Teacher Evaluation

Melinda Sayavedra
Oregon State University and Western Oregon University

Accredited universities normally include a standard that addresses faculty evaluation. It may contain references to performance criteria and procedures and usually emphasizes the need for faculty evaluations to be systematic, regular, fair, objective and relevant to achieving the goals of the institution. Accredited language programs usually have a similar standard that addresses instructor evaluation. However, quite often performance criteria are never spelled out and teacher effectiveness is measured only by student appraisals of the course and the instructor. Public K-12 programs normally have agreed upon standards, but schools all too often measure teacher effectiveness solely by student test scores. The research is very clear that neither of these constitutes a fair and relevant evaluation. Teaching large groups of individuals with varied skills and backgrounds in diverse settings is not easy and neither is measuring its effectiveness.

An effective evaluation system needs to entail more than student test scores and appraisals. It needs to include professional standards with performance criteria and indicators spelled out and endorsed by the language teachers, tools and procedures for self-assessment, peer mentoring, mentor coaching and supervisor evaluation as well as procedures and suggestions for other types of reports and evidence of teacher performance. The aim is to create an evaluation system that promotes teacher growth and gathers evidence for making high-stakes decisions.

Purposes of Evaluation

There are two types of teacher evaluations: summative and formative. Summative evaluation measures and rates teachers. It is used for determining promotion, tenure, merit raises, awards, and dismissal. Formative evaluation focuses on teacher development and improvement (Marzano, 2012). In formative evaluation, teachers are "*participants in*, not *recipients of*, their own evaluations" (NYSED, 2012. p. 3). The goal in formative evaluation is improving instructional practices and continuing growth as a professional. It is collaborative in nature and promotes critical self-reflection and goal-setting. Formative evaluation systems are modeled after clinical supervision, which Mosher & Purpel (1972, p.78) define as "the improvement of instruction." Clinical supervision is multi-directional, supportive and constructive, and emphasizes making positive changes in teacher effectiveness.

Research suggests that the two types of evaluation be kept separate (Marzano, 2012; Rindler, 1994; Felder & Brent, 2004) and that teachers be made fully aware of which type of evaluation is taking place, who will see the reports and where they will be stored. Programs need to determine which procedures and tools are for summative evaluation and which are for formative evaluation. Where formative evaluations and processes are used and do not overlap with summative evaluations, teacher growth is more likely to occur (NYSUT, 2011; Rindler, 1994).

Implementation of an evaluation system must include trust and respect between teachers and evaluators. If formative

evaluations are used for making accountability decisions rather than for improving teaching over time, trust and respect will be undermined. Instead of building a culture of growth, a culture of fear will be established. No one will want to reflect on and share their weaknesses and areas for improvement if the evaluation is being used to determine pay raises, or worse, dismissal (Rindler, 1994).

In general, beginning-service teachers are best evaluated through formative assessment and given time to reflect and make changes as needed. When it is time to do summative evaluation, teachers need to be informed ahead of time and the expectations, process, and tools to be used in evaluating made clear. The goal is to maintain a clear delineation between formative and summative assessment (Rindler, 1994).

Professional Standards and Performance Criteria

Research supports the need for establishing clear teaching and professional development standards that teachers understand and endorse prior to implementation (Rindler, 1994; Heneman et al, 2006; Darling-Hammond, 2012). Having standards does not mean instituting standardization. Effective, sustainable standards invite diversity (Hargreaves & Fink, 2004). We want teachers who are creative and innovative in designing lessons and materials and in solving problems and meeting challenges. We want teachers who meet the students where they are and move them forward in their English proficiency and language skills. But we also recognize the broad and deep body of research that describes what makes a teacher effective and how language is best taught and learned.

Multiple measures should be used to get a full picture of a teacher's performance (Theall, 2002; Felder & Brent, 2004; CTL, Uof NC, 1994, Rindler, 1994), understanding of language teaching and learning, and overall professionalism, but without over-burdening teachers or evaluators. Different types of evidence are needed for establishing whether different standards are being met. These are the biggest challenges in creating an effective teacher evaluation system. If the aim is teacher growth and improvement, then making expectations clear, and setting achievable goals and a reasonable timeframe is essential.

Evaluation Procedures

The aim is to create an evaluation system that promotes teacher growth.

