age, ability and background.</li> <li>• Assigns lessons that are clear and definite to the student.</li> <li>• Assigns lessons that are planned for a sequential pattern of learning.</li> <li>• Selects assignments that are purposeful, worthwhile, and related to the subject matter and skills being taught.</li> <li>• Takes into</li> </ul>	<ul style="list-style-type: none"> <li>• Provide examples of lesson plans which demonstrate meaningful assignments to all the students.</li> <li>• Make supportive suggestions regarding effective out-of-class assignments.</li> <li>• Arrange a meeting between the teacher and another teacher.</li> </ul>

<ul style="list-style-type: none"> <li>• Classwork and homework assignments are made realistic in length and level of difficulty according to the grade, ability, and home background of the student.</li> </ul>	<p>consideration the student and his family when assignments are made.</p>	
--	--	--

Source: *Intern, Intervention, Evaluation*.

Although the Toledo Plan’s standards provide a reasonable amount of information about what is expected, some other districts are using even more detailed standards. Poway Unified School District, in California, probably provides the most detail, with its system of 73 criteria.<sup>8</sup>

### **Distinguishing levels of performance**

The Toledo Plan’s standards provide a general definition of the three possible performance levels. The following definitions apply to all of the evaluation criteria:

- Satisfactory – should be the point of departure. A check there should indicate that the teacher is showing the degree of professional qualities and growth to be expected and desired as a beginning teacher.
- Unsatisfactory – should signify that these weaknesses or deficiencies are of such a serious nature as to indicate that the teacher will not be recommended for reappointment unless substantial improvement is shown.
- Outstanding – should suggest that instances have been observed where unusual skills, abilities, and attitudes are in evidence and where the teacher’s accomplishments are in excess of the basic requirements for a satisfactory rating.

---

<sup>8</sup> Poway describes their standards as including 18 criteria and 73 ‘indicators’, but the indicators are functionally equivalent to criteria. See Anderson and Pellicer, 2001, p. 66.

The threshold for being in danger of termination or nonrenewal under the Toledo Plan is that a teacher is rated ‘unsatisfactory’ on any one of the 23 evaluation criteria. In other words, teachers must be at least ‘satisfactory’ on all measures or face the possibility of termination. Thus, it is very important that the Toledo Plan standards provide a strong basis for distinguishing levels of performance.

We caution that without further guidance, these definitions alone are not likely sufficient to ensure that CTs select levels for each criterion based on the same rules. Elsewhere in this report we discuss the training CTs receive and the questioning of CTs that takes place during meetings of the governing body that oversees the Toledo Plan—the Intern Board of Review (IBOR)—both of which have facilitated shared understanding of what these definitions mean in practice.

One way to provide more detailed guidance is to establish, for each individual performance criterion, a set of specific definitions that define each level of performance. Such a set of definitions is called a scoring rubric, because it provides a system by which an observer can rate or ‘score’ a teacher. Exhibit 5 shows an example of a scoring rubric from an evaluation system that distinguishes between performances that are ‘unsatisfactory,’ ‘basic,’ ‘proficient,’ and ‘distinguished.’ The text under each performance label pertains to only one evaluation criterion, which in this case is “Teacher interaction with students.” Poway Unified School District and Mount Diablo Unified School District—two California districts with PAR programs—use scoring rubrics to distinguish the levels of performance observed by CTs. Scoring rubrics are also used in a variety of national teacher assessment systems (see Porter, Youngs, and Odden, 2000).

Exhibit 5: Example of defining levels of performance for an individual evaluation criterion

Element	Unsatisfactory	Basic	Proficient	Distinguished
Teacher Interaction with Students	Teacher interaction with at least some students is negative, demeaning, or sarcastic, or inappropriate to the age or culture of the students. Students exhibit disrespect for teacher.	Teacher-student interactions are generally appropriate but may reflect occasional inconsistencies, favoritism, or disregard for students' cultures. Students exhibit only minimal respect for teacher.	Teacher-student interactions are friendly and demonstrate general warmth, caring, and respect. Such interactions are appropriate to developmental and cultural norms. Students exhibit respect for teacher.	Teacher demonstrates genuine caring and respect for individual students. Students exhibit respect for teacher as an individual, beyond that for the role.

Source: Danielson and McGreal, 2000.

While there are no scoring rubrics in the Toledo plan other than the general definitions supplied earlier, the practice of formulating and issuing 'performance goals' partially fulfills the role that rubrics can play. Rubrics define exactly what behavior is expected, and the standards give CT samples of performance goals for each criterion. The difference between the two is that a performance goal does not distinguish—in the degree of detail that a rubric does—between unsatisfactory and satisfactory levels of performance.

### Premises about good performance

To distinguish satisfactory and unsatisfactory performances, standards must be based on certain premises about what constitutes good teaching. Each criterion within the standards contains such a premise. For example, the standard "Skill in assessment and evaluation" (see Appendix A) contains the premise that good teachers are able to assess student learning, in order to monitor progress and to focus and tailor instruction.

Our review of the premises in the Toledo Plan's standards indicated that the standards all have face validity. That is, they all make sense in light of existing research and normal practices. Continuing to use

“Skill in assessment and evaluation” as an example, there is literature on the importance of ongoing assessment of student learning in order to target instruction to the students’ level of understanding. In fact, assessment of student learning is also a topic in many teacher preparation programs. Thus, it is reasonable to include this criterion.

However, researchers have not been able to directly test individual premises regarding teaching practices and show effects on student achievement, except in a few cases. These research findings identify very basic characteristics of teaching and teachers. For example, it has been shown that students in mathematics classes learn more from teachers who majored in subjects related to mathematics (Goldhaber and Brewer, 2000). The body of research findings that show effects on student achievement do not call into question any of the premises of the Toledo Plan’s standards (see Wayne and Youngs, 2003).

Another way to examine the premises of the Toledo Plan’s standards is to compare them to other PAR programs’ standards. A recent review of several existing and planned PAR programs examined their standards and found that the criteria were very similar. All standards focused on premises about good teaching and addressed a common set of topics. Anderson and Pellicer (2001, p.31) list the following topics:

- Classroom management
- Planning instructions
- Teaching skills
- Assessment and evaluation skills
- Knowledge of subject matter
- Professional qualities
- Relations with students
- 

Poway’s standards illustrate the commonality among standards. The Table in Appendix C provides a close comparison of the criteria for the Toledo Plan and the PAR program in Poway Unified School District—not to evaluate the Toledo Plan criteria but just to show the degree of alignment (see Appendix C). The process many districts use to develop standards to guide PAR programs or traditional evaluation systems involves a group of teachers, administrators, and

other stakeholders in reviewing and adapting existing standards or standards from other school systems.

## **Gathering and Using Data on Teacher Performance**

Decisions about which teachers to retain depend on good processes for gathering data as well as good processes for interpreting and using those data. The primary source of data on teacher performance under the Toledo Plan is a series of classroom observations, which are conducted by the CT who has been assigned to the teacher.

### **Gathering data**

Intern and intervention teachers are observed by their CT several times throughout the school year. In Toledo the first visit must be preceded by a conference and held at a mutually agreeable time. Later visits must be unannounced, which is important since otherwise teachers may plan special lessons during visit times (Consortium for Policy Research in Education, 1998).

The Toledo Plan guidelines require that observations last at least 20 minutes, but in reality the observations typically last more than twice that long. The total number of observations completed depends on the teacher's performance. For example, an intern who is doing well would be observed at least 5 times in the first semester (before December 20) and at least 2 times in the second semester (before March 30). A teacher who is doing poorly would be observed at least six or seven times in the first semester alone.

We believe that the observation practices used in Toledo allow the CT to gather sufficient data to inform judgments about performance. Research indicates that aspects of instruction vary depending on the time of visitation (Ball, Camburn, Correnti, Phelps, & Wallace, 1999), so traditional evaluation practices which involve only two or three observations over the course of a year are probably not effective at accurately identifying underperforming teachers (Odden, Porter, and Youngs, 2001; Danielson and McGreal, 2003). We note that

Cincinnati uses the PAR system of teacher review and requires six observations over the course of one year (Porter, Youngs, and Odden, 2001). There are PAR programs that require more observations, such as Columbus which requires at least 20 observations in one year (Stedman and Stroot, 1998).

Additional sources of data beyond the observations of the CT can be valuable (Anderson and Pellicer, 2001). For instance, Rochester collects 'portfolios' from teachers, which are self-reflective narratives written by the teacher regarding their own practices (Kerchner, 2001). In Toledo, as in many other PAR programs, principals contribute data by filling out a form about the teacher noting cooperation with others, absences, and whether or not the teacher satisfactorily complies with district and building policy (see Appendix D).

## **Recording data**

The methods used to record observations are also important to understand. In Toledo, CTs fill out an observation report form that provides (a) basic information such as the date and time of the observation and the number of students present, (b) a narrative summarizing what was observed, and (c) comments, areas of concern, and suggestions. CTs provide greater detail on unsatisfactory areas than on satisfactory areas. No later than 5 days after the observation and usually sooner, the CT must hold a conference with the teacher and discuss the observation report.

One strength of the Toledo observation process is that information is always shared with the teacher being evaluated. Before any data gathering begins, the teacher receives a copy of the standards. The teacher receives each observation report and must sign to acknowledge that he or she has seen it. The observation report includes a detailed narrative, comments, concerns, and justifications of the ratings—which is much better than just a series of ratings without significant explanation. Finally, all observation reports must be filed with the Intern Board of Review (IBOR). These procedures all contribute to the fairness and effectiveness of the program (Anderson and Pellicer, 2001). There are also clear rules governing the interactions between the CT and the teacher, which is considered

good practice (American Federation of Teachers and National Education Association, 1998).

