Developing High-Quality Teachers: teacher evaluation for improvement

Author(s): Janet Looney

Source: *European Journal of Education*, December 2011, Vol. 46, No. 4, On becoming a teacher: a lifelong process (December 2011), pp. 440-455

Published by: Wiley

Stable URL: http://www.jstor.com/stable/41343393

Developing High-Quality Teachers: teacher evaluation for improvement

Janet Looney

Introduction

Teaching and learning are at the core of educational practice, and as a significant body of research demonstrates, teacher quality is the most important school-level factor affecting student achievement. The seminal 1966 Coleman report 'Equality of Educational Opportunity' in the US found that teacher characteristics explained more variance in student achievement than any other school factor (Coleman, *et al.*, 1966). Hanushek (1992) found that students learning with the most effective teachers outperformed their peers who were learning with the least effective teachers by as a much as one grade level.

Students from low-income families may benefit the most from learning with very effective teachers (Nye *et al.*, 2004; Sanders & Rivers, 1996) and those with a succession of high-quality teachers may potentially achieve the same levels as their peers from high-income families (Motoko *et al.*, 2007). On the other hand, Sanders and Rivers (1996) found that the residual effects of learning with poor-quality teachers were long lasting, and that students assigned to more effective teachers in later grades were unable to compensate for earlier gaps.

As countries aim to ensure high achievement for *all* students, improving and sustaining the quality of the teacher workforce are thus vital policy priorities. But what is the best approach to achieving these goals? A number of researchers have found that teacher credentials are a poor predictor of quality (Goldhaber & Anthony, 2007; Rockoff, 2004). Moreover, teachers need to continuously update knowledge and skills over the course of their careers to reflect changes in curricula and new knowledge on effective teaching and learning and better meet diverse student needs.

Several studies have found that well-designed teacher evaluation systems, aligned with professional learning and development, can contribute to improvements in the quality of teaching and raise student achievement. At the same time, teachers in many countries report that individual teacher evaluations are not conducted systematically, that evaluators may not be trained or may use ineffective methods or tools, and that professional learning and development are often not aligned with their needs (OECD, 2005a; 2009a). While other modes of evaluation, such as national or regional student assessments or school-level evaluations are conducted more systematically and provide some information on teacher performance, these evaluations may not provide the timely feedback or the detailed information on classroom practices needed to shape improvements in teaching.

These different approaches to teacher evaluation are also shaped by the relative emphasis education systems place on the use of results for accountability (high stakes rewards and/or sanctions) and for improvement (professional development targeted to identified needs and support for innovation). There are real tensions between these dual goals for evaluation, and education systems need to find an appropriate balance.

This article provides an overview of research on teacher evaluation for improvement and suggests directions for policy and research to strengthen systems. It begins with a summary of the research on the characteristics of effective teachers, which may set the standards against which teachers are evaluated. It then outlines different approaches to teacher evaluation and development and the evidence of their impact on student learning. These approaches include formal performance appraisals, teacher peer evaluation and student ratings. School-level evaluations are also used to gauge teacher effectiveness and identify areas for improvement for all teachers in the school. They include external school inspection and internal school self-evaluation. Data from large-scale, national or regional student assessments are also used to identify areas for improvement. The article concludes with a discussion on broad directions for policy and highlights areas where more research is needed. While the research cited in this article can help shape policy, there is a need for more evidence on the impact of different forms of teacher evaluation and development. Given the importance of teacher quality for student learning, this should be a priority for both policy makers and educational researchers.

Defining Teacher Quality

Any system for teacher evaluation needs to be tied to a clear set of standards and competences. While there is currently no single, widely accepted definition of teacher quality, perhaps a reflection of the complexity of teaching and learning, research shows that the most effective teachers:

- Are intellectually able. Verbal skills are particularly important; students achieve more with teachers who perform well on tests of literacy and verbal ability (Gustafsson, 2003; Hanushek, 1989; Rice, 2003).
- Have good knowledge of the subject-area(s) and competences they are teaching as well as a broad repertoire of teaching methods and strategies to meet diverse student needs (Darling-Hammond, 2006; Monk, 1994).
- Develop positive relationships with their students and recognise the crucial role of motivation and emotions in learning (Hinton & Fischer, 2010; Immordino-Yang & Damasio, 2007). Empathy — the teachers' ability to understand the student's perspective, feelings, cultural background, challenges and needs — plays a strong role in supporting student learning. Teachers who show they care about students' learning and set challenging goals for learning are particularly effective (Bishop & Glynn, 1999; Cornelius-White, 2007; Hattie, 2009; Marshall & Wiliam, 2006). Teachers who communicate low expectations for their students, on the other hand, have a negative impact on student achievement (Bransford *et al.*, 1999; Rubie-Davis *et al.*, 2006; Rubie-Davis, 2007).
- Have strong classroom management skills, including clarity in presentation of ideas, well-structured lessons and appropriate pacing (Hattie, 2009). Effective teachers also have a good knowledge of typical learner misconceptions and patterns for progression in the subject(s) they are teaching (Bransford *et al.*, 1999; Pellegrino *et al.*, 1999).
- Are skilled assessors. Teachers use assessment 'formatively' to monitor students and provide timely and specific feedback on what they need to do to improve performance and meet learning goals. They also adapt teaching to better meet identified learning needs (OECD, 2005b).

