Research and Practice in Language Teacher Evaluation

David Hayes (Brock University)
Kyungsuk Chang (KICE)*
Chanbinn Imm (KICE)


This article aims to provide guiding principles for English language teacher evaluation through the review of recent accounts of research and practice. Teacher effectiveness needs to be contextually defined and measured. It is suggested that education providers need to work with all stakeholders in their systems both to define criteria for effective teaching and the means by which it may be assessed. As teacher evaluation is multifaceted, a range of data needs to be collected encompassing the views of all the stakeholders as well as the reflections of teachers themselves. Teacher evaluation should not be regarded as a summative product, the result of single evaluation measures conducted upon teachers, but should be thought of as a formative process in which teachers are actively engaged. Regarding evaluator training, it should not be assumed that those charged with evaluating teachers are, because of their positions, adequately trained to perform the evaluations. Evaluator training should be carried out before any evaluation is conducted under a new system. In order for the evaluation system to be meaningful and contribute to improvements in the education system as a whole, it should be seen primarily as a means of professional development for teachers as well as for anyone else involved in the process.

I. INTRODUCTION

A concern with teacher evaluation1 has existed since the inception of public systems of education in western countries in the nineteenth century but has become increasingly important in the educational landscape in recent years in response to changes in society at large (Arends, 2006). Concerns with the quality of the educational experience that students

---

* 1st author: David Hayes, Corresponding author: Kyungsuk Chang

1 Teacher evaluation should be thought of as coterminous with other descriptive labels used in the literature, such as ‘teacher appraisal’ or ‘teacher assessment.'
in school receive, the quest for continually improving standards and the demands to meet societal needs for a highly qualified, flexible workforce have often placed teachers in the frontline of public – often political – criticism which sees teachers as barriers to achieving educational reform and raising student achievement levels (Hershberg & Robertson-Kraft, 2009). Much of this criticism is ill-founded, based as it is on perceptions of declining standards in schools linked to teacher (in)effectiveness rather than on solid evidence regarding standards of achievement in schools and how this is causally linked to teachers’ skills and classroom performance. Clearly, basing development and change in a major area of education such as teacher evaluation on purely subjective perceptions of this kind would be very poor educational policy-making. Nevertheless, the notion that teachers and schools should be more “accountable” beyond the education profession for the public service that they provide has proved to be an important driver in calls for more direct evaluation of teachers’ performance in the classroom (Arends, 2006; Bartlett, 2000; Hershberg & Robertson-Kraft, 2009; Middlewood & Cardno, 2001; Milanowski & Heneman, 2001).

Within the “accountability” framework, it has long been suggested that a well-constructed and effectively administered teacher evaluation system can contribute, through the professional development of teachers, to improvements in classroom instruction and more generally to the quality of education in schools (Bollington, Hopkins, & West, 1990). Connected to this is the notion that one measure of a teacher’s effectiveness is a gain in student scores on standardized tests (Leigh, 2010). Stronge, Ward, Tucker and Hindman (2007) found that ‘effective teachers’ (i.e. those whose students achieved larger than expected learning gains) as opposed to ‘ineffective teachers’ (i.e. those whose students achieved lower than expected learning gains) exhibited differences in “(1) differentiation and complexity of instructional strategies, (2) questioning practices, and (3) level of disruptive student behavior” (p. 180). The differences were in a recognition that lessons needed to suit particular circumstances, in more extensive use of higher order questioning, and in the occurrence of much less disruptive behavior in the classroom. Yet, though this kind of argument is appealing to policy makers, there are inherent problems in causally attributing teacher performance to student test score gains. Kupermintz (2003), for example, notes that the argument is flawed:

... [B]ecause teacher effectiveness is defined and measured by the magnitude of student gains. In other words, differences in student learning determines – by definition – teacher effectiveness; a teacher whose students achieve larger gains is the “effective teacher”. (p. 289)

This simplistic notion of linking student gains with teacher effectiveness as a means of teacher evaluation must, therefore, be treated with extreme caution. However, if student
gains on tests are not valid indicators of teacher effectiveness, or at least not in and of themselves, then other criteria for evaluation must be sought. As more valid and reliable models are developed, amongst the key questions to be answered are:

- Who sets the standards?
- How are standards for effective teaching determined?
- Should teachers be expected to conform to a single set of standards?
- Are the assessments of a teacher’s performance made against standards fair?
- Do assessments of teachers actually improve instruction in classrooms?