### **Using data to make decisions**

We examined closely the process by which observation data can lead a CT to recommend termination or nonrenewal. That process begins if the teacher has been notified of an unsatisfactory area but does not show improvement—or a concerted attempt at improvement—at the time of the next visit. The CT includes a *performance goal* in the next observation report, along with text warning that the goal must be met.

The CT determines the exact wording of the performance goal. The Consultant Handbook guidelines for writing performance goals appear in Exhibit 6. Note that the CT is expected to use the standards as a guide, but the performance goal is more specific, focused, and tailored to the teacher's needs. The guidelines also indicate that the CT should consult with other CTs. As was shown earlier in Exhibit 4, the *Intern, Intervention, Evaluation* handbook provides sample performance goals for each criterion in the standards, which also improves the consistency of the CT's performance goals.

#### **Exhibit 6: Guidelines for Performance Goals**

1. Performance goals need not be written unless no attempt is made by the intern to respond to suggestions and recommendation in the written report and conference, or the intern is having severe problems.
2. Use the handbook, other consultants [CTs], and previous reports documenting problem areas as a resource.
3. Concentrate on one or two major areas of concern and cite specifics from previous reports.
4. Write specific goals and ideas for achieving the goals.
5. Offset the performance goals in capital letters or bold print on the written report.
6. The intern should sign and date each performance goal. A specific timeline for meeting each performance goal should be established.
7. Once performance goals are set, the status of the goals must

be documented in all subsequent reports until a satisfactory level of performance has been achieved and the performance goals discontinued.

Source: Toledo Public Schools, 1999.

The practice of setting goals in response to continued underperformance is in line with other PAR programs. For example, Poway refrains from giving performance goals until a teacher's second year. New York City's Peer Intervention Program differs somewhat in that the teachers set their own goals, irrespective of a set of standards, although that program is voluntary and cannot result in termination (Anderson and Pellicer, 2001).

Once a performance goal is set, CTs typically do not lift the goal until three satisfactory performances are observed. Typically a teacher with a performance goal wants to be observed more frequently because each observation is a chance to demonstrate growth before the next Intern Board of Review (IBOR) meeting (December 20 and March 30). CTs are required to give teachers 5 days before checking to see if they have successfully implemented a recommended change or met a performance goal.

At the IBOR meeting, CTs provide board members with a summary report for each teacher (see Appendix A), and IBOR reviews these reports. IBOR questions each CT about each teacher assigned him or her. This process is an important check on whether the CT has acted appropriately and reached the correct conclusions, as well as a form of supervision of the CT, which will be discussed later.

**Mentoring/Assistance**

The Toledo PAR Plan provides significant assistance to interns and intervention teachers. One of the benefits of the program is that all interns receive this assistance, regardless of whether they are having difficulty or not. In practice CTs direct the most assistance to teachers who are having the most difficulty.

The assistance received by interns includes training sessions that are tailored to the needs of new teachers, which experts argue can make

new teachers more effective (Shields et al, 2004). This training begins with one week of workshops in the summer. Interns and intervention teachers may also attend a variety of inservices provided by the CTs during the school year. These have included sessions for sharing ideas as well as sessions on practical topics of interest to new teachers, such as classroom management and handling parent conferences and computerized gradebooks. Other PAR programs provide similar training for new teachers, although each program has some unique components. For example, new teachers in Cincinnati participate in monthly, subject-specific training sessions called 'practica' (Porter, Youngs, and Odden, 2001).

The main form of assistance—provided to both intern teachers and intervention teachers—is the mentoring provided by the CTs. Most of the interaction between the CTs and the teacher occurs in the conferences that teachers have with CTs after each session. On average CTs spend about 15 hours per semester in such conferences with the teacher and also spend some time in other conversations with the teacher. The amount of time CTs are able to dedicate to each individual teacher makes them effective assistance providers relative to administrators, who have less time available due to their other responsibilities (Kelly, 2000).

There is not a pool of scientifically rigorous studies showing that mentoring is effective at helping new teachers learn (Lopez, Lash, Schaffner, Shields, & Wagner, 2004), but reviews of less rigorous studies find suggestive evidence that mentoring improves teacher retention and teaching practices (Ingersoll and Kralik, 2004; Strong, 2004).

In their discussion of best practices for PAR programs, Anderson and Pellicer (2001) recommend that CTs work with teachers on particular issues but also act as 'brokers' to connect the teachers to key resources that will help them with particular challenges. CTs in Toledo do this, for instance by referring teachers to particular instructional videos, and helping teachers organize a professional day to observe a teacher with a particular strength.

The fact that CTs are matched to teachers in their subject or approximate grade level makes them particularly effective at giving

assistance and identifying key resources. A recent study which examined the patterns of turnover among first-year teachers who received varying types of support found that one of the important supports was having a mentor in the same field (Smith and Ingersoll, 2004).

During post observation conferences, CTs also provide assistance in the form of ongoing feedback on performance. Some feedback is positive, to reinforce good practices, and some feedback addresses improvements that are needed. Some experts worry that the mentor-mentee relationship is hampered when the mentor has responsibility for evaluation. The advantage of this arrangement may be that teachers take the feedback more seriously. However, teachers may behave defensively and not be open to learning from the mentor (Shields et al., 2004; Haselkorn and Fideler, 1999). Indeed, a teacher who has been given a performance goal understands that his or her job is at stake and is likely to be guarded if not adversarial when interacting with the CTs.

## **Consulting Teacher Selection, Training, and Supervision**

The success of the Toledo Plan relies heavily on the skills and capabilities of the Consulting Teachers (CTs). CTs' observations are the primary basis for accurately identifying teachers who are struggling. The Toledo Plan also depends on CTs to help teachers develop and improve their skills. In addition to observing and assisting teachers, CTs must also produce extensive documentation of their interactions and present the data to the IBOR.

There is literature on the skills and capabilities needed by those who mentor other teachers, but it does not include any scientifically rigorous research (Lopez et al., 2004). Experts are in agreement that good mentoring requires high quality mentors—people who not only are good teachers but also have special qualities that make them good mentors (Alliance for Excellent Education, 2004; Shields et al., 2004).

Because so much depends on CT's skills and capabilities, we examined closely how CTs are selected, trained, and supervised.

The first discussions below addresses how Toledo selects CTs and what qualities are sought, in light of best practice research and practices in other PAR programs. The second discussion focuses on how Toledo trains CTs after they are selected. A third discussion addresses the supervision of CTs. CT selection, training, and supervision are the primary mechanisms available to Toledo for controlling the quality of CTs.

## **Selecting Consulting Teachers**

According to a handbook published jointly by the American Federation of Teachers and the National Education Association (1998, p. 19), the criteria typically used by PAR programs for CT selection include the following:

- have taught successfully for a specified number of years;
- be recognized as an outstanding classroom teacher;
- demonstrate deep knowledge of the disciplines(s) they teach;
- possess a repertoire of effective classroom management strategies and instructional techniques;
- have strong verbal skills, both orally and in writing; and
- have the ability to work cooperatively and effectively with others.

In Toledo, the hiring of CTs is overseen by a committee selected jointly by the district and the union. The guidelines specify the following minimum criteria for all applicants:

- (a) five years of outstanding teaching service;
- (b) favorable references from three fellow teachers who are in the same building; and
- (c) a favorable reference from the building principal.

Applicants who meet these criteria will not necessarily be selected. For instance, although the minimum number of years of experience is five years, a typical CT has at least 12 years of experience. The committee has no other specific guidelines for hiring decisions, but the following are important according to a publication of the Toledo Public Schools and Toledo Federation of Teachers:

- Has ability to work well with adults
- Has ability to communicate well

- Is respected as a teacher and a leader
- Is committed to a teaching career
- Is comfortable with decision to terminate or non-renew a contract (Lawrence, 2003)

These criteria all appear relevant to the responsibilities of the consulting teacher. They also conform to what has been described as 'best practice' with the exception that the Toledo criteria for CT selection do not include subject matter knowledge (Anderson and Pellicer, 2001). Toledo does try to assign CTs to teachers in the same assignment area (e.g., secondary mathematics). Experts agree that it is important for mentors to have the same assignment area as those they mentor (Alliance for Excellent Education, 2004; Kelly, 2000; Shields et al., 2004).

New York City's Peer Intervention Program lists some additional criteria. First, the criteria include not only interpersonal skills but also "demonstrated knowledge of, and successful experience with, adult learners." Adults learn in a manner that is different from K-12 students, which makes this criterion particularly relevant. Second, the program requires "Exemplary knowledge and evidence of creativity and initiative with respect to curriculum content, materials and methods" (AFT & NEA, 1998, pp. 11-12). Clearly a CT should have this capability in order to provide effective assistance.

We also examined Toledo's procedure for judging applicants according to these selection criteria. The process begins with a paper application that includes 4 references—one from the principal and 3 from other teachers. The committee scrutinizes these applications. Applicants must describe their reasons for applying to be CTs. Applicants who pass the initial paper screening are formally interviewed, and a writing sample is taken on-the-spot at the time of the interview. Applicants then submit to two unannounced observations of their teaching.

We believe the procedure described above provides valid data on whether the applicant has good oral and written communication skills, whether the applicant has a strong reputation as a teacher and a leader, and whether the applicant is a good teacher. This procedure

is similar to the selection process used in other PAR programs (NEA & AFT, 1998; Anderson and Pellicer, 2001).

## **Training Consulting Teachers**

Experts on PAR programs underscore the need to provide CTs with thorough training (Costa and Garmston, 2000; Kelly, 2000; Koppich, 2000). Although the hiring process selects persons with certain qualities, many skills still have to be taught. The training topics highlighted by experts are (1) how to observe and evaluate teachers and (2) how to mentor and provide assistance to teachers (Anderson and Pellicer, 2001; Koppich, 2000; Shields et al., 2004).