Before any evaluation procedures take place, as noted above, teachers must be informed of and endorse the standards that describe effective language teachers and teaching, and be familiar with the instruments used for the evaluations. Felder and Brent, (2004) and Rindler (1994) and also suggest that if teacher development is the objective, then providing opportunities for self-evaluation and peer feedback as well as collecting student appraisals and supervisor evaluations is important. No matter who is doing the evaluation, formative evaluation should always be kept confidential.

Self-Assessment

An important aspect of formative assessment to promote teacher development is the opportunity for teachers to reflect on their own teaching, to self-assess and discuss their findings with peers, mentors, supervisors or teaching consultants without fear of these being misused in high-stakes decision-making. Self-assessment should be systematic, not haphazard. "It requires discipline and

perseverance, but the results are worth it" (Brown, 2007, p. 492).

A logical first step in self-assessment is to have teachers rate themselves on the standards. Teachers should be encouraged to identify one or two specific elements they would like to improve (Brown, 2007). They should then monitor those elements, set specific goals to change behavior, identify steps to meet their goals, and collect evidence to assess progress. Taking time to reflect and record, or discuss, the self-assessment is important for synthesizing what was learned and for making real changes. More experienced teachers may need to focus on addressing the questions, "What am I doing that I've probably done the same way for too long?" (Weimer, 2013, p. 1), 'How can I update and challenge myself?', 'What have I avoided learning more deeply?'

Self-assessment should be systematic, not haphazard.

After that, teachers may identify one or two specific performance criteria they feel they might improve. Examples include: increase knowledge of theory and best practices in teaching speaking; implement the use of technology appropriately and effectively in instructional activities; increase the amount of student-talk vs. teacher-talk; develop and implement routines to manage activities and transitions; maintain clear standards for student behavior; and build community in the classroom - whatever a teacher identifies as an aspect of his or her teaching that could be improved or updated.

Goals can also grow out of information gained through self-recorded videos of lesson delivery, peer mentoring, or a mentor coach's or supervisor's observations. Teacher development is multi-faceted and multi-directional. Information sought and gained can come from multiple sources.

Peer Mentoring

Felder and Brent (2004, p. 201) caution that peer evaluation needs to be carefully constructed, that "a single observed class may not be representative of someone's normal teaching," and that teachers may have very different ideas of what good teaching looks like. Although the latter can be dealt with by having clear standards that everyone understands and endorses, teachers still need to feel that the purpose of peer mentoring is not to judge, but to support professional growth.

Feedback from teachers who have participated in peer mentoring is often quite positive. Teachers frequently come away from the experience with greater insight about teaching and appreciation for the diverse ways in which material can be covered (Huston & Weaver, 2008). The observing teacher often learns as much as, if not more than, the teacher being observed, garnering new ideas, staying fresh and sharpening skills (Brown, 2007).

Peer mentoring is not about evaluating, judging or rating. Having teachers rate one another can undermine trust and collegiality. Peer mentoring is about helping one another. This can be done by focusing on one or two aspects of a teacher's instructional plans, lesson delivery, or classroom environment, discussing what it looks like, and how it might look different.

Self-reflections can be used to help focus peer mentoring. Teachers identify some aspect they'd like a peer to pay attention to during the observation. For example, the teacher might ask the peer mentor to note how she or he gives directions for an activity to students and to write down exactly what the teacher says and does when giving

directions. The peer mentor might be asked to make a map of teacher movement during class – where was the teacher during various parts of the lesson? Or the peer mentor might be asked to look at a lesson plan and give suggestions for scaffolding it through a gradual release model, and then observe how it played out in the classroom.

Beddes et al. (2012) created and presented a model for peer mentoring based on the work of Huston and Weaver (2008) as well as information from the National School Reform Faculty at www.nsrfharmony.org. They use the following approach for peer mentoring. First, teachers identify a partner for reciprocal peer coaching. They agree to give one another three hours of their time during the term: 30 minutes for a pre-observation conversation, a one-hour focused observation (reciprocated), and a 30-minute post-observation discussion.

During the 30-minute pre-observation discussion both teachers identify the focus area for their observation; they agree on a note-taking strategy (a map, a T-chart of teacher-talk and student-talk, etc.); and they set up ground rules such as where the observer will sit. Each teacher observes the other for an hour or more focusing just on the identified aspect and takes the agreed upon type of notes. The observer may also reflect and jot down ideas or suggestions if the other teacher wishes to receive additional feedback.