It must also be recognized that systems of teacher evaluation are not in themselves truly objective, i.e., there is always subjectivity in the development of the system and, most particularly, in the process of evaluation of teachers performed against pre-determined standards within individual contexts of teaching (Woods, 1996). Both the validity and reliability of systems for teacher evaluation based on ratings by administrators have been questioned over the years (Danielson & McGreal, 2000; Weber & McBee, 1990).

Some writers go beyond notions of individual subjectivity and see teacher evaluation as inherently ideological – part of a political process. This process is generally held to consist of attempts by governments to exert more control over teacher’s work (Bartlett, 2000; Grace, 1985). Anderson (1998) indicates that “Integral to the process of evaluation and appraisal is the issue of power and authority: Who is appraising whom, and what power and authority do the participants have (p. 165)?” Bartlett (2000), in reviewing the history of teacher appraisal in the United Kingdom has no doubt where power lies, seeing a shift over the years in the way the process has been conducted as “a means of exerting greater control over teaching by assessing performance and linking it to pay” (p. 36). Dadds (2001) maintains that government education reforms in the UK which have been directed at raising standards in order to enhance global economic competitiveness have resulted in a situation in which “teacher professional judgment is dislodged by government in favor of centralist prescription of curricular and pedagogy” (p. 44). Similar views are to be found in Bennett (1999) who foretold that a “supportive, collegiate climate will be replaced by the threatening atmosphere of accountability and competition” (p. 424). Within such a view of education, teacher evaluation would seem to be useful only as a means of ensuring that teachers are adhering to the centrally prescribed pedagogy.
II. RESEARCH AND PRACTICE

1. Purposes and Processes of Teacher Evaluation

"Teacher evaluation – a term that brings fear, anticipation, stress, anxiety, or even boredom to the hearts of teachers and administrators everywhere" (Danielson & McGreal, 2000, endmatter).

As this comment indicates, evaluation (as it is typically conceived), when looked at from the perspective of the principal actors involved, is not regarded as a positive process that is central to a teacher’s professional development. This seems to be both a long-standing (Dadds, 1986) and a widespread experience, whatever the teaching context and often in contradiction of the ideals of the teacher evaluation policy. In their investigation in Botswana, Monyatsi, Steyn and Kamper (2006) concluded that “many teachers viewed the current teacher appraisal system as demoralizing, even threatening. Evidently, Botswana teacher appraisal policy ideals (non-disciplinary, accurate, open, based on proper training, part of continuous support and staff development …) are not sufficiently met in practice” (p. 225). Why should teacher evaluation be so contentious? Danielson and McGreal (2000) go on to comment that “Because of the factors described here – unclear or inappropriate evaluative criteria, limited administrative expertise, one-way communication – current teacher evaluation is a meaningless exercise. It yields little of value to either the teachers or the schools in which they work” (p. 7). But, given that teacher evaluation will likely remain part of the educational landscape for years to come, ways must obviously be found to make the systems and processes more meaningful to both teachers and administrators.

A starting point is to consider the purposes of evaluation. Inevitably, systems of teacher evaluation will be dependent on the purposes of the evaluation: Why does the evaluation need to be carried out and who is it for? Middlewood and Cardno (2001, p. 5) offer a ‘continuum of emphasis’ in teacher evaluation, as follows:

| Emphasis on assessing performance outcomes | Emphasis on teachers’ professional development |

Yet this continuum may also be seen to consist of irreconcilable tensions (Bartlett, 2000; Danielson & McGreal, 2000; Goddard & Emerson, 1992). Hannay, Telford and Seller (2003) note that “There is an acceptance that professional development needs to be long-term, reflective and self-embedded. However, performance appraisal strategies are frequently tied to contractual needs (Searfoss & Enz, 1996) with constrained time lines” (p.
A similar view is put forward by White (1997) who believes that “it is essential to achieve a balance between the requirements of accountability to stakeholders, on the one hand, and those of professionalism, on the other, in order to nurture and encourage the autonomy of the individual practitioner” (p. 138). How is it possible to achieve this balance, to reconcile long-term processes of professional development with short-term appraisal processes when, often, a single system is required to serve these varied purposes at different levels within the education system, as Table 1 (Middlewood & Cardno, 2001, p. 6) illustrates?