There are several training mechanisms for CTs in Toledo. Before he or she stops teaching, a future CT attends IBOR several times (3-4), observing the procedure and process and familiarizing themselves with the documents. Another training mechanism is one day of “shadowing,” in which the prospective CT is paired with an experienced CT to observe how observations and teacher conferences are conducted.

The primary formal training for CTs lasts two days during the summer. Our review of the syllabus for the two day training suggests that at least 3 hours are dedicated to observation and evaluation. At least one hour is spent on conferences. The remainder of the time is spent introducing the CTs to the program and its procedures. During those two days, CTs are sometimes broken out into groups according to assignment (e.g., mathematics, elementary, etc.).

CTs also receive ongoing training, much of which is informal. CTs frequently discuss their work in the office, which is a single room shared by all CTs. All of the CTs also participate in a regularly scheduled monthly meeting. In addition, new CTs are paired with veteran CTs. Through all of these mechanisms CTs continually share knowledge and learn from each other. In terms of formal ongoing training opportunities, CTs can take a 3 credit graduate course on current teaching and classroom management methodologies. CTs must also repeat the two-day seminar every year, although sometimes they participate as instructors.

Other PAR programs provide similar training initially. Some provide ongoing training that is more extensive, most notably Columbus, which provides a weekly seminar for CTs which often serves as an opportunity for discussing specific cases as a group (Stroot et al., 1998). A handbook on PAR published by American Federation of Teachers and the National Education Association (1998) indicates that training “. . .not only orients them fully to the program, its policies and guidelines but also addresses and enhances skills regarding working with adults, documenting observations, deepening subject matter knowledge, and assuring up-to-date, research-based information on instruction and learning” (p. 20).

### **Supervising Consulting Teachers**

A final program element affecting the quality of CTs is supervision. IBOR meetings provide a context in which the board members can examine the performance of the CT, to some extent. When CTs present information about the teachers to whom they have been assigned, IBOR members pose questions that prod the CT to take particular actions (Kerchner, Koppich, & Weeres, 1998). They can also review the observation reports that have been filed by the CT at any time if it is concerned about the work of a CT. IBOR is responsible for supervising CTs and has on occasion terminated consulting teachers.

Another mechanism for learning about CT performance is a written evaluation of the program, filled out by all interns and submitted anonymously. Evaluations are usually negative when an intern has not succeeded in the program. In practice, these evaluations have not led to the identification of CTs who should be terminated. CTs do receive feedback through the evaluations, as the evaluations are compiled and given to the CTs in summary form.

Finally, veteran CTs are paired with newer CTs to provide a source of informal mentoring for newer CTs. These relationships also may be a mechanism by which IBOR may learn about CT performance.

The fact that much of the work of CTs cannot be directly observed by IBOR makes it hard to increase monitoring of CT performance. In Rochester, CTs undergo an evaluation cycle just like teachers do, and principals contribute to CT's evaluations. Principals in Rochester continue to conduct observations during the program, so the idea of using input from principals may make more sense in Rochester than in Toledo (Institute for Education Reform, 2000).

## **Task 2: Compare the Toledo Plan’s guidelines and procedures with similar efforts to judge the performance of probationary and veteran teachers.**

The University of North Carolina has posted a series of articles on its web site about peer review in teaching.<sup>9</sup> While the material appears to address peer review in higher education, it is relevant to Toledo:

- Peer review for “summative” purposes and peer review for “formative” purposes are carefully distinguished. “They are for different purposes and should be conducted independent of each other.”
- Summative evaluation describes activities that are conducted to gain information needed to make decisions such<sup>10</sup> as promotion, reappointment or tenure.
- Formative evaluation describes activities that are to provide teachers with information they can use to improve their teaching. The information is intended for their personal use, rather than public inspection.
- Formative peer observation is highly recommended for non-tenured faculty and for all faculty prior to a summative observation of teaching.

### **Best Practices:**

- It is critical that good peer review processes include the following characteristics:
  - Openness
  - Mutually agreed upon criteria
  - Adherence to developed procedures
  - Written feedback at all stages
  - Discussion of the results
  - Methods for monitoring and revising the process
- Both formative and summative review must be accomplished. Formative review that provides information used by the teacher

---

<sup>9</sup> The North Carolina State University web site is at [http://www.ncsu.edu/procost/peer\\_review](http://www.ncsu.edu/procost/peer_review)

<sup>10</sup>

for improvement and professional development of teaching begins during the first year of employment. Summative review focuses on information needed for personnel decisions and accountability begins when a personnel decision is pending.

- Teaching mentors should be assigned to new faculty during the first year of their contract. New teachers should be given support for developing their teaching skills during their first year from experienced, successful teachers.
- Peer reviewers can be mentors, developers, and judge depending on the type of review. Peer reviewers need to study the literature related to in their roles as formative *or* summative reviewers.
- In summative evaluations, it is typical for an ad hoc committee consisting of *several* faculty members to perform the observation.
- Observers must respect the observed instructor. A faculty with a strong difference of opinion or dislike for a peer may have difficulty being a fair observer.

The Education Policy Studies Division of the National Governors' Association for Best Practices has issued a paper entitled *Exploring Teacher Peer Review*.<sup>11</sup> This paper addresses peer review in K-12 public schools. Relevant points are the following:

- The AFT and NEA have jointly declared that peer review without peer assistance for teachers in the program does not represent sound educational policy.
- Peer review programs do not have the same general support among districts and unions as peer assistance programs.<sup>12</sup>

---

<sup>11</sup> 1/12/2000. Authors are Phillip Escamilla, Theresa Clarke, Dane Linn. The contact person listed is Phillip Escamilla, (202) 624-3623.

<sup>12</sup> In 1997 delegates to the NEA Representative Assembly voted to drop their long-standing opposition to peer review. The AFT has long supported peer review.

- Some critics of peer review systems have argued that consulting teachers are not held accountable to any authority.<sup>13</sup>
- The National Commission on Teaching and America's Future claims that more teachers needing instructional assistance have been helped and more teachers needing career redirection have been dismissed under peer review than under traditional methods of evaluation.<sup>14</sup>

Appendix E compares certain features of peer assistance and review programs in five districts: Toledo, Cincinnati, Minneapolis, Rochester and Columbus. The information comes from documents we have obtained, and interviews. The chart speaks for itself, but here are observations that I think are most interesting:

### **Probationary Teachers' Payment of Union Dues and Rights of Union Representation**

Toledo's 50% dues assessment against probationary teachers who are members of the union but who do not have full rights of representation, is not extreme. Of the other four districts reviewed, all charge 100% dues to probationary teachers. However, two of the districts do provide more representation to probationary teachers than does Toledo. One district, Columbus, provides "complete access to union representation during peer review." Nevertheless, even in Columbus the fact remains that union members are evaluating (and

---

<sup>13</sup> The report cites Myron Lieberman, *Teachers Evaluating Teachers: Peer Review and the New Unionism* (Bowling Green, Ohio: Social philosophy and Policy Center and Transaction Publishers, 1998): 49.

<sup>14</sup> ERIC Digest 126 (May 1999). According to the U.S. Department of education, almost twice as many teachers were dismissed under peer review as under administrator evaluations in a Cincinnati program.

sometimes recommending termination of) union members who are probationary employees.

### **Mixing Summative and Formative Evaluations/ Ability to Change CTs**

Every district mixes summative and formative evaluation in the peer review process. This is interesting, given that common sense and the best practice recommendations for peer review that we have seen indicate that combining the two is problematic, especially for the teacher under review.<sup>15</sup> However, a significant mitigating factor in the districts that allow it, is the ability of a teacher under review to change CTs or mentors when things go wrong. In this important respect Toledo is in the minority. Of the five districts reviewed, Toledo is the only one where teachers undergoing evaluation have no say in who their CT is, and no ability to at least request a change.

### **Make Up of Peer Review Panel**

Toledo and Columbus are the two districts of the five reviewed that have a union majority on the panel or board that recommends renewal or non-renewal of probationary teachers. Cincinnati and Rochester have an even distribution between union and administration on the board, and Minneapolis has an administration majority. Significantly, in Columbus it takes two-thirds vote (of seven members) for the panel to act. Accordingly, it takes 5 of the 7 members to agree on any action, which means that the union majority cannot act on its own

### **Selection of CTs/Mentors**

In Toledo, the TFT selects CTs from an applicant pool, following an application process. In each of the other four districts reviewed, it is the panel that appoints CTs or mentors, following an application process.

---

<sup>15</sup> See Sutherland to Kaboolian memo of 2/28/05 on Best practices in Peer Review of Teachers. It references work done at North Carolina State University, available online at [http://www.ncsu.edu/procost/peer\\_review](http://www.ncsu.edu/procost/peer_review).

## **Role of Principal in Evaluation**

Toledo is the only district of the five reviewed where the principal is unable to participate in the evaluation of a probationary teacher's teaching. In Minneapolis and Rochester the principal's role is significant. In Cincinnati and Rochester the principal's role is somewhat discretionary on the part of the principal, but it can be aggressive should the principal wish to become involved.

## **Conclusions**

Certainly, there are more similarities than differences between the Toledo plan and peer review plans in other districts. Generally speaking, they are the same. A select group of experienced teachers both mentors and evaluates the performance of newly hired teachers and reports to a joint panel on their performance. There are performance standards, and there is an orderly process. The panel has recommendatory authority only. Final decisions are made by the board of education. The union and the administration jointly participate in design and control of the process.

Generally speaking, the Toledo Plan fits the pattern. There is no one thing about it that dramatically distinguishes it from other programs. Still, on several key issues, Toledo comes out on the "control" end of the spectrum. Viewed as a whole, perhaps the Toledo Plan is distinguished. Cumulatively, it is more tightly controlled by the district and the union than the other four programs reviewed.