Teachers meet again in a 30-minute post-observation discussion to show the notes, discuss what was noted, and share suggestions. Having the discussion focus on both teachers' focus questions helps avoid judgments and builds an environment of sharing discoveries and ideas.

> *Feedback to teachers needs to be comprehensive, useful and specific.*

A common concern about peer mentoring is the amount of time it takes. However, three hours out of a term does not seem taxing. Vidmar (2006) suggests that since teachers pick the peer they want to work with, they can choose someone whose time schedule meshes well with their own, and that pre- and post-conferencing can be likened to taking a coffee break together. Teachers could plan their lessons together as suggested by Zwart, Wubbels, Bergen, & Bolhuis (2009), and use this time to set a focus for an observation.

Mentor Coaching

Mentor coaching is formative feedback done by a supervisor, teaching consultant or a more experienced teacher colleague assigned to the role. Mentor coaching can be done much the same way as peer mentoring, though reciprocation may or may not be part of mentor coaching. Mentor coaching can entail choosing, with the teacher, one or two elements to focus on during the lesson. The observation might also be more general in nature – a pre-conference to establish what the lesson will cover, an observation, and feedback on what was done well and what might be done differently. This can also be a good opportunity to discuss professional goals with teachers, whether those are in lesson design, lesson delivery, professional development, or focused on some other standard. Teachers can share with the mentor the areas they would like to improve or develop, their implementation plans and how they will assess whether they've met their goals.

A mentor coach can also use a more formal, comprehensive evaluation tool to observe and discuss teachers' strengths and areas for improvement, but it must be very clear that it will not become part of a teach-

er's summative evaluation. It is strictly for growth purposes. Mentor coaching, when done by the supervisor who will also be doing the summative observation and evaluation on the teacher, can help build trust, good communication, and credibility that is important in effective supervisor evaluation.

What happens during feedback is key to teacher growth. According to Rindler (1994), more important than the number of observations done by the mentor coach or supervisor is the time spent on the evaluation process. Teachers value and trust programs that allow for more time on evaluation and use that time well. Filling out lengthy forms or pulling together large portfolios only to have a supervisor briefly scan the material or make little or no effort to carefully examine and understand the contents is felt by teachers to be a waste of time (Rindler, 1994). Feelings of resentment, rather than openness to growth can result.

> *Student appraisals, while important, cannot give a full picture of the effectiveness of a teacher.*

Feedback to teachers needs to be comprehensive, useful and specific. Significant to teacher growth, according to Rindler's study (1994) was the usefulness of suggestions provided, its basis in current research, and the specificity of the feedback. Also important to the impact on teacher growth was the credibility of the evaluator, the evaluator's ability to model suggestions and the level of trust. As mentioned earlier, having the evaluation focus on clear standards that are endorsed by the teacher had a significant impact on teacher improvement as well.

Evaluators need to be well-trained and have good interpersonal skills (Darling-Hammond, 2012; Rindler, 1994). Mentor coaches and supervisors need to be well-versed in current second language acquisi-tion theory and practice to establish credibility. They need to be good listeners to establish trust. Mentor coaches and supervisors need to let teachers know specifically what was done well and offer specific ideas on what might be done differently based on the standards. Feedback should be given in writing as well as discussed during conferencing. Coaches and supervisors also need to be able to point teachers to specific resources for making changes or improvements and help ensure that these are available to the teachers (Darling-Hammond, 2012).

Student Appraisals

Student appraisals, while important, cannot give a full picture of the effectiveness of a teacher. Many student appraisal forms don't provide enough meaningful feedback for teachers to make effective changes in their teaching. Rindler (1994) reports that teachers found that a summary of students' comments was more useful than bubbled-in responses. Research also suggests that sharing the results of the student appraisals with a trusted colleague or teaching consultant may increase "the degree of improvement" (Center for Teaching and Learning, University of North Carolina, 1994, p. 3) in teacher effectiveness. However, relying solely on student appraisals for making high-stakes decisions is unsound. "Student ratings are often misinterpreted, misused, and not accompanied by other information…" (Theall, 2002, p. 1).

That is not to say that student appraisals shouldn't be taken into account in evaluating a teacher. Students are in a unique position to report on teacher behaviors in the classroom and the frequency of the behaviors. They can report on the amount and difficulty of the work required for the

course, the usefulness of the materials used in the course, the effectiveness of the teacher's communication, the teacher's availability and helpfulness outside of class, what they learned and whether they were satisfied with the course (Theall, 2002). Written appropriately, student appraisals do not merely indicate teacher popularity, but point to what makes a teacher popular. Effective appraisals address teacher behaviors.