**TABLE 1**

| Multi-level Purposes of Appraisal System (Middlewood & Cardno, 2001, p. 6) |
|------------------|------------------|------------------|
| **Levels**       | **Purposes**     | **Purposes**     |
| EDUCATION SYSTEM | Accountability   | Development      |
|                  | School review and audit | Improving the quality of teaching |
| ORGANISATION     | Charter goals    | School improvement |
| INDIVIDUAL      | Management decisions | Performance improvement |
| Professional    | Professional responsibility | Self-reflection and improvement |
| INDIVIDUAL      | Personal         |                  |

Though this multi-level analysis emphasizes the “complexity inherent in appraisal system development and in the interpersonal dynamics that dominate the implementation of these systems” (Middlewood & Cardno, ibid), the complexity is not reflected in teacher evaluation systems. It remains true that by far the majority of teacher evaluation in schools “consists of a principal’s report of teacher performance, usually recorded on a checklist form, and sometimes accompanied by a brief meeting” (Peterson, 2000, p. 18). This seems to be the case for a variety of contexts worldwide. Murdoch (2000), reflecting on teacher evaluation specifically in ELT, notes that “Where it exists at all, it is based largely on irregular observations conducted by an over-burdened director of studies or senior teacher who does not have sufficient time to prepare for, and follow up on, classroom visits” (p. 54). Yet it is also an unfortunate truth that “Seventy years of research on principal ratings of teachers shows that they do not work well” (Peterson, 2000, p. 19). Evidence indicates that a single method of teacher evaluation cannot at once satisfy every teacher’s need for it to provide professional self-development for him or her as an individual as well as the education system’s requirement for accountability in terms of how teachers are performing which can then be fed into plans for improvement. If this is the case, arguments for more differentiated models of teacher evaluation would seem to have some validity. Campbell,
Kyriakides, Muijs and Robinson (2004) list “five potential dimensions of differentiation in teacher effectiveness”, viz.:

“The range of role activities expected of teachers in modernizing education systems is extremely broad. These include social, pastoral, welfare dimensions and leadership of other adults, and other work outside classrooms, in addition to the formal instructional dimension. Second, there is the issue of differentiated effectiveness across different subjects in the curriculum, or across different components. Third, teachers may be differentially effective in promoting the cognitive progress of different groups of students according to background variables. The principal ones are ability, age, sex, socio-economic status and ethnicity. Fourth, teachers may be differentially effective in promoting the learning of students according to the students’ personal characteristics, such as their personality, cognitive learning style, and extent of motivation and self-esteem. Fifth, teachers may be differentially effective in response to the different cultural and organizational contexts in which they work, such as the school culture, department structure and school size”. (pp. 9-10)

Within ELT specifically, there seems to be limited recognition of this range of differentiation. For example, Murdoch (2000, pp. 55-56) believes that teacher evaluation needs to have five aims, viz.:

- To encourage reflective practice
- To empower and motivate teachers
- To assess all aspects of a teacher’s professional activity
- To take account of students’ views
- To promote collaboration

However, his model focuses very much on the teacher’s pedagogic activity and does not take account of those teacher roles beyond instruction in the subject which are of such importance to writers on general education. Campbell, Kyriakides, Muijs and Robinson (2003) comment that under modernizing tendencies and as societies become more secular, schools become the main site of moral and social value formation. Under these trends the role of teachers is expected to be broad, affective, moral and welfare in orientation as well as cognitive. Further, teacher evaluation ought then to be founded on a model of teacher effectiveness which incorporate[s] measures of effectiveness across these different roles rather than as now limited to aspects of the cognitive.

The dichotomy between evaluation for accountability and evaluation for personal professional development also has an impact on teachers’ perceptions of evaluation processes and their resulting behavior. Petersen and Comeaux (1990) observe that from the
perspective of teachers, evaluation may be viewed as an opportunity for positive interaction and professional growth among colleagues or as a negative requirement that leads to embitterment. Alternatively, teachers may view standardized evaluation procedures as simply something to “pass”. Teachers may develop model lessons that are reserved only for formal evaluations, and their typical teaching may be unaffected by either the substance or the format of the teacher evaluation procedures that they experience as beginning and practicing teachers.

The notion that standardized teacher evaluation based on observation of lessons is not a reflection of typical practice, and thus unreliable, is supported by a comment from a Korean teacher in Hayes’ (2008) survey of in-service education that observation “is just for showing what we prepared for them” (p. 47). Another problem is that the results of such evaluations are rarely used for developmental purposes, serving only to fulfill a bureaucratic requirement that the evaluation has indeed taken place (Monyatsi, Steyn, & Kamper, 2006).