**Task 3. Compare and contrast both from teacher quality and a cost per non-renewal perspective, the effectiveness and efficiency of the TP to teacher evaluation and non-renewal systems utilized by other large urban districts.**

Peer review programs save money by avoiding costly losses in a number of significant ways, only one of which is the avoidance of the high cost of terminating a tenured teacher.

- The programs weed out weak teachers while they are probationary, avoiding the expense of termination later after they become tenured.
- Peer review programs improve retention, sometimes dramatically. This is important because poor retention is extremely costly to school districts in financial terms alone.<sup>16</sup> These costs are incurred as districts recruit, hire, orient and evaluate new teachers to replace other relatively new teachers who leave, having just been recruited, hired, oriented and evaluated.
- Less relevant but an important point—Poor retention is also costly for states. States typically contribute significantly to the education of teachers in public college and universities. This investment is lost every time a teacher quits the profession.<sup>17</sup>

These cost savings are hard to measure. There is very little data on cost alone. (See Appendix F for comparative districts.) Any analysis would include a comparison of the cost of operating an effective peer review program with the savings realized in the avoidance of litigation and improving retention, but there are other considerations. First, it is important to keep in mind that the financial benefit of effective peer

---

<sup>16</sup> One researcher says that the aspect of peer review that is likely to become important in the future is its ability to retain new teachers longer. The problem of poor retention will become worse as student enrollments are growing and increasing numbers of teachers are reaching retirement. ERIC: ED429343. Peer Review of Teachers. ERIC Digest, Number 126. Cites Bradley, Ann. "Peer Review Programs Catch Hold as Unions, Districts Work Together." Education Week on the Web (June 3, 1998): 1-7. The National Commission on Teaching and America's Future has declared that teacher retention has become a national crisis, and that high rates of turnover are the biggest barrier to teacher quality. *No Dream Denied*. Summary Report. 2003. Page 8.

<sup>17</sup> In Hawaii, as an example, the 1999-2000 cost to the state for every upper level undergraduate student at the state university in the college of education was \$10,670 per year. For graduate students it was \$17,923. Hawaii Educational Policy Center Policy Brief, July 2002.

review programs cannot be measured in terms of net savings alone. Part of the financial benefit is the ability to re-direct expenditures, from the negative processes associated with termination and replacement to the positive processes focused on professional growth and support. Second, peer review is often part of a larger program (professional development/ career in teaching) expense in a district so the costs are not easy to isolate without doing a detailed time and rate analysis for all the professionals that participate in the process. Third, costs not incurred because of less litigation and less teacher replacement can only be guessed at.

For purposes of the present analysis it is possible to estimate that that effective peer review programs do reduce litigation costs associated with terminating tenured teachers, and that they significantly improve retention, thus avoiding large expenditures associated with replacing teachers who quit.

## **National Trends and Anecdotal Information**

### Termination

The National Commission on Teaching and America's Future claims that more incompetent teachers have been dismissed under peer review than under traditional methods of evaluation.<sup>18</sup>

### Retention

A National Center for Education Statistics study found that nationally, in all schools, the attrition rate of new teachers who had not received induction support was 26% within the first three years of teaching. For teachers who had participated in an induction program, the attrition rate was reduced to 15%.<sup>19</sup>

The same study found that, nationally, 19% of teachers leave teaching altogether after the first year.<sup>20</sup> One study estimates that 30% of all teachers leave the profession within five years, and

---

<sup>18</sup> ERIC 126.

<sup>19</sup> Koppich, 57

<sup>20</sup> Henke, Chen, and Geis, 2000.

another concludes that 46% of teachers leave teaching by the end of their fifth year.<sup>21</sup>

It has been estimated that in urban districts without peer review as many as 50% of new hires leave after 5 years.<sup>22</sup>

In 2000, The National Center for Educational Statistics determined that teachers without induction support leave teaching at a percent almost 70% higher than those who receive it.<sup>23</sup>

Similarly, it has been found that receiving feedback on teaching reduced the rate of attrition for first year teachers from 25.7% to 13%.<sup>24</sup>

Typically, large urban districts with the highest percentages of poor and minority students have the highest turnover rates.<sup>25</sup>

## **Cost**

Estimates of the cost of terminating a tenured teacher are in the \$50,000-\$200,000 range, per teacher, when a teacher litigates.<sup>26</sup> It usually takes 3-6 years for litigation over termination to run its course. Some cases go longer.<sup>27</sup>

---

<sup>21</sup> Koppich, 7. Also, National Commission on Teaching and America's Future. *No Dream Denied*. Summary report. 2003. Page 10. Richard M. Ingersoll, source.

<sup>22</sup> Bradley, Ann. Peer Review Programs Catch Hold as Unions, Districts Work Together. *Education Week* June 3, 1998.

<sup>23</sup> National Commission on Teaching and America's Future. *No Dream Denied*. Summary report. 2003. Page 27.

<sup>24</sup> National Commission on Teaching and America's Future. *No Dream Denied*. Summary report. 2003. Page 21.

<sup>25</sup> National Commission on Teaching and America's Future. *No Dream Denied*. Summary report. 2003. Page 14.

<sup>26</sup> Tom Gillett, former VP Rochester Education Association, cites Murray, Grant and Swaminathan, *Rochester's Reforms: The Right Prescription?* in estimating that the average cost of terminating a tenured teacher under New York law has been as high as \$100,000. The cost has declined after 1994 when the law was changed. Adam Urbanski estimated the amount to be \$50,000-\$70,000. Robert Wroth (1998), cited in ERIC 126, placed the number at \$60,000. Richard L. Logan, a labor relations consultant for the Columbus Education Association, estimated the cost to be \$75,000-\$200,000 in Ohio in 1998. *Ed. Week*. June 3, 1998.

<sup>27</sup> Bob Chase estimated five or six years in 1999. EIA (Education Intelligence Agency) Communique Archives, August 16, 1999. The Sturdivant case in Toledo was filed in 2002, and is still pending in two separate actions in two Ohio courts. The case of Ivy Featherstone against the Columbus Public Schools took seven years to resolve. EIA.

The cost of teacher replacement caused by poor retention, even when there is no litigation, is significant. While true costs are difficult to capture, and models for cost calculation vary greatly, such cost should be considered to include expenses associated with departure, the recruitment and hiring of new teachers, and the orientation and training of new teachers.<sup>28</sup> This financial cost is on top of lost productivity, disruption and loss of continuity caused by heavy teacher turnover.

A 2000 Texas study of the cost of teacher turnover shows the cost in Texas to be between \$2,000 and \$5,000 per teacher that resigns, depending on geographic, economic and community characteristics, as well as the ability to pay signing bonuses and stipends for shortage areas.<sup>29</sup> Compared with other sources, this is a very conservative estimate. Some models would place the cost much higher, anywhere between 25% to 100% of the leaving teacher's salary. The US Department of Labor estimates that costs to replace an employee average 33% of the new hire's salary.<sup>30</sup>

A full cost analysis of the Toledo Plan for FY02 was performed by Cazares and Harris under the supervision of Linda Kaboolian. An excel spreadsheet with their cost estimates is included on the final report CD. Their conclusion was that the costs of the program were outweighed by savings in retention.<sup>31</sup> The test of this conclusion is to compare the retention rates of teachers who hired into Toledo during the year that the program was in abeyance, 1995-96, with the retention rates of teachers hired in the year immediately preceding and the year immediately following that year. The retention rate after 3 years and 5 years is significantly lower for those in the 1995-96 cohort.

---

<sup>28</sup> *The Cost of Teacher Turnover*, prepared for the Texas State Board for Education Certification by Texas Center for Education Research. October 2000.

<sup>29</sup> *The Cost of Turnover*

<sup>30</sup> *The Cost of Turnover*

<sup>31</sup> Cazares, L. and Harris, A. *Professionalism Through Collaboration: A Social Cost-Benefit Analysis of the Toledo Plan*. Kennedy School of Government, Policy Analysis Exercise completed as a requirement for the Master's Degree in Public Policy. April, 2002.

**Task 4. Interview school district and union leadership regarding their perceptions of the effectiveness and efficiency of the TP.**

**Task 5 Interview community stakeholders regarding their perception of the effectiveness and efficiency of the Toledo Plan.**

These two tasks involved interviews with current and past participants in the Toledo Plan. Care was given to interview people who had been renewed as well as those who were not renewed. Interviewees were selected from employment lists provided by the district using random digit methodology in order to find a diverse sample. The numbers of current, and recent interns and Consultant Teachers were too small to be representative in a statistical sense, nevertheless they were diverse. In addition, members of the staff, union officials and public leaders were interviewed. In all, more than 70 interviews were conducted. Interviewees were encouraged to send any supporting documents they wanted examined, and many did.

In addition, a comprehensive review of the coverage of three separate newspapers over the past five years (where available) was also conducted.

Several observations:

- Race has played a role in the life of Toledo and its residents for a long time.
- Many people carry the scars of battles fought over race.
- The Toledo Public School, like many urban districts is on the front line in the struggle of Toledo to come to terms with how race relations have fared in the past, how diverse people choose to live together constructively in the present and plan for a prosperous future for all.
- The Toledo Plan has long been perceived as being unfair to black interns. New interns hire into the district having heard this perception.

## *Is the Intern Program Racially Biased against Black Intern Teachers?*

The records of intern teacher were examined for the academic years 1996-97 through 2003-04 to learn whether success in the intern program was determined by the race of the intern. In all, we were able to assemble a dataset from records held in office of human resources, the archives of the Toledo Plan and union membership role. The final data set contains 1,556 observations. We are very confident that this data set is the most complete and accurate representation of the interns who participated in the program during the years we studied.

Several different tests demonstrated that success was not determined by race. The percentage of teachers in the intern program who were ultimately successful is exactly the same for Blacks and Whites (87%).