- Work collaboratively with their peers to develop a positive school climate, to improve overall school performance, and to engage in mutual support and professional learning (OECD, 2001; O'Day, 2002; Seashore Louis *et al.*, 2010).

Teachers may develop expertise in each of these areas with experience. Effective evaluation systems will take this progression into account. Moreover, ongoing opportunities for individual and collective teacher development are particularly important if teachers and schools are to improve over time.

The Different Purposes of and Approaches to Teacher Evaluation

The different forms of teacher evaluation focus on individual teaching performance in classrooms (*teacher appraisal*; *teacher peer evaluation* and *student ratings*)[1]; the school context (*external school inspection* and *internal school self evaluation*[2]) and student outcomes (*national or regional student assessments*[3] and *value-added assessments*[4] to measure gains in learning over time) (See Box 1 below for definitions of these and other relevant terminology used in this article).

Box 1. Terminology

The key terms used to describe different approaches to teacher evaluation as used in this report are:

Teacher appraisal — Formal performance reviews, usually conducted by a school level supervisor, to judge individual teacher performance. The results of appraisals may be used formatively to identify specific needs for professional development, or summatively for decisions related to promotion, rewards or sanctions.

Teacher peer evaluation — Teachers working in the same subject area (in the same or a different school) may observe their peers and offer feedback and suggestions for improvement. They may use evaluation tools to guide the process. The results may be used formatively, with primary emphasis on providing feedback for improvement, or summatively, as a complement to formal appraisals.

Student ratings — Students are invited to complete evaluation forms, providing feedback on teacher effectiveness at the end of a term.

School evaluation (external and internal) — External evaluations (*e.g.* school inspections are targeted project evaluations) are conducted by an individual or team who are not part of the school staff. Internal evaluations (school self-evaluations) are conducted by project or school staff. The distinction is in regard to who conducts the evaluation and for what purposes (school accountability vs. improvement).

NB: Education specialists in English-speaking countries often distinguish between "assessment" and "evaluation". As used in this report, the term "assessment" refers to judgements of individual student performance "Evaluation" is used for judgements on the effectiveness of policies, schools and school system and/or specific learning programmes. It includes external school inspections and programme evaluations, and internal school self-evaluations.

> ***Large-scale national or regional assessments*** — These tests of student achievement have the primary purpose of evaluating the performance of the school system. They may also be used to hold schools accountable for student learning and/or to provide feedback to schools and teachers on their performance. The results may be published and used to guide school choice and create incentives for schools to improve performance. They do not have civil effect for students (*e.g.* graduation or university admissions).
>
> ***Value-added assessments*** — Valued-added measurements of student achievement refer to gains over a given year, which can be attributed to the contributions of the local education area, the school or individual teachers. These gains are the "value added". The approach is intended to show how educators promote student progress beyond the level predicted by the student's socio-economic status.

Most systems use some combination of these measurements to evaluate teacher performance (although value-added assessments have thus far only been implemented in the UK and the US). This is important, as no single measurement can capture the full range of teacher performance in different contexts or conditions, or the qualities that are important for effective teaching. Different assessment methods will also provide different kinds of information as to how teacher characteristics and instructional strategies are influencing learning and can help build the knowledge base on 'what works' (Baker, 2004; Herman, 2005; Abu-Alhija, 2007). Moreover, all assessments and evaluations contain some degree of error, and as Haertel and Lorié (2004) assert, can only provide 'an imperfect estimate of student performance'; multiple measurements may help to avoid the risk of incorrect decisions based on error (Koretz, 2005).

The tools and processes for each of these different forms of evaluation should also be valid and reliable, i.e. they should be appropriate for their intended purposes (valid) and evaluators' judgements should be consistent across repeated observations (reliable). However, at this point, there is little comparative information on the processes used to validate different tools or to train evaluators in different countries.

Alignment and Coherence within Broader Frameworks for Assessment and Evaluation

The effectiveness of teacher evaluation will also depend on its alignment and coherence within the overall educational assessment and evaluation framework. Indeed, the core logic of assessment and evaluation frameworks rests upon alignment of standards (for students as well as teachers), curriculum and assessment and evaluation (Gordon *et al.*, 2009). If systems are misaligned, it is impossible to draw valid conclusions about the success of teaching and learning or to develop effective strategies for improvement.

At the same time, systems should avoid too much alignment. Systems focused on tight alignment may undermine new and innovative approaches to teaching and learning. They will also need to recognise how teachers develop expertise over time, and in turn, how they contribute to their peers' development and learning.

The Impact of Teacher Evaluation and Development of Student Learning

Large-scale international surveys on teacher evaluation, as noted at the beginning of this article, indicate that in many systems teachers are not evaluated systematically and that professional development is rarely linked to specific needs targeted through the evaluation process (OECD, 2005a; 2009a). Nevertheless, smaller scale studies have shown that evaluation focused on improvement can help to raise student achievement.

Evaluations of Individual Teacher Performance

Teacher evaluation (or appraisal) is common across OECD countries. However, according to findings of the OECD's (2005a)[5] *Teachers Matter* report on attracting, retaining and developing teachers, individual teacher evaluations are not necessarily conducted systematically. The study found that teacher evaluation was conducted on a regular basis in approximately half of the 25 participating countries. School leaders often did not have appropriate or valid tools for evaluation and did not spend the necessary time observing classroom practices to gain any real insight on teacher effectiveness. Many teachers interviewed for the study also noted that they did not receive systematic feedback from evaluators or guidance regarding professional development (OECD, 2005a).