2. Measures of Teacher Effectiveness

Whatever the purpose of teacher evaluation, any system must be founded on what it means to be an effective teacher, i.e. there must be agreed evaluative criteria against which teachers’ performance can be measured (Arends, 2006; Danielson & McGreal, 2000). As Arends (2006) succinctly puts it, “Effective teacher assessment is dependent upon a knowledge base and clear definitions of what constitutes effective teaching” (p. 17). Further, if the system is to be workable, these clear definitions must be publicly available in (and beyond) the teaching profession and there must be clear statements regarding how they are to be evaluated – the process of evaluation – with precise articulation of the standards against which teachers’ performance is to be evaluated. Without these the system will not have validity in the eyes of the key participants in the process – teachers themselves.

Surprisingly, the validity of evaluation instruments and procedures amongst teachers is not something that those who design and implement the systems often have at the forefront of their thinking, even though teachers’ views have been found to be at odds with those of their appraisers or evaluators (Bartlett, 1998; Kyriacou, 1997). The situation does not seem to have changed since Peterson and Comeaux (1990) concluded that the views of teachers are generally absent from discussions on the development of teacher evaluation systems even though it is their ‘performance’ which is being evaluated. It is little wonder that many evaluation or appraisal systems leave teachers feeling confused as to their purpose(s) which, in turn, leads to lack of commitment in their implementation (Gratton, 2004).

Equally surprising is that there is little agreement on what constitutes effective teaching.
Campbell et al. (2003) note that in the past the "operational definition of effectiveness [has] usually [been] restricted to the teacher’s classroom instructional behavior and its association with cognitive outcomes" (p. 351). This is in spite of the fact that the complexities of teaching itself have been recognized and, with particular respect to English language teaching, newer methodologies such as Communicative Language Teaching (CLT) being seen to require fundamental changes to teachers’ traditional patterns of classroom behavior (Vanci-Osam & Aksit, 2000). An example of standardized criteria is those used in teacher evaluation for the Cambridge Certificate in English Language Teaching to Adults (CELTA). Thaine (2004) discusses his experience of using CELTA criteria in effect in 2000 which, at that time, were as follows:

- Demonstrate classroom teaching skills by:
  - Establishing rapport and developing motivation
  - Adjusting their own language to meet the level and needs of the learners
  - Giving clear instructions
  - Providing accurate and appropriate models of language
  - Focusing on appropriate specific language and/or skills
  - Conveying the meaning of new language with clear and appropriate context
  - Checking students’ understanding of the new language
  - Clarifying forms of language
  - Identifying errors and sensitively correcting students’ oral language
  - Identifying errors and sensitively correcting students’ written language
  - Monitoring and evaluating students’ progress

- Demonstrate an awareness of teaching and learning processes by:
  - Teaching a class with the sensitivity to the needs, interest and background of the group
  - Organizing the classroom to suit the learners and/or the activity
  - Setting up and managing pair, group, individual and plenary work
  - Adopting a teacher role appropriate to the stage of the lesson and the teaching context
  - Teaching in a way which helps to develop learner self-awareness and autonomy

As can be seen, these are indeed "restricted to the teacher’s classroom instructional behavior and its association with cognitive outcomes” as Campbell et al. (2003, p. 351) put it. Perhaps more importantly, Thaine (2004) found that there were significant issues with interpretation of the criteria by those evaluators who had to use them in assessment of teachers’ classroom practice. He found that “just under half of the assessment criteria are easy to interpret, and are therefore transparent to trainers. The remaining criteria were in
some way problematic [...]. There was either a lack of consistency of interpretation, or the criteria were perceived as being difficult to define and/or exemplify” (Thaine, 2004, p. 339). Elsewhere, Peterson (1987) found many years ago that administrator reports tended to low degrees of criticality, providing a narrow range in assessment. Thus, not only do assessments of this kind focus primarily on classroom instructional practice and neglect the teacher’s wider educational roles in moral and social value formation, they also prove difficult to operate with the degree of consistency and criticality that is essential if they are to be reliable indicators of teachers’ effectiveness.