Among those who were not successful, some interns were “not renewed” after their first year as a result of a recommendation by the Internal Review Board of the Intern Program (IBoR) while others resigned before recommendations were made about their renewal. Non-renewals get the most attention however, we worried that resignations might be a covert form of “non-renewal”. We do not know if intern teachers who resign do so because they have a foreboding sense that they will not be renewed or because of other, perhaps more pleasant personal considerations. Working with the hypothesis that resignations were the result of discouragement or negative treatment by consulting teachers – a proxy for discrimination, we tested to see whether resignations showed a racial pattern and found that they did not.

A more sophisticated analysis of success in the intern program examined the roles played by race, gender, age and the interactions between these variables. In other words, since we knew from the comparisons above that success did not vary by race, we wanted to know if race, gender and age were operating in combination. So, for example, were older men of color were more less likely to be successful than younger white women? The model which considers the combination of these demographic variables is represented by the equation:

$$\text{Success} = B_0 + B_1\text{Race} + B_2\text{Gender} + B_3\text{Age} + B_4\text{RaceXGender} + B_5\text{GenderxAge} + \text{Constant}$$

After testing several models, it became apparent that the interaction variables were not useful. The best fit was offered by the equation:

$$\text{Success} = B_0 + B_1\text{Race} + B_2\text{Gender} + B_3\text{Age} + \text{Constant}$$

The actual data runs for this equation and tests can be found in the file "Main Test of Racial Bias".

The only variable with a significant coefficient is age, suggesting that older interns have a significantly lower success rate. However, these results should be interpreted with great caution. Data on age is missing for nearly half of the interns. As result, any conclusions about the effect of age on success in the intern program should not be drawn without additional research. Even if interns of different ages had different success rates, it is not clear that this alone represents discrimination. We do not know if the capacity to teach in an urban school district is similarly distributed among different age cohorts. In addition, reports from other school districts suggest that 2<sup>nd</sup> career teachers and older teachers, have a higher failure rate in intern and mentoring programs. If older teacher have significantly lower success rates, then it would be important to examine why they are less successful to see if additional education, professional development and support can improve the success rate. In addition, the cost-effectiveness of recruiting and training older interns should be assessed. Clearly, this is an issue that needs additional research.

Unsuccessful interns experience personal disappointment and the costs to students and districts are considerable. Considering the cost investment made in an intern teacher, it is extremely expensive to recruit, mentor and evaluate someone who has a poor chance of success. |

## Conclusions

Our review focused on several distinct elements of the Toledo Plan: the Plan's standards; its methods for gathering and using data on teacher performance; the mentoring and assistance provided through the Plan; and finally the selection, training, and supervision of CTs. Each element affects either (a) the fairness or accuracy of the Plan's judgments about teacher performance, or (b) the quality and effectiveness/efficacy of the assistance provided to struggling teachers.

One clear conclusion of our review is that the Toledo Plan measures teacher performance more accurately than do traditional evaluation systems. The documentation of each observation under the Toledo PAR plan is more extensive than is typical under traditional evaluation systems, and the observers (CTs) bring specialized expertise in an assignment area. Moreover, observations by CTs are more than twice as frequent as those conducted under traditional evaluation systems. We also note that the assistance provided through the Toledo Plan is also more extensive than the assistance delivered through traditional evaluation systems.

Nevertheless, we conclude that the Toledo Plan—like most programs—can be made better. Our review of the Toledo Plan guidelines and procedures has led us to make several recommendations. We summarize them here and in some cases link related recommendations that were reported separately in the body of this report.

- First, we believe that Toledo can modify its standards to draw a clearer line between satisfactory and unsatisfactory performance. We hope Toledo will try to provide more specified and detailed definitions of performance levels for specific criteria within the standards (i.e., rubrics). We also recommend that Toledo consider whether or not the standards should be made more explicit—more specific—about what good teaching entails. Regardless, we recommend that Toledo examine the standards from time to time, using a collaborative process, to ensure that the standards are informed by the latest

research on effective teaching and are aligned with ongoing efforts to improve education in Toledo.

- CTs are central to the success of the Toledo Plan. Although we do not question the qualifications of current CTs, we suggest that Toledo adopt more detailed hiring criteria for CTs. Additional time set aside for training may help CTs provide greater, more effective assistance at helping teachers improve their instruction. In addition, we recommend additional training focused on observation and rating exercises, which would improve the accuracy and consistency of CTs' judgments.
- Finally, we note that the setting and fulfillment of performance goals for underperforming teachers is a key element of the Toledo Plan. To facilitate better assistance when such a goal is set, Toledo may wish to consider assigning an additional CT so that one person is not playing the role of both mentor and evaluator. Because so much rests on the wording of performance goals, we also recommend that the goal setting process include review of the CT's observation reports by a second CT who would then be required to 'sign off' on the wording of the performance goal.

We believe the Toledo plan currently provides an effective basis for termination decisions—a much better basis than traditional evaluation practices evident today. No system of teacher evaluation is perfect. However, to the extent that Toledo can judge teachers in a way that captures performance more perfectly, the system will be fairer and more effective for Toledo, for the teachers in the program, and ultimately for the children of Toledo.

Some of our recommendations may require investments that may not be feasible at this time. We sincerely hope that Toledo will be able to pursue some of these recommendations, as each might improve the system's ability to improve instruction and its ability to make judgments about teacher performance.

## **Task 6: Provide recommendations regarding the Toledo Plan**

### **Standards for evaluation**

#### ***Recommendations***

We believe Toledo's standards are sufficiently specific but encourage Toledo to consider the potential advantages and disadvantages associated with making the standards more specific (i.e., by expanding on the current 23 criteria). Greater specificity may generate more opportunities for the teacher to learn from the process. Greater specificity may also translate into more reliable measurement of teacher performance. The disadvantage is that making the standards more specific may make them cumbersome to implement (Anderson and Pellicer, 2001). More specific standards may necessitate lengthy training for CTs, or longer observation periods in order to capture all of the information necessary. We recommend that Toledo consider the specificity of its standards, being mindful of these tradeoffs.

### **Distinguishing Levels of Performance**

#### ***Recommendations***

We recommend that Toledo weigh the potential costs and benefits of creating rubrics. There are several advantages to using rubrics. One advantage is that they can facilitate communication between CTs and the teachers they are evaluating. Besides helping teachers better understand what is expected of them, using rubrics facilitates conversations about teaching and improves the consistency of CT's ratings. If a teacher does not understand or disagrees with a rating given, the teacher and the CT can talk in precise terms about what distinguishes an 'unsatisfactory' rating from a 'satisfactory' rating (Danielson and McGreal, 2000; Anderson and Pellicer, 2001). Those conversations allow the teacher being rated to learn about the substance of what good teaching requires.

However, the development of rubrics requires a substantial investment of time. If Toledo is interested in their development, we

recommend that Toledo write and ‘pilot test’ rubrics for a few of its performance criteria—focusing on those criteria that CTs and interns currently find difficult to rate or understand. Toledo would then be able to evaluate whether or not the development of additional rubrics would be worthwhile.

## **Premises about Good Performance**

### ***Recommendations***

We find no flaws in the Toledo Plan’s standards and the premises that they embody. However, we are somewhat concerned that Toledo may not have taken into account shifts in research or evolving views in Toledo about what comprises good teaching. With minor exceptions, the 1984 version of the Teacher Summary Evaluation form is identical to the form used today (see Wise, Darling-Hammond, McLaughlin, and Bernstein, 1984). For example, under the Teaching Procedures category, a criterion called “Voice quality” is now represented by two more specific criteria called “Oral and written communication skills” and “Speech, articulation and voice control.”<sup>32</sup> Since 1984, research on teaching has shifted its attention toward cognitive learning theory and toward the interplay between content and teaching methods (Danielson and McGreal, 2000).

While we do not call for a major revision of the standards, we recommend occasional district-level reviews of the standards to ensure that they continue to meet district objectives. The research base on good teaching is sufficiently diverse to permit disagreement about exactly what premises ought to be contained in PAR standards. However, experts have argued that it is very important for a system of standards to be clear about the premises upon which it is based. In addition, the standards must of course be accepted by those who use them (Danielson and McGreal, 2000).

---

<sup>32</sup> Additional changes include the following: Under the Classroom Management category, one criterion was added (“Confidence/stability”), and another criterion was taken from Personal Characteristics and Professional Responsibility (“Is reasonable fair and impartial in dealing with students”). Under Personal Characteristics and Professional Responsibility lost “Appropriate interaction with pupils” and the criterion “is punctual” became “is punctual and regular in attendance.” Finally, the new form prevents CTs from indicating “Outstanding” levels of performance in Knowledge of Subject and in Personal Characteristics and Professional Responsibility.

## **Using Data to Make Decisions**

### ***Recommendations***

We believe that the procedures used in Toledo for gathering, recording, and using data are strong. Our recommendations focus on the setting of performance goals. Performance goals allow the CT to specify the level of performance that would be ‘satisfactory’ for a given teacher with a given problem. Performance goals thus function as the standard for an individual teacher to meet. While we believe that CTs need discretion in setting those goals, we also believe that significant oversight is warranted.

The Toledo Plan may wish to consider an extra check to ensure that the performance goal issued by a CT is warranted and reflects the standards, including the expected levels of performance. This might be accomplished, for example, by having another CT sign off any time a performance goal is drafted. We expect that this happens informally already, but making it a formal requirement for CTs would guard against mistakes and would improve the consistency and fairness of the process without creating additional administrative burden.

## **Mentoring/Assistance**

### ***Recommendations***

Our review indicates that the assistance provided to interns and intervention teachers is substantial and consistent with research and best practice. While we do not recommend changes to the assistance component of the Toledo Plan, we do note that having the CT fulfill both the role of evaluator and the role of mentor has advantages and disadvantages, as discussed above. An alternative arrangement involving multiple CTs—or one evaluator and one mentor—may be more expensive but may enable the program to better assist teachers after they have been given performance goals.