According to the OECD's (2009a) Teaching and Learning International Survey (TALIS)[6], 13% of teachers across the 23 participating countries reported that they did not receive any appraisal or feedback. Nearly half the teachers responding to TALIS also indicated that they did not believe school leaders used effective methods to evaluate their performance and 75% said that they would not receive any recognition for improvements in their performance (OECD, 2009a). Despite these concerns, teachers responding to TALIS indicated that they appreciated feedback from school leaders. They also noted that they were more likely to focus on priorities for instruction as highlighted in evaluations. At the same time, most systems do not set priorities, tending to put equal weight on different areas evaluated (with the exception of priorities related to special learning needs and teaching in multicultural settings). Most teachers responding to TALIS viewed individual and collaborative research, informal dialogue to improve teaching and qualification programmes as having the greatest impact on their development. These survey results, however, do not reveal the impact of different development approaches on student learning (OECD, 2009a).

Other smaller scale studies found that the effectiveness of formative teacher evaluation and development depended in large part on the manner in which feedback was given and on whether teachers had opportunities to discuss teaching methods, and were able to take on new approaches over time. For instance, timely and specific feedback appears to be important in the evaluation as well as in training and professional development. Freiberg, Waxman, and Houston (1987) found that student teachers participating in a controlled experiment felt that feedback from supervisors and peers was most effective when it was followed by seminars where they could discuss teaching performance and ways to improve instructional strategies with peers. Wade (1985) found that the most effective training involved classroom observation, microteaching, video/audio feedback on performance, and practice.

Timperley and colleagues (2007) found that professional development was most effective when it challenged teachers regarding their conceptions of student capabilities or focused on approaches to teaching specific parts of the curriculum; occurred over time; focused on methods to improve student outcomes; engaged teachers in a professional community of practice; and when school leaders supported teachers' opportunities to learn and to process new information, and then offered relevant expertise.

A few studies have focused on the effectiveness of **teacher peer evaluation**. Goldstein (2002), in an examination of a programme known as the 'Peer Assistance Review' in a large urban school district in the US, found that teachers were generally positive about the process, but also wanted school leaders to maintain their central role in evaluation. Munson (1998) found that teachers believed that peers could provide constructive feedback on their teaching, and that it was also important for building collegiality. However, Kumrow and Dahlen (2002) reported that they were unable to identify any studies showing a relationship between teacher peer review and better student performance.

Box 2. Tools for teacher evaluation

It is important that tools and processes for teacher evaluation, whether conducted by the school leader or peers, are valid and reliable and that criteria for evaluation are linked to clear standards on teaching. This is particularly true for summative evaluations that have an impact on the teachers' record and career prospects. (Validity of teacher evaluation means that the instruments and processes for evaluation meet their intended purposes. Reliability means that evaluators' judgements are consistent across repeated observations).

Classroom observation tools are usually based on checklists or rubrics setting out parameters for teacher performance in different contexts. However, as Haertel (1991) points out, the context specificity of teaching means that it is difficult to achieve reliability of judgements for different teachers working in different settings. Ultimately, evaluations that are based on classroom observations will rely on the evaluators' professional judgement. Instruments for classroom observations should be tested for validity, and evaluators should be trained.

Student ratings of teacher performance are relatively rare (currently only in Mexico, the Slovak Republic, Spain and Sweden) and their use is not compulsory (Isoré, 2009). No English-language studies of their impact on teaching were identified in the writing of this article. However, there is evidence that younger students may provide effective feedback on the quality of their learning experiences. For example, The Middle East Voice of Children survey, implemented from April to June 2006 in Jordan, Lebanon and Palestine, which surveyed approximately 1,500 15- and 16-year-old students, found that students valued teachers who nurtured their curiosity, helped them to develop their thinking skills and encouraged their active participation in class. They were concerned with the quality of their relationships with teachers, expressing their dissatisfaction with

rote learning and teachers who berated them. These younger students were eager to learn and appreciated their teachers' support and encouragement (Awartani *et al.*, 2007).

Studies on student ratings conducted in university settings found that students provided accurate feedback on the quality of their teachers (Marsh, 2007). As with younger students, the qualities they value are in line with those practices identified as important for improving student learning. For example, students rate those teachers who have high expectations and strong subject knowledge more favourably. Cohen (1980; 1981) found that student feedback had a medium impact on the improvement of teaching at the college level. Hattie (2009) notes that students generally provide accurate feedback regardless of the quality of the evaluation tool used. This may be because they do not feel constrained by the tool. They observe teacher practices on a daily basis, and are clearly well placed to comment on the teacher's impact on their learning.

School Level Evaluations (External and Internal)

Several OECD countries require both external (school inspections, which typically are conducted tri-annually) and internal school self evaluations (typically conducted annually) Although appraisal of individual teacher performance is beyond their scope, these school-level evaluations include classroom observations and review of other school-level data. Evaluators have opportunities to identify different strengths and areas for improvement across a school faculty.