3. Methods of Evaluation

From the foregoing discussion, we can reasonably conclude that it is likely that a single measure will not provide a valid and reliable indicator of teacher effectiveness. Peterson (1987) suggested many years ago that multiple measures may tap into different constructs of quality in teaching. Thus, differentiated models of teaching effectiveness inevitably imply differentiated models of teacher evaluation. However, typical models of teacher evaluation tend to concentrate on instructional behavior assessed through classroom observation of a teacher’s performance. In recent years, classroom observation models have emphasized ‘reflective’ orientations and are usually conducted within a ‘clinical supervision’ framework. Ali (2007, pp.17-18) provides an example of this with her 6-step process for a reflective teacher observation program (where ‘trainees’ = in-service teachers) which is designed to be ‘reflective’, ‘collaborative’ and ‘developmental’:

- Step 1: Analyze trainees’ observation needs
- Step 2: Prepare trainees for observation procedures
- Step 3: Trainees observe trainers teaching (focused questions)
- Step 4: Observers contact trainees’ school to build rapport
- Step 5: Observers observe trainees and trainees reflect on teaching
- Step 6: Summative assessment

The ‘teacher observation and evaluation criteria’ used are firmly focused on classroom performance as a teacher of a particular subject, with a limited sense of the teacher as an educator in a school system responsible for “moral and social value formation” (Campbell et al., 2003, p. 351). Ali’s (2007, pp. 24-25) criteria, drawn from a public school setting in the USA, are given in Table 2 below:
### TABLE 2

#### I. INSTRUCTIONAL SKILLS – The teacher demonstrates, in his or her performance, a competent level of knowledge and skill in designing and conducting an instructional experience.

Indicators:
- Writes and teaches to clear objectives – Utilizes principles of learning – Provides a variety of instructional experiences
- Uses appropriate instructional strategies for students, subjects, and goals – Monitors ongoing performance to adjust lessons
- Uses school’s goals and guide effectively – Demonstrates creativity in the teaching process

#### II. CLASSROOM MANAGEMENT AND ORGANIZATION – The teacher demonstrates, in his or her performance, a competent level of knowledge and skill in organizing the physical and human elements in the educational setting.

Indicators:
- Provides a classroom climate conducive to learning
- Provides a model in demeanor and appearance that does not detract from teaching effectiveness
- Assesses individual differences, provides appropriate student grouping and uses appropriate instructional resources to meet individual needs
- Involves students in planning and evaluating their own work where appropriate

#### III. STUDENT DISCIPLINE AND ATTENDANT PROBLEMS – The teacher demonstrates the ability to manage the non-instructional human dynamics in the educational setting.

Indicators:
- Communicates clearly established parameters – Recognizes conditions that lead to problems – Assists students toward self-discipline
- Responds reasonably to discipline problems – Effectively utilizes the assistance of administrators or support personnel

#### IV. KNOWLEDGE OF SUBJECT MATTER – The teacher demonstrates a depth and breadth of knowledge of theory and content in general education and subject matter specialization(s) appropriate to the grade level.

Indicators:
- Gives evidence of subject matter competency in area(s) to be taught
- Recognizes the relationship between one’s subject matter field and other disciplines or subjects
- Keeps abreast of new developments in the subject matter area

#### V. INTEREST IN TEACHING PUPILS – The teacher demonstrates an understanding of and commitment to each pupil, taking into account each individual’s unique background and characteristics. The teacher demonstrates enthusiasm for or enjoyment in work with pupils.

Indicators:
- Plans educational experiences based on students’ unique background and characteristics
- Enjoys working with students
- Provides prompt, meaningful communication among parents
VI. PREPARATION AND SCHOLARSHIP – The teacher exhibits, in his or her performance, evidence of having a theoretical background and knowledge of the principles and methods of teaching and a commitment to education as a profession. Indicators:
- Keeps abreast of current and effective emerging principles of teaching
- Contributes to school and professionalism
- Maintains professional rapport with colleagues, parents, and community

VII. EFFORT TOWARD IMPROVEMENT WHEN NEEDED – The teacher demonstrates an awareness of his or her limitations and strengths and demonstrates continued professional growth. Indicators:
- Participates in career development
- Utilizes self-evaluation as a tool for professional growth
- Responds constructively to recommendations

Though they make an attempt to show how they might be measured, these criteria suffer from having indicators of achievement which assume agreement as to what constitutes effective teaching but which, in reality, are open to multiple interpretations. Just what does it mean to use “schools’ goals and guide effectively”? Just exactly what is “a model in demeanor and appearance that does not detract from teaching effectiveness”? The element of subjectivity in interpretation of these indicators is nowhere acknowledged and, in practice, is liable to render judgments made according to them meaningless. A similar criticism could be made of a ‘Field Experience Feedback Form’ presented by Cheng and Tang (2008, p. 226) where users felt that if it was to be used for summative assessment (as intended) it was difficult to determine what constituted evidence of skill in the various domains of teaching as well as to assign grades and levels on the basis of such evidence as could be determined, as shown in Figure 1 below. This kind of observation instrument might be more suited to developmental, formative assessment rather than summative.