## **Selecting Consulting Teachers**

### ***Recommendations***

We do not see any problems with the CT selection procedures. The criteria themselves appear job related but could be made more specific, which would be consistent with what experts on PAR programs recommend when defining selection criteria for CTs (Kelly, 2000; Koppich, 2000). Although Toledo may already take into account the important factors discussed here, putting more detailed criteria in writing might strengthen the reputation of the Toledo Plan as a selective program.

## **Training Consulting Teachers**

### ***Recommendations***

We believe that Toledo might increase the effectiveness of its program by expanding the training given to CTs. In the current system, the hiring process ensures that CTs have skills in working with adults, and they expand on those skills during their service as CTs.

We believe that additional training in mentoring and related topics might make mentors more effective at providing assistance to new teachers.

We also recommend that Toledo incorporate more training on observations and scoring, to ensure accurate identification of struggling teachers. One strong practice of the Toledo training is that during the 'shadowing' day, the new CT and the veteran CT observe the same class and then compare their notes. Additional processes like this—perhaps involving review of videotapes—should be used to ensure that CTs have internalized the standards and rate performances consistently. We note that all CTs currently receive training on how to implement the state-mandated Praxis III assessments. That training concludes with a test of their ability to accurately use the Praxis III scoring system and takes four days. Although that training focuses on a different scoring system, Toledo may wish to review the Praxis III training practices to see what can be done to extend the existing training in use of the Toledo Plan's scoring system.

## **Supervising Consulting Teachers**

### ***Recommendations***

We found that the Toledo Plan adequately addresses the supervision of CTs. We offer no recommendations for improvement at this time.

### **Right to Request a Replacement of a Consulting Teacher:**

#### **Recommendation**

In line with some other peer review programs, interns should know that they have the right to request the replacement of their Consulting Teacher. The procedures for such a request should be enumerated in the orientation materials distributed to interns.

### ***Statistics***

#### **Recommendation**

It is very expensive for TPS to hire people who don't succeed in their intern year on the other hand TPS wants high standards for its teachers. A healthy tension has to be kept between these two considerations. In any event, it sends a dual message to rely on the "failure" rate as point of pride. To some, it demonstrates the commitment of the union to teacher quality; to others it is a statement that there is a "quota" of failures that will occur, no matter the ability of the interns. A much more robust measure of success would be a combination of success rate of interns AND retention rate of interns after 3 and 5 years at TPS.

### ***Preventing Failures***

#### **Recommendation**

It appears non-renewals are often for weaknesses in classroom management and that interns in special education classrooms are most likely to have this difficulty. This is important data about what additional professional training opportunities are necessary to increase the success rate of intern teachers. Additional professional inputs to help interns with special education needs and classroom management are necessary. In addition, the Toledo Public Schools

should provide feedback to college and university-based teacher preparation programs on the success rate of their graduates.

### ***Institutionalization of the Toledo Plan***

#### **Recommendation**

The Toledo Plan has been governed for over two decades through a series of operating procedures that have evolved to accommodate new circumstances, are well reasoned but have not been formalized on paper. Instead the institutional memory resides in the relationship between union and administration leaders. The eventual retirement of the leaders from either management or union requires that all aspects of daily operations should be described and memorialized on paper for future participants.

### ***Involvement of School Board Members and Superintendent***

#### **Recommendation**

School board members as well as the superintendent should be required to shadow a consulting teacher for a day and should be required to sit in on a full day of IBoR hearings.

#### ***Governance***

#### **Recommendation**

There is no right or wrong way to govern the IBoR. A comparison of the Toledo Plan with other significant peer review programs shows similarities regarding the presence of the union president, periodic rotation of the chair, and the union/administration split.

However, many programs also have power balancing mechanisms to offset the larger union representation on the governance committee. This is already present in TPS and all other districts in Ohio where state law grants to the superintendent and the school board the right to ultimately decide whether to renew a teacher. The IBoR decision is legally only a recommendation to the superintendent.

Where power balancing mechanisms are built into the governance system, they often take the form of requiring a supermajority of votes to not accept the recommendation of a CT or to not renew an intern. The rationale for this is that to not accept a CT's recommendation should require both management and union votes. The only way to achieve this is to require at least a 6 votes to not accept. Additionally, the decision to not renew has great impact on the intern in question. As a result, it is reasonable that the decision to not renew should be made by a supermajority of votes (even if "not renew" is recommended by the CT).

It is thereby recommended that a 6/3 vote be required to *not accept a CT's recommendation* or to *accept a CT's recommendation to not renew an intern*.

### ***Recruitment of Minority Faculty***

#### **Recommendation**

The pool of minority candidates for positions in TPS is very small and the competition by other districts for these teachers is very strong. Even if a large number of positions opened up in TPS, it is very unlikely that the demographics of the teaching corps will change quickly. In addition, long standing criticism of the district by advocates of diversity paradoxically has lead candidates to believe that TPS is a school district where they would be welcomed or will not succeed. It is strongly recommended that the Superintendent and TFT President, with the participation of appropriate CTs and past interns, jointly recruit minority candidates, advertising the mentoring and professional development opportunities provided by the Toledo Plan. It is also recommended that TPS and TFT explore ways other ways to jointly recruit minority teachers.

## References

- Alliance for Excellent Education. (2004). *Tapping the potential: Retaining and developing high-quality new teachers*. Washington, D.C.: Author.
- American Federation of Teachers and National Education Association. (1998, September 25-27). Peer assistance & peer review: An AFT/NEA handbook. *Shaping the profession that shapes the future: An AFT/NEA conference on teacher quality*. Washington, DC: Author.
- Anderson, L., & Pellicer, L. (2001). *Teacher Peer Assistance and Review: A practical guide for teachers and administrators*. Thousand Oaks, CA: Corwin Press, Inc.
- Ball, D.L., Camburn, E., Correnti, R., Phelps, G., & Wallace, R. (1999). *New Tools for Research on Instruction and Instructional Policy: A Web-based Teacher Log*. Seattle, WA: Center for the study of teaching and policy.
- Calderón, M. (1999). Teachers learning communities for cooperation in diverse settings. *Theory into Practice*, 38(2), 94-99.
- Chrisco, I.M. (1989, May). Peer assistance works. *Educational Leadership*, 31-32.
- Consortium for Policy Research in Education (1998). Peer involvement in knowledge and skill assessment. Handout from seminar "Resigning Teacher Compensation" held in Madison, WI. on October 21, 1998.
- Costa, A., & Garmston R. (2000). Peer assistance and review: Potentials and pitfalls. In Bloom, G., & Goldstein, J. (Eds.). *The peer assistance and review reader*. (pp. 63- 79) Santa Cruz, CA: New Teacher Center.

Cromwell, S. (1998, February 9). Pushing for peer-review. *Education World*. Retrieved

December 22, 2004, from

[http://www.education-world.com/a\\_admin/admin047.shtml](http://www.education-world.com/a_admin/admin047.shtml)

Danielson, C., & McGreal, T.L. (2000). *Teacher evaluation: To enhance professional practice*. Princeton, NJ: Educational Testing Service.

Dell'Olio, J.M. (1998, Winter). 'Hearing myself': Reflection-in-action in experienced teachers' peer-assistance behaviors. *Journal of Curriculum & Supervision*, 13(2), 184.

Editorial Projects in Education. (2005). Quality Counts 2005, No small change: Targeting money toward student performance. Bethesda, MD: Author.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter?: High school certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-146.

Goldstein, J. (2004, Summer). Making sense of distributed leadership: The case of peer assistance and review. *Educational Evaluation and Policy Analysis*, 26(2), 173- 197.

Haselkorn, D., & Fideler, E. (1999). Learning the ropes: Urban teacher induction programs and practices in the United States. Belmont, MA: Recruiting New Teachers, Inc.

Hertling, E. (1999, May). Peer review of teachers. *ERIC Digest* 126. Retrieved December 13, 2004, from <http://eric.uoregon.edu/publications/digests/digest126.html>

Ingersoll, R., & Kralik, J.M. (2004, February). *The impact of mentoring on teacher*

*retention: What the research says.* Denver CO: Education Commission of the States.

Institute for Education Reform (2000). Peer assistance and review: Working models across the country. Sacramento, CA: California State University. Retrieved February 15, 2005, from <http://www.calstate.edu/ier/reports/PARReport.pdf>

Kelly, P.P. (2000). Teacher peer review recommendations. In Bloom, G., & Goldstein, J. (Eds.). *The peer assistance and review reader.* (pp. 32-49) Santa Cruz, CA: New Teacher Center.

Kelly, P.P. (1998). A Comparative Analysis of Teacher Peer Review Programs in Four Urban Districts: Professional Unionism in Action. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Kerchner, C.T. (2001, Fall). Deindustrialization: Why teachers must come to regard—and organize—themselves as mind workers. *Education Next*, 46-50.

Kerchner, C.T., Koppich, J.E., & Weeres, J.G. (1998). *United mind workers: Teachers and unions in the knowledge society.* San Francisco, CA: Jossey-Bass Publishers. Retrieved December 23, 2004, from <http://63.197.216.234/crci/mindworders/udpages/newck5.htm>

Koppich, J.E. (2000). Peer assistance and review: Enhancing what teachers know and can do. In Bloom, G., & Goldstein, J. (Eds.). *The peer assistance and review reader.* (pp. 19-31) Santa Cruz, CA: New Teacher Center.

Lawrence, D. (2004, October). Ask the right questions: Rethinking high school reform

[Electronic Version]. In I. Bond & J. Ayers (Eds.), *Profiles in Leadership: Innovative Approaches to Transforming the American High School*. (pp.65-72). Retrieved January 28, 2005, from <http://www.all4ed.org/publications/ProfilesInLeadership/ProfilesInLeadership.pdf>

Lawrence, D. (2003). *The Toledo Plan: Practical advice for beginners*. Toledo, OH: Toledo Public Schools and Toledo Federation of Teachers.