There is some evidence that school-level inspections may contribute to improvements in teacher and student performance. Luginbuhl and colleagues (2007), in a study conducted for the Dutch Central Planning Agency found that in the first two years following an inspection, student performance improved by 2% to 3% of a test score's standard deviation. More intensive inspections produced greater improvements in school performance. Also working in the Netherlands, DeWolf and Janssens (2007) found that the visits did appear to lead to improvements. However, they also conjecture that this may be in part a result of schools' strategic behaviour (*e.g.* reshaping the test pool, 'indicator fixation', or fraud). Grubb (2000) argues that teachers' positive and negative perceptions of the inspection process will also shape perceptions regarding effectiveness.

Most studies on school self-evaluation focused on the process of the evaluation itself rather than on impact on teaching and student outcomes (Blok *et al.*, 2007). In a large-scale survey of Dutch teachers, Hofman and colleagues (2009) went further than most studies to examine processes and improvements in teaching. They found a significant relationship between the quality of teaching and learning processes and school self evaluation. However, the study does not provide information on changes in student achievement related to these improvements.

A number of OECD countries now implement **large-scale national or regional examinations** to monitor student progress toward central standards. They include Australia, Belgium (French and Flemish communities), Canada, England, Estonia, Finland, France, Hungary, Italy, Korea, Luxembourg, Mexico, Norway, Scotland, Spain, Sweden, Turkey and the US. Students in selected year levels (*e.g.* years 4 and 9) are tested in priority subject areas, such as language, mathematics and science. The data are generally not used to evaluate individual teacher performance, but rather for school-level accountability, to guide school choice, and to identify areas for school improvement.

These large-scale assessments do not capture teaching processes within classrooms, but the focus on student outcomes is nevertheless important. Data from these assessments help school leaders decide how to direct resources (time, money and personnel). Teachers may collaborate to improve instructional strategies in areas where students perform less well.

Several studies from the US have found that standards-based approaches have been highly effective in focusing attention of schools and teachers on priorities for student learning (Herman & Baker, 2009). At the same time, external assessments may lead to a narrowing of curriculum, particularly in high stakes environments (*e.g.* rewards or sanctions or publication of assessment results)(Koretz, 2005; Popham, 2002).

Value-added assessments are currently implemented only in the UK and US, although there are regional and pilot initiatives in other OECD countries. The approach, which relies upon complex statistical methods to isolate the contributions of specific teachers, schools or local education areas, can be useful for identifying effective practices (OECD, 2008c). They are intended to address concerns that student assessments do not take into account prior achievement levels or student characteristics, such as their socio-economic background. Attention to students' prior achievement also contributes to a better understanding of student performance over time (Doran & Izumi, 2004).

Well-designed value-added assessments can provide more accurate quantitative information on school and teacher performance and target areas for school and system-wide improvement. They may also facilitate learning across schools or systems that have very different results for similar populations of students (Baker *et al.*, 2010; OECD, 2008c).

At the same time, there are concerns that value-added models weaken policies to promote equity of student outcomes. A potential message of these models is that teachers should not be held to account when students with lower levels of prior attainment do not meet standards. Factors beyond school, including students' socio-economic background and the effects of poor teachers in earlier years, certainly do have a powerful impact on student learning. But as noted earlier, lowered expectations for student performance also have an important negative impact on learning. Polices promoting value-added analyses need to balance these issues.

Baker and colleagues (2010) have argued that student test scores, including those that take teacher value-added into account, should not be used as the primary measure of teacher performance. They note that value-added estimates are unstable across models, years, and the classes teachers teach. A teacher identified as ineffective one year could have significantly different results the following year. They conclude that value-added assessments are indeed valuable, but should supplement evaluations conducted by competent supervisors and peers.

The Role of Incentives

As noted earlier, countries tend to emphasise either accountability or the improvement functions of assessment and evaluation. Certainly, both approaches are focused on improvement, but they reflect fundamentally different ideas as to how to promote change.

In systems emphasising accountability, teacher evaluation may include high-stakes incentives such as performance-based advancement, salary increases, bonus

pay, or alternatively the threat of teacher job loss or school sanctions for under-performance. These material incentives, it is believed, will motivate teachers to improve teaching and raise student outcomes. Researchers found that teachers also behaved as if assessments had high stakes when the results were published (even if there are no specific sanctions attached), as teachers worked to avoid the stigma of low rating (Corbett & Wilson, 1991; Madaus, 1988; McDonnell & Choisser, 1997)[7].

The empirical evidence on the impact of high-stakes assessments on classroom instruction is mixed, although a number of studies report neutral or negative effects[8]. Santiago and Benavides (2009) argue that teacher evaluation systems that emphasise accountability may be counterproductive to improvement. Teachers are less likely to reveal any weaknesses in their performance, and therefore miss opportunities for feedback or professional learning and development. Baker and colleagues (2010) have argued that competition for monetary incentives may weaken peer learning.

High stakes student assessments may also create incentives to 'teach to the test'. Teachers may significantly narrow learning if they focus only on areas most likely to be on the test, as no single test will measure the full range of knowledge, skills and competences set out in central standards or competences. Teachers may use strategies such as coaching, (focusing on test taking strategies and tricks), realignment (focusing on the content and kinds of problems most likely to appear in the test), and/or reallocation (of time spent on higher priority subjects). These strategies may lead to 'score inflation' (overstatement of students' actual learning), and undermine efforts to develop appropriate strategies for improvement (Koretz, 2005; Looney, 2009).