FIGURE 1
Field Experience Feedback Form (Cheng & Tang, 2008, p.226)

<table>
<thead>
<tr>
<th>Participant’s name:</th>
<th>Time:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programme year:</td>
<td>Class:</td>
</tr>
<tr>
<td>Date:</td>
<td>Subject:</td>
</tr>
<tr>
<td>School:</td>
<td>Topic:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain of teaching</th>
<th>Evidence</th>
<th>Target for improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. PROFESSIONAL ATTRIBUTES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Commitment and dedication to teaching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Passion for continuous learning</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Moving beyond a model restricted to observation Danielson and McGreal (2000, p. 53) link various domains of teacher effectiveness with sources of information which could be collected to provide evidence of an individual’s degree of effectiveness (see Table 3 below). Here there remains a focus on instructional behavior but there is also at least a recognition of the need to include information concerning other stakeholders.

### TABLE 3

**Sources of Information on Teacher Effectiveness (Danielson & McGreal, 2000, p. 53)**

<table>
<thead>
<tr>
<th>Component of the Framework</th>
<th>Sources of Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1: Planning and Preparation</strong></td>
<td></td>
</tr>
<tr>
<td>1a: Demonstrating Knowledge of Content and Pedagogy</td>
<td>Classroom observation, sample unit plan, sample lesson plan, interview, log</td>
</tr>
<tr>
<td>1b: Demonstrating Knowledge of Students</td>
<td>Interviews, sample lesson plan</td>
</tr>
<tr>
<td>1c: Selecting Instructional Goals</td>
<td>Sample unit plan, sample lesson plan, teaching artefact</td>
</tr>
<tr>
<td>1d: Demonstrating Knowledge of Resources</td>
<td>Sample unit plan, sample lesson plan</td>
</tr>
<tr>
<td>1e: Designing Coherent Instruction</td>
<td>Sample unit plan, sample lesson plan, teaching artefact</td>
</tr>
<tr>
<td>1f: Assessing Student Learning</td>
<td>Sample unit plan, sample lesson plan, teaching artefact</td>
</tr>
<tr>
<td><strong>Domain 2: The Classroom Environment</strong></td>
<td></td>
</tr>
<tr>
<td>2a: Creating an Environment of Respect and Rapport</td>
<td>Classroom observation; student surveys; parent surveys</td>
</tr>
<tr>
<td>2b: Establishing a Culture for Learning</td>
<td>Classroom observation; teaching artefact; samples of student work; student surveys; parent surveys</td>
</tr>
<tr>
<td>2c: Managing Classroom Procedures</td>
<td>Classroom observation; interview</td>
</tr>
<tr>
<td>2d: Managing Student Behavior</td>
<td>Classroom observation; interview; records of students sent to the office</td>
</tr>
<tr>
<td>2e: Organizing Physical Space</td>
<td>Classroom observation</td>
</tr>
</tbody>
</table>
Domain 3: Instruction
3a: Communicating Clearly and Accurately
3b: Using Questioning and Discussion Techniques
3c: Engaging Students in Learning
3d: Providing Feedback to Students
3e: Demonstrating Flexibility and Responsiveness

<table>
<thead>
<tr>
<th>Classroom observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom observation</td>
</tr>
<tr>
<td>Classroom observation; teaching artefact; samples of student work</td>
</tr>
<tr>
<td>Classroom observation; samples of student work</td>
</tr>
<tr>
<td>Classroom observation</td>
</tr>
</tbody>
</table>

Domain 4: Professional Responsibilities
4a: Reflecting on Teaching
4b: Maintaining Accurate Records
4c: Communicating with Families
4d: Contributing to the School and District
4e: Growing and Developing Professionally
4f: Showing Professionalism

<table>
<thead>
<tr>
<th>Interview, reflection form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance records, field trip records</td>
</tr>
<tr>
<td>Phone logs, letters to parents, back-to-school night handouts, parent survey, phone calls from parents</td>
</tr>
<tr>
<td>Logs of professional activities; copies of documents to which teacher has contributed, with explanation of role</td>
</tr>
<tr>
<td>Logs of professional goals and improved practice; copies of conference programs attended or at which presented</td>
</tr>
<tr>
<td>Interview, feedback from colleagues</td>
</tr>
</tbody>
</table>

It can be seen from this table that, once teacher evaluation concerns itself with anything other than observation of classroom behavior, the process entails a great deal of systematic data collection from a variety of sources on a continuing basis. The burden this places on whoever has to collect the data should not be underestimated and is likely to militate against implementation of a system based on such a wide range of information, no matter how desirable it might be as a means of reflecting the full range of perspectives on a teacher’s work.