Lopez, A., Lash, A., Schaffner, M., Shields, P., & Wagner, M. (2004). *Review of research on the impact of beginning teacher induction on teacher quality and retention*. Menlo Park, CA: SRI International.

Medley, D.M., Coker, H., & Soar, R.S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach*. New York, NY: Longman, Inc.

National Association of Secondary School Principals. 2001. *Priorities and barriers in high school leadership: A survey of principals*. Santa Monica, CA: Milken Family Foundation and Reston, VA: National Association of Secondary School Principals.

National Research Council. (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.

Porter, A. Youngs, P., & Odden, A. (2001). *Advances in teacher assessments and their uses*. In V. Richardson (Ed.), *Handbook of research on teaching* (4<sup>th</sup> ed., pp. 259- 297). Washington, DC: American Educational Research Association.

Schools and Staffing Survey, 1999-2000. [Data file]. Washington, DC: National Center

for Education Statistics.

Shields, P., Haslam, B., LaGuarda, K., Lopez, A., Lash, A., Schaffner, M., Wayne, A., &

Wagner M. (2004). Identification and description of promising models of teacher induction. Menlo Park, CA: SRI International.

Smith, T.M., & Ingersoll, R.M. (2004, Fall). What are the effects of induction and

mentoring on beginning teacher turnover? *American Educational Research Journal*, 41(3), 681-714.

Stedman, P., & Stroot, S.A. (1998, February). Teachers helping teachers. *Educational*

*Leadership*, 55(5), 37-38.

Strong, M. (2004, October 25). *Induction, mentoring and teacher retention: A summary*

*of the research*. Santa Cruz, CA: The New Teacher Center.

Stroot, S., Fowlkes, J., Langholz, J., Paxton, S., Stedman, P., Steffes, L., & Valtman, A.

(1999, January/February). Impact of a collaborative peer assistance and review model on entry-year teachers in a large urban school setting. *Journal of Teacher Education*, 50(1).

Stroot, S., Keil, V., Stedman, P., Lohr, L., Faust, R., Schincariol-Randall, L., Sullivan, A., Czerniak, G., Kuchcinski, J., Orel, N., & Richter, M. (1998). *Peer assistance and review guidebook*. Columbus, OH: Ohio Department of Education.

Toledo Public Schools (1999). Consultant handbook. Toledo, OH: Author.

Wayne, A.J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, 73(1), 89-122.

Weiss, E.M., & Weiss, S.G. (1998, December). *New directions in teacher evaluation*.

ERIC Digest. Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education. (ERIC Document Reproduction Service No. ED 429 052). Retrieved December

15, 2004, from

<http://edstar.ncrel.org/mn/ViewEssay.asp?IssueID=38&EssayID=181>

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

## **Appendix A: Teacher Summary Evaluation Report**

[replace this page with PDF printout of teacher summary evaluation report]

## Appendix B: Explanation of Levels Within Standards

Level	Explanation	Examples
Level 1	<p><b>Statements of vision, core principals, propositions</b>            Statements at this level are highly generalized, abstract. They are important as statements of values, but are not designed for making valid inferences from evidence of practice.</p>	<p>Highly accomplished teachers are committed to their students and their learning (NBPTS)</p> <p>Effective teachers understand that learning is an active process of constructing knowledge (Praxis)</p>
Level 2	<p><b>Statements of category</b>            These statements define the main categories of teaching practice. Together they should represent a coherent framework of what teachers know and do – an “architecture” of practice. They should reflect the Level 1 norms and values.</p> <p><i>Variously called “Domains” “Areas of competence” “Dimensions”</i></p> <p>These statements constitute the main components of the in the “domain” of what is to be assessed.</p>	<p>Communicating, interacting, and working with students and others (National Competency Framework for Beginning teaching)</p> <p>Organizing content knowledge for student learning (Praxis III)</p>
Level 3	<p><b>Statements of elements</b>            Statements at this level are descriptions and elaborations of Level 2 statements. They should elucidate the Level 2 statements and reflect the Level 1 norms and values.</p> <p>Statements at Level 3 should be useful in making judgments about a teacher’s performance. They point to elements of observable appropriate behavior, but transcend references to specific practices.</p> <p><i>Variously called “Elements” “Indicators” “Components” “Criteria”</i></p>	<p>Establish learning environments that acknowledge the concerns, values, and priorities of students’ families, cultures, and communities (Professional Standards for Teachers, Queensland)</p> <p>Take account of ethnic and cultural diversity to enrich the curriculum and raise achievement (Induction standards, UK)</p>

Source: Ingvarson and Kleinhenz, 2003.

## Appendix C: Comparison of the criteria for the Toledo Plan and the PAR program in Poway Unified School District

Comparison of the criteria for the Toledo Plan and the PAR program in Poway Unified School District		
<i>Toledo (Ohio)</i>	<i>Poway (California)</i>	<i>NBPTS</i>
Skill in planning	Designs long-range plans to accomplish yearly goals	
	Selects appropriate lesson objectives	
	Designs lessons that include elements for essential learning	
	Establishes clear academic standards	
		Meaningful applications of knowledge
Skill in assessment and evaluation	Monitors student learning	Assessment
Skill in making assignments		
Skill in developing good work-study habits		
Resourceful use of instructional materials	Uses appropriate learning materials	Instructional resources
Skill in using motivating techniques		
Skill in questioning techniques	Delivers effective instruction	
Ability to recognize and provide for individual differences	Delivers effective instruction; uses a variety of instructional strategies to meet the needs of students; creates learning opportunities for all students	Knowledge of students; respect for diversity; multiple paths to knowledge
Oral and written communication skills	Delivers effective instructional uses assessment results to give students and /or parents timely, accurate, and constructive feedback	

**Comparison of the criteria for the Toledo Plan and the PAR program in Poway Unified School District (*continued*)**

<i>Toledo (Ohio)</i>	<i>Poway (California)</i>	<i>NBPTS</i>
Adequate speech, articulation, and voice quality	Delivers effective instruction	
Effective classroom facilitation and control	Maintains a positive learning environment, which promotes appropriate behavior and self-esteem	Learning environment
Effective interaction with pupils		
Efficient classroom routines		
	Organizes physical space	
	Effectively manages instruction time	
Confidence/stability		
Reasonableness, fairness, and impartiality in dealing with students		
Knowledge of subject; academic preparation	Demonstrates subject matter competence	Knowledge of content and curriculum
Genuine interest in teaching		
Adequate personal appearance		
Skill in adapting to change	Adjusts teaching and learning based on assessment	
	Grows and develops professionally	Reflection
Adherence to accepted policies and procedures of Toledo Public Schools	Complies with established rules, regulations, policies, and laws	
Acceptance of responsibility both inside and outside the classroom	Shares the responsibility for the smooth operation of the school	Contributions to the profession
<i>(continued)</i>		

**Comparison of the criteria for the Toledo Plan and the PAR program in Poway Unified School District (*continued*)**

<i>Toledo (Ohio)</i>	<i>Poway (California)</i>	<i>NBPTS</i>
Cooperative approach toward parents and school personnel		Family involvement
Punctuality and regularity in attendance		

Source: Anderson and Pellicer, 2000, pp.63-5.

## **Appendix D: Principal Summary**

[replace this page with PDF printout of form "Principal Summary"]

Appendix E.

**COMPARISON OF PEER REVIEW PROGRAMS IN FIVE DISTRICTS**

	<b>TOLEDO</b>	<b>CINCINNATI</b>	<b>MINNEAPOLIS</b>	<b>ROCHESTER</b>	<b>COLUMBUS</b>
<b>PROGRAM NAME</b>	Toledo Plan	Peer Assistance and Evaluation Program	Achievement of Tenure Process (AofT)	Career in Teaching Plan	Peer Assistance and Review Program
<b>START DATE</b>	1981	1985	1992	1999	1985
<b>PRE-TENURE PERIOD FOR TEACHERS</b>	2 years: intern year and a second year.	2 years, maximum	3 years, maximum New hires must complete the AofT process by the end of 3d year.	3 years. 2 years for new hires with prior experience .	1 year of probation. 3 additional years of one year contracts.
<b>DO PRE-TENURE TEACHERS PAY FULL UNION DUES?</b>	50%	100%	100%	100%	100%
<b>DO NEW HIRES HAVE THE RIGHT TO UNION REPRESENTATION IN ANY APPEAL OR CHALLENGE TO THE PEER REVIEW PROCESS?</b>	No. As probationary employees they are considered not to have this right.	Limited. Probationary teachers can grieve <i>procedural</i> issues during evaluation, and union field reps will advise them but field reps. will not accompany teachers to any meeting. If probationary teacher is recommended for	No. As probationary employees they are considered not to have this right	No. As probationary employees they are considered not to have this right.	Yes. Interns have "complete access to union representation during peer review." May be accompanied by union re. And/or lawyer in hearing before HR Director to challenge PAR recommendation.

		termination by PRP, teacher can appeal to Superintendent and Bd. of Ed. and can be accompanied by union field rep.			
--	--	--	--	--	--