In systems emphasising improvement, information gathered in evaluation processes is used to identify teacher strengths and weaknesses and appropriate professional development opportunities. Incentives are based on teachers' desire to help students succeed (purposive incentives) and to be respected in their professional community (solidary incentives).

There is growing evidence that teacher collaboration and peer learning have a strong impact on improved student achievement over time (Jackson & Bruegmann, 2009). In a large-scale longitudinal study conducted in the US, Seashore Louis and colleagues (2010) also found that collective leadership at both the school and district levels was associated with stronger impacts on student achievement. At the school level, collective leadership focused on instructional improvement had a significant positive impact on teachers' working relationships and on student achievement.

In another US-based study, Finnigan and Gross (2007) conducted a survey of teachers in ten low-performing schools in Chicago which had been placed on probation in order to learn more about how they responded to different incentives. They found that rewards and sanctions had less influence on teachers than their desire to help students or to maintain their status in their professional community. The survey data also showed that accountability policies with high stakes might have a counterproductive effect on teacher motivation over time, particularly in schools that continue to struggle. The Consortium for Policy Research in Education (CPRE), also working in Chicago, found that among low-performing schools that had been placed on probation, those that had previously developed strong cultures of peer collaboration were able to exit

probationary status relatively rapidly (DeBray, Parson, & Woodworth, 2001; Elmore, 2001).

These findings point to a number of unintended consequences related to higher stakes assessment and evaluation. However, the message here is not that systems emphasising accountability are bad and those emphasising improvement are good. Accountability is an extremely important goal for education systems. Rather, care needs to be taken to balance different approaches to ensure that systems incorporate multiple measures so that judgements are based on strong evidence, to distinguish evaluations intended for formative versus summative purposes, and to ensure that teachers receive appropriate support for improvement.

Directions for Policy

While the evidence based on whether and how teacher evaluation can lead to improvements in teacher quality and student outcomes is still relatively underdeveloped, a few broad directions for policy are clear:

First, teacher evaluations need to tie to clear standards and competences. These standards set the bar for the teacher workforce. Teacher evaluation, as a formative process, will be most effective when it measures current teacher performance against defined standards and competences for high quality teaching, and then identifies strategies and development needs to help teachers meet standards.

Second, teacher evaluations need to be integrated with broader assessment and evaluation frameworks. As part of this framework, standards and competences for teacher performance and systems of measurement need to align with standards and competences for student learning. It is also important to consider how systems balance goals for accountability and improvement, and ensure that teachers' professional development is aligned with broader goals for school improvement.

Third, teacher evaluations should be based upon multiple measurements. Evaluations by competent peers and supervisors, supplemented by student ratings, school-level evaluations and measures of student outcomes can provide a more complete picture of performance. These may also include opportunities to observe less easily quantifiable characteristics, such as relationships with students, how teachers communicate their expectations for student performance, and how they guide formation of values.

Multiple forms of measurement also provide more information regarding performance and can guide better decision-making regarding strategies for improvement. At the same time, policy makers will need to determine how much redundancy systems should include, how different assessments and evaluations should be weighted, how these measures take into account the different characteristics of high quality teachers, and how to design systems with complementary measures (Baker, 2004).

Fourth, evaluations should emphasise timely feedback linked to specific ideas for improving instructional strategies and opportunities to practice. The most effective teacher evaluation and development will, when appropriate, challenge teacher conceptions regarding student learning and capabilities.

Fifth, professional development should align with identified needs for development and encourage the development of communities of practice within and among schools. Large-scale student assessments and school inspections and evaluations help to identify key areas where local education authorities and schools need to improve practice. School improvement efforts focused on building instructional capacity

not only examine practices in individual classrooms, but also allow teachers to move out of their individual classrooms to engage in collective efforts to improve student achievement. Peer learning may also extend well beyond individual schools to encompass learning across a local education authority and beyond[9].

Sixth, evaluation for improvement (formative) should align with evaluations with summative impact (for career advancement). Teacher evaluation aimed at improvement should also support teachers' career advancement and progress towards greater expertise and ability to mentor more junior teachers at later points in their careers. This can support efforts to improve the quality of the teacher workforce.

A Research Agenda

Given the importance of teacher quality for student achievement, efforts to identify the most effective approaches to teacher evaluation and development are vital. Priorities for future research include:

- *The impact of different incentives on teacher motivation and improvement.* There is a need for further research on how teachers respond to different incentives — for financial rewards, professional status and the desire to help students succeed. This research may also help systems as they balance goals for accountability and for improvement.
- *Validated tools and processes for teacher evaluation (and training for evaluators).* Systems will need to invest in the design and evaluation of more effective evaluation tools as well as training for evaluators. Teacher evaluation will have little impact if these processes do not capture important elements of quality teaching, or consider the impact of different approaches on student learning.
- *How teachers progress from novice to expert status.* A better understanding of how teachers develop expertise over time can help shape professional development programmes, as well development of effective career ladders and reward systems.
- *Effective approaches to collective teacher learning and development.* Research in this area should examine how teachers working in different settings (urban vs. rural, low vs. higher socio-economic status, younger vs. older students) build effective learning communities, focused on improving outcomes for all students.

Effective teacher evaluation and development can potentially play a vital role in improving the quality of teacher workforce, but there is still much work to do, at the conceptual and empirical levels, and in designing effective strategies for policy and implementation.