Similar concerns regarding practicality may be voiced about portfolio evaluation. This method of evaluation is traditionally connected with the assessment of student learning (Hamp-Lyons & Condon, 2000; Lo, 2010; Nunes, 2004) and, under various guises such as ‘coursework’ or ‘controlled assessment’ may be a component of the final school-leaving examination system in some countries, such as the United Kingdom (see www.qca.org.uk). However, there have also been instances where portfolio evaluation has been applied to the evaluation of teachers and teacher learning. Ahellal (2002, p. 5) contends that Teacher Development Portfolios (TDP) are “one of the most effective ways to shift emphasis from subjective, directive and domineering teacher supervision to teacher development”, allowing teachers opportunities to reflect on their practices, achievements and areas where they recognize improvement is needed over time. Arends (2006, pp. 12-13) lists the components of a Teacher Portfolio in use in Connecticut, USA, as:
Materials used in instruction
Examples of student work
Videotapes of teaching
Written records and exhibits
Reflections that described the teachers’ thinking processes

Similarities can be seen with Danielson and McGreal’s ‘Sources of Information’ in Table 3 above. However, if the collection of data for the portfolio is seen as an integral part of a teacher’s work, which is carried out on a daily (or perhaps weekly) basis, then it may be less burdensome than if the data is collected at a single point in the academic year purely for the purposes of evaluation.

Whatever method of evaluation is chosen, a key element that needs to be considered in the system is the potential benefits it brings to the individual teacher (as well as to the evaluator). Middlewood and Cardno (2001) claim that evaluation:

“creates opportunity for dialogue about performance based on observation and reflection on practice [...]. Hence, the giving and receiving of feedback is fundamental ... [which] implies that the most important appraisal activity is interpersonal and not technical. When technical purposes overtake the improvement purposes we find that the appraisal processes may be honed down to become just a mechanism for check-listing that minimum criteria have been met”. (p. 11)

Nevertheless, improvement should not be thought of as a goal confined to the individual teacher. Teachers are part of institutions whose overall purpose is to provide a high-quality educational experience for their students. Thus, as Anderson (1998) maintains, for evaluation to be successful, “the establishment of a climate in which all members of the school are committed to ongoing evaluation of their activities and work, as a cornerstone of both individual and organizational improvement, is crucial” (p. 184). If teachers are the only sector of the educational workforce subject to evaluation, there is likely to be little commitment amongst them to the system.

We should also note in this section that there is an assumption with many systems of teacher evaluation that those who carry out the process – the evaluators – are qualified for this task, often through achieving a certain position in the system. This, of course, cannot be assumed but needs to be demonstrated. Murphy (1996) argues that training in evaluation is a neglected part of the process, one which not only requires time and patience but also, as one of his interviewees indicated “For formative evaluation to succeed it must be part of the school culture, which must infiltrate slowly, not just be imposed down on them” (p. 336). Changes in school culture are also likely to require much time and patience to achieve.
4. Contextual and Cultural Appropriacy

The foregoing discussion needs to be framed according to the constraints and opportunities of the context(s) in which the system of teacher evaluation is to be implemented. The phenomenon of “travelling reforms” or “best practices” that are transferred from one country to another has been well documented. For example, the interest in the Finnish education system since its success in topping the Program for International Student Assessment (PISA) tables bears witness to the notion that once a particular country has been deemed ‘successful’ in international rankings, then it becomes the object of intense study and countries elsewhere are subject to calls to emulate that model (Pusawiro, 2010; Sahberg, 2007; Simola, 2005). Burridge (2010) reports that more than 100 delegations from other countries visited Helsinki in 2009 in order to find the secrets of Finnish success in education. However, less well-documented is the effect on other education systems of innovations introduced outside their country of origin as a result of these visits.