For student appraisals to be considered useful, most researchers suggest that:

- there should be no more than 25 items on the appraisal form. More than that and students could begin to experience fatigue (CTL, UNC, 1994, p. 2).
- student appraisals should be anonymous.
- the instructor should not be in the room when the appraisals are being done.
- the instructor should not see end-of-semester appraisals until the term is over and grades have been given.
- the facilitator should read the instructions <u>and</u> the rating scale to students and answer any questions students may have about the appraisal and what the scale means.
- the facilitator should point out open-ended comments sections and encourage students to answer these.
- the facilitator should make clear that the appraisals will be used to help the teacher improve the course.
- students should be given sufficient time to complete the form; 20 to 30 minutes is usually ample time.

Research suggests that a minimum of 10 students is best for the data to be truly useful and that if less than 75% of the students complete the appraisal, one must be very careful in interpreting the data (CTL, UNC, 1994). Teachers should also be given the opportunity to respond in writing to student appraisals.

Most importantly, data from student appraisals should be collected and summarized over time to get a broad view of students' reactions to instructor performance and to help promote teacher growth (Felder & Brent, 2004; CTL, UNC, 1994).

Teachers might also consider giving students mid-semester appraisals to help inform their teaching. This could be done formally by having students answer open-ended questions on a form or by simply asking students from time to time to write a one-minute response to questions such as "What did you learn today?" or "What are you confused or unclear about?" These types of formative assessment can aid the teacher in improving instruction immediately rather than waiting for the end of the semester to make changes.

Finally, student ratings should be accompanied by other information (other types of evaluation and evidence) for summative evaluation of instructors and for faculty review. They should not be the sole form of evaluation. They cannot and do not provide a full picture of instructor effectiveness, but they should be part of the picture.

Supervisor Evaluation

Evaluation by a supervisor is usually considered part of summative evaluation, and indeed it is, but it doesn't have to be limited to that role. Supervisors can act as mentors and participate in mentor coaching, as described above. Supervisors can conference with teachers about goals and meeting the standards. They can observe teachers presenting warm-up activities or other instructional activities rather than always observing a complete lesson. Any observation should, however, be accompanied by a pre-

and post-conference to establish the objectives and share what was learned.

Some teachers prefer to have supervisors visit their classes several times over the course of the year, not only to help their development as teachers, but so that the supervisor will get a fuller picture of the teacher in action over several occasions. This, in turn, allows the supervisor to "take an average."

It can't be emphasized enough that teachers need to know whether an evaluation procedure or tool is being used for summative or for formative evaluation. In both cases, teachers need to be familiar with the instruments being used. Formative evaluation feedback should be kept by the teacher for improvement and growth. Summative evaluation becomes part of a teacher's professional record. Teachers should be told where it will be kept, who has access to it, and how it will be used. They should get a written copy and have the opportunity for written comments or rebuttals.

Effective appraisals address teacher behaviors.

Service and Professional Development

Two important areas of teacher growth that are often included in an evaluation system are service to the institution or the profession and professional development activities. These affect both the teaching environment and teacher performance. Most teaching standards include continuing education and service in their criteria for a good reason: teachers who serve on committees, develop curriculum, work on special projects, present at conferences, and attend conferences and workshops stay current and are exposed to more ideas and possibilities to improve programs and their own teaching.

Reports on engagement in service and professional development can be part of formative evaluation for identifying goals and later submitted as a report for summative evaluation and faculty review. Including reports on service and professional development as part of the teacher evaluation system gives teachers the opportunity to demonstrate that they are actively engaged in honing their skills and in gaining a broader understanding of the profession or institution. For supervisors, including these important elements in the evaluation system brings the topic to the table for discussion, which can be especially important when working with teachers who don't or won't participate in continuing education opportunities.

Conclusion

When reviewing faculty for determining promotion, tenure, merit raises, awards, and dismissal, supervisors and review committees should consider supervisor summative evaluations, student end-of-term appraisals, reports on service to the institution and the profession, and reports on professional development.