Janet Looney, 14, rue de l'Ecole Polytechnique, 75005 Paris, France, janetlooney1@yahoo.fr

NOTES

1. Student ratings are used at the school level in Mexico, the Slovak Republic, Spain or Sweden (Isoré, 2009). In most OECD countries, are more commonly used at the university level.
2. Ten countries reported to the OECD's *Education at a Glance* survey (2009b) that they require external school inspection as well as internal school self-evaluation (usually on an annual basis) (the States and Territories of Australia, the Czech Republic, England, Iceland, Korea, New Zealand, Portugal, Scotland, Sweden and Turkey) (UNESCO, 2006/07).

3. More than half of OECD countries monitor student performance through periodic national or regional assessments [Australia, Belgium (French and Flemish communities), Canada, England, Estonia, Finland, France, Hungary, Italy, Korea, Luxembourg, Mexico, Norway, Scotland, Spain, Sweden, Turkey, and the US]. In Germany, the Länder recently cooperated to develop external examinations to provide comparable information on student performance (UNESCO, 2006/07).

4. Currently, valued-added assessments are implemented only in the UK and the US. However, as an indication of the growing interest in this approach, 13 countries and regions took part in the 2008 OECD study on best practices to assess the value-added of schools. They were: Australia, Belgium (Flemish Community), Czech Republic, Denmark, France, Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden and the UK.

5. The 2008 TALIS surveyed teachers and principals in lower secondary schools in 16 OECD countries and 7 partner countries. They included: Australia, Austria, Belgium, Brazil, Bulgaria, Denmark, Estonia, Hungary, Iceland, Ireland, Italy, Republic of Korea, Lithuania, Malta, Malaysia, Mexico, Norway, Poland, Portugal, Spain, Slovak Republic, Slovenia and Turkey.

6. The study included 20 student teachers working in secondary schools. They were randomly assigned to three groups. Supervisors and peers in the first group gave feedback only. Teachers in a second group received written feedback based on three classroom observations, followed by a one-hour seminar discussing their profile. They did not have an opportunity to discuss ways to improve their instructional strategies.

7. At least 18 OECD countries publish the results of external assessments and/or evaluations (inspections and/or school self-evaluations). They include the Czech Republic, Denmark, England, France, Hungary, Iceland, Korea, the Netherlands, New Zealand, Norway, Portugal, Scotland, Slovenia, Sweden and Turkey. Australia, Belgium (Flemish Community) Ireland and Italy also publish results, but avoid the use of tables that compare school performance. Results are not published in Finland, Mexico and Luxembourg.

8. In the US, two macro-level studies on results of the National Assessment of Educational Progress (NAEP) came to very different conclusions. Linn (2000) compared NAEP with state level assessments and found no clear trends. Hanushek and Raymond (2005) found gains in student performance on the NAEP (an effect size of 0.2 standard deviations). The gains were only in states with high stakes for student performance on state-level assessments, which Hanushek and Raymond claim as support for the role of stakes in improving student achievement.

9. Internet enabled networks provide opportunities for teachers to share best practices across countries and continents. For example, iNet, based in the UK, focuses on effective practices related to personalising student learning. The network extends well beyond the UK, to include networks on four continents (www.sst-inet.net/countrynetworks.aspx).

REFERENCES

ABU-ALHIJA, F. N. (2007) Large scale testing: benefits and pitfalls, *Studies in Educational Evaluation*, 33, pp. 50–68.

AWARTANI, M., WHITMAN, C. V. & GORDON, J. (2007) *The Voice of Children: student well-being and the school environment* (Paris, Brussels and US, Universal Education Foundation). www.uef-l4wb.org/wp-content/uploads/2010/12/VoCMiddleEast-Pilot_Preliminary-SurveyResults_FullReport.pdf.

BAKER, E. L. (2004) *Aligning curriculum, standards, and assessments: fulfilling the promise of school reform.* (Los Angeles, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)).

BAKER, E. *et al.* (2010) *Problems with the Use of Student Test Scores to Evaluate Teachers.* EPI Briefing Paper, 278 (Washington, DC, Economic Policy Institute).

BISHOP, R. & GLYNN, T. (1999) *Culture Counts: Changing Power Relations in Education* (Palmerston North, New Zealand, Dunmore Press).

BLOK, H., SLEEGERS, P. & KARSTEN, S. (2007) Looking for a balance between internal and external evaluation of school quality: Evaluation of the SVI model, *Journal of Education Policy*, 23, pp. 379–395.

BRANSFORD, J. D., BROWN, A. L. & COCKING, R. R. (Eds) (1999) *How People Learn: brain, mind, experience, and school* (Washington, DC, National Academy Press).

COLEMAN, J. S. *et al.* (1966) *Equality of Educational Opportunity* (Washington, DC, National Center for Educational Statistics).

COHEN, P. A. (1980) Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings, *Research in Higher Education*, 13, pp. 321–341.

COHEN, P. A. (1981) Student ratings of instruction and student achievement: a meta-analysis of multisection validity studies, *Review of Educational Research*, 51, pp. 281–309.

CORBETT, H. D. & WILSON, B. L. (1991) Two state minimum competency testing programs and their effects on curriculum and instruction, in: R. E. STAKE (Ed) *Advances in Program Evaluation: Vol. I., Effects of Mandated Assessment on Teaching* (Greenwich, CT, JAI Press, pp. 7–40).