“Travelling reforms” or “best practices” moving around the globe affect teacher evaluation systems as they do other areas of education and are rarely questioned (Murphy, 1996). One reason for this may be that there is a paucity of research on teacher evaluation in non-western settings. As Kelly, Ang, Chong, and Hu (2008) note: “Empirical research examining factors that determine the effectiveness of teacher appraisal in Asian settings is sparse, with much of the debate on the efficacy of different appraisal approaches limited to Western settings” (p. 40). Thus, for example, teacher appraisal in Hong Kong – like other aspects of the education system – has been intimately affected by western experience, in this case from the United Kingdom due to historical, colonial links (Mo, Conners & McCormick, 1998). Walker and Dimmock (2000) suggest that applying Western models of teacher appraisal to Hong Kong is problematic as “the enduring values, norms, and beliefs composing a society’s culture may influence the shape and operation of appraisal in schools” (p.156). They complain that “There appears to be a naïve belief among many policymakers and practitioners that policies and practices designed in one context can be unproblematically transported elsewhere” (ibid, p. 157). In other areas of the educational system, the experience of teachers in some non-western contexts with the implementation of Western English language teaching methodologies indicates that this kind of transportation is indeed often problematic: one size does not fit all (Holliday, 2005).
III. GUIDING PRINCIPLES FOR ENGLISH LANGUAGE TEACHER EVALUATION

Based on the preceding review and taking full account of the remarks made about contextual and cultural appropriacy, some guiding principles can be suggested for English language teacher evaluation within the scope of the current review. Firstly, English language teacher effectiveness needs to be contextually defined and measured. There is little agreement on what constitutes ‘effective’ English teaching; still less on how this should be measured. What is suitable for one educational context may not be suitable for another. Education providers need to work with all stakeholders in their systems both to define criteria for effective English teaching and the means by which it may be assessed.

Secondly, English language teacher evaluation is a multifaceted enterprise and, to reflect this, systems of evaluation must be equally multifaceted. The effectiveness of an English teacher is unlikely to be evaluated adequately with a single measure such as English language classroom observation. A range of data will need to be collected encompassing the views of all the stakeholders – students, parents, educational administrators – as well as the reflections of teachers themselves.

It should also be taken into account that English language teacher evaluation covers areas beyond the classroom. Given that up to fifty percent of a teacher’s work is comprised of non-teaching duties and given the importance of teachers and schools in moral and social value formation, an English teacher’s effectiveness is gauged by his/her performance not just inside the classroom but beyond it. This must be accounted for in a system of English teacher evaluation. Next, teacher commitment and confidence should not be neglected in English language teacher evaluation. For any system of evaluation to work effectively, it must have the confidence of the teachers that it is a fair and equitable system and they must be committed to its successful implementation. To this end, English teacher feedback on how a system is working needs to be collected and used to improve the system.

The fifth principle is that English language teacher evaluation should be considered as a process. It should not be regarded as a summative product, the result of single evaluation measures conducted upon teachers as passive objects of the evaluation, but should be thought of as a formative process in which teachers are actively engaged. Though summative assessment may be required in an education system as part of accountability, formative assessment which influences a teacher’s work in (and beyond) the classroom on a daily basis is likely to result in greater long-term improvements in the educational experience of students. Following from this, it is suggested that English language teachers take part in evaluation as partners. If evaluation is seen as a process, teachers must be active partners in the system from its inception, through its developmental phases to its
final form(s). This requires language teachers’ representatives to be included in working committees to develop evaluation systems and instruments and in monitoring committees to oversee the implementation of the system.

The seventh principle is that training for evaluators is necessary. It should not be assumed that those charged with the task of evaluating English language teachers are, because of their positions, adequately trained to perform the evaluations. Evaluator training should be carried out before any evaluation of teachers is conducted under a new system. Eighth, evaluation for development should be a basic consideration. Continuing professional development should not be thought of as a concept solely applicable to classroom teachers. In order for the system of evaluation to be meaningful and contribute to improvements in the education system as a whole, it should be seen primarily as a means of professional development for English teachers as well as for anyone else – such as the evaluators themselves – involved in the evaluation process.

Finally, evaluation needs to be contextually and culturally appropriate to the given situation where English teaching takes place. Language teacher evaluation processes and procedures derived from study of international experience should not be applied to contexts outside those of their creation without adequate analysis of their contextual and cultural relevance. Good practice from other contexts may need to be modified before it can be applied elsewhere.

REFERENCES


Applicable levels: teacher education  
Key words: teacher evaluation, teacher appraisal, teacher effectiveness, accountability, teacher observation, professional development, evaluator training
Imn, Chanbinn
Korea Institute of Curriculum & Evaluation,
Jungdong Building, 15-5 Jung-dong,
Jung-gu, Seoul 100-784, Korea
Tel: +82-(0)2-3704-3660/ H.P.: +82-(0)-10-9048-4887
Fax: +82-(0)2-3704-3954
E-mail: cbimm@kice.re.kr

Received in December, 2010
Reviewed in February, 2011
Revised version received in March, 2011