Summative evaluation may also include evidence collected and voluntarily submitted by the teacher as indicators of the level at which the teacher has met the standards. This might include: sample lesson plans with annotations; checklists or written comments from post-observation conferencing with peers or mentors; self-recorded videos of their teaching; student self-evaluations of their own progress; student pre- and post-test scores; samples of student projects; etc. These multiple measures help the supervisor and review committee get a fuller picture of a teacher's effectiveness and level at which they are meeting the standards.

Teacher evaluation can be a very sensitive topic for teachers and program administrators alike. Evaluations need to be fair and relevant to both teachers and programs. By including clear standards, opportunities and tools for various types of evaluations and evaluators, multiple measures, and frequent and useful feedback, language programs can establish a productive evaluation system that supports learning and teaching and lays the groundwork for making good personnel decisions.

References

Beddes, S., Cutler, E., Manning, C. & Russell, B. (Nov. 2012). *Point of Encounter: Lessons from a peer coaching pilot group.* A presentation at ORTESOL Fall Conference, Portland, OR.

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy.* (3rd ed.). White Plains, NY: Longman.

Center for Teaching and Learning, University of North Carolina (CTL, U of NC). (Sept. 1994). *Student evaluation of teaching*. Number 16. Chapel Hill, N.C. Retrieved from *cfe.unc.edu/pdfs/FYC16.pdf*

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. & Rothstein, J. (2012). Evaluating teacher evaluation. *Education Week*. Retrieved from http://www. edweek.org/ew/articles/2012/03/01/kappan_hammond.html?tkn=WTUF Published Online: February 29, 2012.

Felder, R.& Brent, R. (2004). Random thoughts…How to evaluate teaching. *Chemical Engineering Education, 38* (3). pp. 200-202. Retrieved from http://www.ncsu.edu/felder-public/Columns/Teacheval.pdf

Hargreaves, A. & Fink, D. (2004, April). The seven principles of sustainable leadership. *Educational Leadership, 61* (7) pp. 8-13. Retrieved from http://www.ascd.org/publications/educationalleadership/apr04/vol61/num07/The-Seven-Principles-of-Sustainable-Leadership.aspx

Heneman, H.G. III, Milanowski, A., Kimball, S.M., & Odden, A. (2006). Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay. *CPRE Policy Briefs No RB-45*. Philadelphia, Pa.: Consortium for Policy Research in Education (as cited in New York State United Teachers' *Teacher Evaluation and Development Handbook;* see reference below).

Huston, T., & Weaver, C.L. (2008). Peer coaching: Professional development for experienced faculty. *Innovative Higher Education, 33*, (1), pp. 5-20.

Marzano, R.J. (Nov. 2012). Teacher Evaluation: What's fair? What's effective? The two purposes of teacher evaluation. *Educational Leadership 70* (3) pp. 14-19. Association for Supervision and Curriculum Development (ASCD). Retrieved from http://www.ascd.org/publications/educational-leadership/nov12/vol70/num03/The-Two-Purposes-of-Teacher-Evaluation.aspx

Mosher, R. & Purpel, D. (1972). *Supervision: The reluctant profession.* Boston: Houghton Mifflin.

New York State United Teachers (NYSUT). (2011). *Teacher evaluation and development (TED) handbook and workbook.* Retrieved from www.nysut.org/ted.

Rindler, B. (1994). *The attributes of teacher evaluation systems that promote teacher growth as perceived by teachers of intensive English programs.* Boston University Press, Boston, MA.

Theall, M. (2002). Student ratings: Myths vs. research evidence. *Focus on Faculty.* Brigham Young University Faculty Center. Retrieved from http://studentratings.byu.edu/info/faculty/myths.asp

Vidmar, D.J. (2006). Reflective peer coaching: Crafting collaborative self-assessment in teaching. *Science Direct. 20* (3) pp. 135-148.

Weimer, M. (Jan. 23, 2013). *Six steps to mak-*

ing positive changes to your teaching. *Faculty Focus.* Magna Publications. Retrieved from http://www.facultyfocus.com/articles/teaching-professor-blog/six-steps-to-making-positive-changes-in-your-teaching/

Zwart, R.C., Wubbels, T., Bergen, T., & Bolhuis, S. (2009). Which characteristics of a reciprocal peer coaching context affect teacher learning as perceived by teachers and their students? *Journal of Teacher Education., 60.* pp. 243-257.

Melinda Sayavedra has been teaching English language learners and training teachers for over 30 years in a variety of settings. She recently developed a comprehensive teacher evaluation program for beginning service ESOL teachers headed to Japan.