CORNELIUS-WHITE, J. (2007) Learner-centered teacher-student relationships are effective: A meta-analysis, *Review of Educational Research*, 77, pp. 113–143.

DARLING-HAMMOND, L. (2006) *Powerful Teacher Education: Lessons from Exemplary Programs* (San Francisco, Josey-Bass).

DEBRAY, E., PARSON, G. & WOODWORTH, K. (2001) Patterns of response in four high schools under state accountability policies in Vermont and New York, in: S. FUHRMAN (Ed) *From the Capitol to the Classroom: standards-based reform in the states* (pp. 170–192). (Chicago, University of Chicago Press).

DORAN, H. C. & IZUMI, L. T. (2004) *Putting Education to the Test: A Value-Added Model for California* (San Francisco, Pacific Research Institute).

ELMORE, R. (2001) *Psychiatrists and light bulbs: educational accountability and the problem of capacity.* Paper presented at the annual meeting of the American Educational Research Association, Seattle.

FREIBERG, H. J., H. C. WAXMAN & W. R. HOUSTON (1987) Enriching feedback to student teachers through small group discussion, *Teacher Education Quarterly*, 14, pp. 71–82.

FINNIGAN, K. & GROSS, B. (2007) Do accountability sanctions influence teacher motivation? Lessons from Chicago's low-performing schools, *American Educational Research Journal*, 44, pp. 594–629.

GOLDHABER, D. & ANTHONY, E. (2007) Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching, *Review of Economics and Statistics*, 89, pp. 134–150.

GOLDSTEIN, J. (2002) *Results from the implementation of teacher peer assistance and review in California: A case study of one urban school district*. Paper presented at the annual meeting of the American Education Research Association (New Orleans).

GORDON, J., HALASZ, G., KRAWCZYK, M., LENEY, T., MICHEL, A., PEPPER, D, PUTKIEWICZ, E. & WISNIEWSKI, J. (2009) *Key Competences in Europe: Opening doors for lifelong learners across the school curriculum and teacher education (2009) Study undertaken for the Directorate General Education and Culture of the European Commission* (Warsaw, CASE-Center for Social and Economic Research). (http://ec.europa.eu/education/more-information/moreinformation139_en.htm)

GRUBB, N. (2000) Opening classrooms and improving teaching: Lessons in school inspections in England, *Teachers College Record*, 102, pp. 696–723.

GUSTAFSSON, J. E. (2003) What do we know about effects of school resources on educational results? *Swedish Economic Policy Review*, 10, pp. 77–110.

HAERTEL, E. H. (1991) New forms of teacher assessment, *Review of Research in Education*, 17, pp. 3–29.

HAERTEL, E. H. & LORIÉ, W. A. (2004) Validating standards-based test score interpretations, *Measurement: Interdisciplinary Research and Perspectives*, 2, pp. 61–103.

HANUSHEK, E. A. (1989) The impact of differential expenditures on school performance, *Educational Researcher*, 18, pp. 45–62.

HANUSHEK, E. (1992) The trade-off between child quantity and quality, *Journal of Political Economy*, 100, pp. 84–117.

HANUSHEK, E. A. & RAYMOND, M. E. (2005) Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24, pp. 297–327.

HATTIE, J. (2009) *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (New York, Routledge).

HERMAN, J. L. (2005) *Making Accountability Work to Improve Student Learning* (Los Angeles, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)).

HERMAN, J. L. & BAKER, E. L. (2009) Assessment policy: making sense of the Babel, in: G. SYKES, B. SCHNEIDER & D. N. PLANK (Eds) *Handbook of Education Policy Research* (London and New York, American Educational Research Association).

HINTON, C. & FISCHER, K. W. (2010) Learning from the developmental and biological perspective, in: H. DUMONT, D. ISTANCE & F. BENAVIDES (Eds) *The Nature of Learning. Using Research to Inspire Practice* (Paris, OECD).

HOFMAN, R. H., KIKSTRA, N. J. & HOFMAN, W. H. A. (2009) School self-evaluation and student achievement, *School Effectiveness and School Improvement*, 20, pp. 47–68.

IMMORDINO-YANG, M. H. & DAMASIO, A. (2007) We feel therefore we learn: The relevance of affective and social neuroscience to education, *Mind, Brain and Education*, 1, pp. 3–10.

ISORÉ, M. (2009) *Teacher Evaluation: current practices in OECD countries and a literature review.* OECD Education Working Paper, 23 (Paris, OECD).

JACKSON, C. K. & BRUEGMANN, E. (2009) *Teaching Students and Teaching Each Other: the importance of peer learning for teachers.* NBER Working Paper, 15202 (Cambridge, MA, National Bureau of Economic Research).

KORETZ, D. (2005) Alignment, high stakes, and the inflation of test scores, CSE Report, 655 (Los Angeles, University of California, National Center for Research on Evaluation, Standards, and Student Testing).

KUMROW, D. & DAHLEN, B. (2002) Is peer review an effective approach for evaluating teachers? *The Clearing House,* 75, p. 238–241.

LINN, R. L. (2000) Assessments and accountability, *Educational Researcher,* 29, pp. 4–16.