	<b>TOLEDO</b>	<b>CINCINNATI</b>	<b>MINNEAPOLIS</b>	<b>ROCHESTER</b>	<b>COLUMBUS</b>
<b>DOES DISTRICT USE CONSULTING TEACHERS OR AN EQUIVALENT</b>	Yes. There are currently 13 consulting teachers. They are appointed as needed, each having a case load of 10-12.	Yes. Up to 20 part-time consulting teachers. District prefers using "Lead Teachers," but not all CTs are lead teachers	Yes/modified. They are mentors. Each probationary teacher has his/her own AofT <i>team</i> . The teacher can also have a mentor who is assigned <i>when the teacher needs help</i> .	Yes/modified. Interns are evaluated by "Lead Teachers/Mentors" <i>and</i> by their (teacher's) supervisors.	Yes. There are typically 25-30 appointed CTs that evaluate intern teachers.
<b>ARE EVALUATIONS SUMMATIVE OR FORMATIVE OR BOTH SIMULTANEOUSLY</b>	Both simultaneously	Both simultaneously	Both simultaneously AofT team can recommend non-renewal, but only after a teacher needing help has had a mentor for at least 3 months. But, mentor is AofT member.	Both simultaneously Lead Teachers/Mentors are one of two primary sources of input on advancement and renewal recommendations to the CIT Joint Governing Panel	Both simultaneously

	<b>TOLEDO</b>	<b>CINCINNATI</b>	<b>MINNEAPOLIS</b>	<b>ROCHESTER</b>	<b>COLUMBUS</b>
<b>IS THERE PEER REVIEW/EVALUATION FOR VETERAN TEACHERS</b>	Yes. Veteran teachers who are struggling can be placed in intervention, either at their request or involuntarily.	Yes. There is intervention for deficient veteran teachers.	Yes. All teachers have a professional development plan and a professional development team in place at all times. Teachers select their PDP teams to serve as "critical friends" through the year.	Yes. It can be voluntary or involuntary.	Yes. It can be voluntary or involuntary. "This year we are evaluating 270 to 300 teachers. About 25 of them are experienced veteran teachers."
<b>DO VETERAN TEACHERS HAVE MORE DUE PROCESS RIGHTS THAN PROBATIONARY TEACHERS</b>	Yes. Non probationary teachers are considered to have property rights in their jobs, and are thus entitled to due process protection.	Yes. Non probationary teachers are considered to have property rights in their jobs, and are thus entitled to due process protection.	Yes. Non probationary teachers are considered to have property rights in their jobs, and are thus entitled to due process protection.	Yes. Non probationary teachers are considered to have property rights in their jobs, and are thus entitled to due process protection.	Yes. Non probationary teachers are considered to have property rights in their jobs, and are thus entitled to due process protection.

	<b>TOLEDO</b>	<b>CINCINNATI</b>	<b>MINNEAPOLIS</b>	<b>ROCHESTER</b>	<b>COLUMBUS</b>
<b>MAKE UP OF PEER REVIEW PANEL</b>	Union majority. Intern Board of Review (IBOR). 5 teachers selected by TFT President and 4 administrators selected by SOS.	Even distribution. 5 teachers selected by union and 5 administrators selected by superintendent	Administrators on majority. AofT REVIEW Team. (not same as teacher's personal AofT team). REVIEW team consists of "the site Professional Development Process (PDP) coordinator, union steward, 2 other administrators and 1 or more members of teacher's AofT team.	Even distribution. CIT Joint Governing Panel— (Career In Teaching). 6 members appointed by RTA; 6 members appointed by SOS. Quorum is 7. Decisions require affirmative vote of at least 7.	Union majority. 7 members: 3 administrators and 4 union members. Administrators are Director of HR, a building principal, and the Director of Contract _____. Union members are the Union President and one rep from elementary, middle and high school levels. 5 votes needed to carry.
<b>CHAIR OF REVIEW PANEL</b>	Rotates annually between Union President and Assistant SOS	Rotates. Federation President and Deputy SOS serve alternately on an annual basis as facilitator of PRP meetings and are	"There isn't really a chair. HR does the leg work and the meetings are run by a mediator from the State Bureau of Mediation	Rotates annually between RTA and District. Year begins July 1.	Columbus Intended to rotate annually. "Actually, they both co-chair all the time."



	<b>TOLEDO</b>	<b>CINCINNATI</b>	<b>MINNEAPOLIS</b>	<b>ROCHESTER</b>	<b>COLUMBUS</b>
<b>ARE CTs COMPENSATED</b>	Yes. \$6,150 in 2004-2005.	Yes. Lead Teachers automatically receive \$5,500 (04-05) over base. CTs who are not Lead Teachers receive \$3,000.	Yes. Mentors receive \$5,000 and a professional account (\$800?) to purchase materials.	Yes. Lead Teachers who are activated as mentors receive a 5% or 10% increase, depending on duties.	Yes. CTs get a supplemental contract for an amount equal to 20% of a 1 <sup>st</sup> year teacher's base pay.
<b>CAN TEACHERS APPEAL ASSIGNMENT OF CTs?/ MENTORS</b>	No. TP does not allow for change of CT.	Yes. Teachers can ask the PRP for a change. Changes are rarely granted, but sometimes they are.	They choose their own. Uncertain whether they can change.	Yes. Upon request a teacher can change her mentor changed. Easily done.	Yes. The PAR Panel will consider requests for change. It sounds like a flexible approach.
<b>ARE CTs/ MENTORS EVALUATED?</b>	Yes (from DL 3/4/05) Interns do formal evaluations on various aspects of their experience. Results compiled by research department and given to Sheila and Fran. Also, CTs are informally evaluated every time they present before IBOR	Yes. CTs are monitored and evaluated by PRP on a regular basis. PRP can return a CT to classroom at any time because of performance concerns.		Yes. Each member of the CIT Panel has a caseload of mentors to observe and evaluate every year. Interns also evaluate mentors, based on a check list of performance criteria. Mentors must re-apply every 2 years.	Yes. PAR Panel reviews CTs every year. Panel receives input from a selected group of interns. Some CTs are evaluated out of the program before their term expires.

<b>ROLE OF PRINCIPAL IN EVALUATION OF NEW HIRES</b>	<p>None. Principals have no part in evaluating teaching in first year. Classroom presence is discouraged. As of 1998 they were allowed to refer teachers to TP without seeking union approval.</p>	<p>Some. Principals shall receive copies of CT monthly evaluation reports as completed. May visit any teacher; may disagree with CT; may inform PRP of disagreement with CT</p>	<p>Significant. Principals are on the AofT team . They are involved in evaluation and assessment of probationary teachers.</p>	<p>Significant. Teacher supervisors jointly evaluate Intern Teachers with "Lead Teacher/Mentors."</p>	<p>Varies. Principals can observe interns, or leave it to CTs. CTs are required to keep principals informed of their activities. Principals must write a report to PAR Panel, and can disagree with CT.</p>
---	--	---	--	---	---

## Appendix F. Comparison of Districts on Termination

### **Cincinnati**

**Termination: 4.5% from 1997 to 2002**

**Retention:**

**Cost:**

#### Termination

Almost twice as many teachers in Cincinnati were dismissed under peer review as under administrator evaluations, before peer review was instituted.<sup>33</sup>

From 1985-2002, 3.5% of interns were dismissed through peer review (77 out of 2199).<sup>34</sup>

From 1997 through 2002, 4.5% of interns were dismissed through peer review (30 out of 666). Annual dismissal rates ranged from 2.9% (1997-1998) to 7.0% (2000-2001).<sup>35</sup>

From 1986 through 2002, 198 veteran teachers were referred to intervention. 83 were removed from the classroom (42%). 34 improved to a satisfactory level. 71 left the district voluntarily (36%). 10 were still in progress at the time the statistic was released.<sup>36</sup>

#### Retention

No data.

---

<sup>33</sup>ERIC 126, Pg.3.

<sup>34</sup> From CFT document. Data received from different sources about Cincinnati is slightly inconsistent, but close enough for our purposes.

<sup>35</sup> CFT Summary Statistics. For the five year period beginning 1997-98 30 of 666 apprentices were evaluated out of the district and 41 of 89 veterans who entered intervention were evaluated out of the district.

<sup>36</sup> From CFT document.

## Cost

Before peer review began in 1985, litigation costs were much higher than they have been since. Two reasons: termination cases are no longer arbitrated; and lawsuits over termination are now few and far between. There have been three suits since 1985.<sup>37</sup>

Cincinnati is permitted to budget up to 20 FTEs and \$1 million for the peer assistance and evaluation program.

### **Columbus**

**Termination: 5%-7% from 1985 to present. 1% - 2% before 1985.**

**Retention: Average 80% over five years.**

**Cost:**

## Termination

The termination rate has been 5-7% over 19 years of program. Now it is a little higher, because classroom management is becoming more of an issue. Before the program the termination rate was 1-2%.<sup>38</sup>

20% of veteran teachers who go through intervention leave the system.<sup>39</sup>

## Retention

80% of new teachers are still at the district after 5 years. 5-7% get evaluated out, and the rest choose to leave. "In other urban districts without peer review, 50% on new hires leave after five years. That alone is justification for the program."<sup>40</sup>

---

<sup>37</sup> Telephone interview with Denise Hewitt, CFT. 3/18/05.

<sup>38</sup> Telephone interview with John Grossman. 3/15/05.

<sup>39</sup> ERIC 126. Pg. 3.

<sup>40</sup> Telephone interview with John Grossman. 3/15/05.

## Cost

Peer review program in Columbus costs \$2-3 million annually. Most of the cost is in paying CTs, and the number of CTs depends on number of new hires. There were 2 years following buy-out programs that skew the results because more teachers were hired to replace those who bought out.<sup>41</sup>

### **Rochester**

**Termination: 8.2%**

**Retention: 85% over five years 1998-2003**

**Cost:**

Termination - Rochester terminates 8-12% of new teachers every year.<sup>42</sup>

Retention -In 1998, 95% of those who entered teaching in Rochester ten years earlier and who received mentoring were still there.<sup>43</sup>

In 1998, Rochester's five year retention rate overall was 86%.<sup>44</sup>

Cost – No data

---

<sup>41</sup> Telephone interview with John Grossman. 3/15/05.

<sup>42</sup> Adam Urbanski. Telephone interview. 3/17/05.

<sup>43</sup> Koppich, v.

<sup>44</sup> Koppich, 26.