LOONEY, J. (2009) *Assessment and innovation in education.* OECD Education Working Papers, No. 24. (Paris, OECD). doi: 10.1787/222814543073

LUGINBUHL, R., WEBBINK, D. & DE WOLF, I. (2007) *Measuring the effect of school inspections on primary school performance: A study based on CITO test scores.* CPB discussion paper (Den Haag, the Netherlands, CPB).

MADAUS, G. (1988) The influence of testing on the curriculum, in: L. TANNER (Ed) *Critical Issues in Curriculum: 87th Yearbook of the NSSE, Part 1* (Chicago, University of Chicago Press, pp. 83–121.)

MARSH, H. (2007) Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness, in: R. P. PERRY & J. C. SMART (Eds) *The Scholarship of Teaching and Learning in Higher Education. An Evidence-based Perspective* (Dordrecht, Springer).

MARSHALL, B. & WILIAM, D. (2006) *Inside the Black Box: Assessment for learning in the English classroom,* (London, Kings College London, Department of Education and Professional Studies).

McDONNELL, L. M. & CHOISSER, C. (1997) *Testing and Teaching: local implementation of new state assessments.* CSE Report, 442 (Los Angeles, University of California, National Center for Research on Evaluation, Standards, and Student Testing).

MONK, D. H. (1994) Subject area preparation of secondary mathematics and science teachers and student achievement, *Economics of Education Review,* 13, pp. 125–145.

MOTOKO, A., LETENDRE, G. K. & SCRIBNER, J. P. (2007) Teacher quality, opportunity gap, and national achievement in 46 countries, *Educational Researcher,* 36, pp. 369–387.

MUNSON, B. R. (1998) Peers observing peers: the better way to observe teachers, *Contemporary Education,* 69, pp. 108–10.

NYE, B., KONSTANTOPOULOS, S. & HEDGES, L. V. (2004) How large are teacher effects? *Educational Evaluation and Policy Analysis,* 26, pp. 237–257.

O'DAY, J. (2002) Complexity, accountability and improvement, *Harvard Educational Review,* 72, http://gseweb.harvard.edu/~hepg/oday.html.

OECD (2001) *Knowledge and Skills for Life* (Paris, OECD).

OECD (2005a) *Teachers Matter: Attracting, Developing and Retaining Effective Teachers* (Paris, OECD).

OECD (2005b) *Formative Assessment: Improving Learning in Secondary Classrooms* (Paris, OECD).

OECD (2008a) *Teaching, Learning and Assessment for Adults: Improving Foundation Skills* (Paris, OECD).

OECD (2008b) *Education at a Glance* (Paris, OECD).

OECD (2008c) *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-added of Schools* (Paris, OECD).

OECD (2009a) *Creating Effective Teaching and Learning Environments: First Results from TALIS* (Paris, OECD).

OECD (2009b) *Education at a Glance 2009: OECD Indicators* (Paris, OECD).

PELLEGRINO, J. W., BAXTER, G. P. & GLASER, R. (1999) Addressing the 'two disciplines' problem: linking theories of cognition and learning with assessment and instructional practice, *Review of Research in Education*, 24, pp. 307–353.

POPHAM, W. J. (2002) Right task, wrong tools, *American School Board Journal*, 189, pp. 18–22.

RICE, J. (2003) *Teacher Quality: Understanding the Effectiveness of Teacher Attributes* (Washington, DC, Economic Policy Institute).

ROCKOFF, J. (2004) The impact of individual teachers on students' achievement: Evidence from panel data, *American Economic Review*, 94, pp. 247–252.

RUBIE-DAVIS, C. M. (2007) Classroom interactions: Exploring the practices of high- and low-expectation teachers, *British Journal of Educational Psychology*, 77, pp. 289–306.

RUBIE-DAVIS, C. M., HATTIE, J. A. C. & HAMILTON, R. (2006) Expecting the best for students: teacher expectations and academic outcomes, *British Journal of Educational Psychology*, 76, pp. 429–444.

SANDERS W. & RIVERS, J. (1996) *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement: research progress report* (Knoxville, Tennessee, University of Tennessee Value-Added Research and Assessment Center).

SANTIAGO, P. & BENAVIDES, F. (2009) *Teacher Evaluation: a conceptual framework and examples of country practices* (Paris, OECD). www.oecd.org/dataoecd/16/24/44568106.pdf

SEASHORE LOUIS, K., LEITHWOOD, K., WAHLSTROM, K. L. & ANDERSON, S. E. (2010) *Investigating the Links to Improved Student Learning: final report of research findings* (St. Paul, University of Minnesota). www.cehd.umn.edu/CAREI/

TIMPERLEY, H., WILSON, A., BARRAR, H. & FUNG, I. Y. Y. (2007) *Teacher Professional Learning and Development: best evidence synthesis iteration* (Wellington, New Zealand Ministry of Education).

UNESCO INTERNATIONAL BUREAU OF EDUCATION (2006/07) *World Data on Education Country Reports* (Geneva, UNESCO). www.ibe.unesco.org/Countries/WDE/2006/index.html.

WADE, R. K. (1985) What makes a difference in in-service teacher education? A meta-analysis of research, *Educational Leadership*, 42, pp. 48–54.

WOLF, I. F. (DE) & JANSSENS, F. J. G. (2007) Effects and side effects of inspections and accountability in education: An overview of empirical studies, *Oxford Review of Education*, 33, pp. 379–396.