

13

Evaluation of School Principals

Petros Pashiardis and Stefan Brauckmann

Evaluation can be defined as the process through which information and data are collected in order to reach decisions concerning the improvement of staff or the identification of the most effective personnel. It is important to realize that from the day we are born till the end of our lives we go through a series of evaluations and judgments both at the personal and at the professional levels. Evaluations sometimes aim at improving us and at other times aim at ranking or providing us with professional rewards. Evaluations can also be used to impose sanctions. Stemming from the above, it is important to provide some answers to the basic question of why it is necessary to carry out evaluation.

When discussing the evaluation of principals in this chapter there will be necessarily a strong emphasis on what is objective, measurable, and controllable in practice as far as is feasible. Even within this inevitably reduced frame of reference there are many different perspectives and we acknowledge the fact that even the most “objective” data can (and will) be subjectively interpreted. Therefore, when using sophisticated evaluation schemes, all we can hope for is to reduce the inevitable subjectivity that follows. The purpose and practice of the evaluation of principals is likely to be viewed differently by national policy makers, school governors and managers, teachers, students, the parents of students and, of course, researchers, although there is likely to be a common core of interest in educational outcomes that is relevant for all.

Ginsberg and Thompson (1993, as cited in Lashway, 2003) maintain that principal evaluation is difficult because of the complexity and ambiguity of their work (Lashway, 2003). Moreover, the nature of their role is highly context sensitive and thus standard procedures could not work reliably in all situations. In a study in northern California interviews were obtained from 14 principals and 6 superintendents on the politics of principal evaluation (Davis & Hensley, 2000). Both principals and superintendents admitted that “most feedback was qualitative and subjective in nature” (p. 391). Furthermore, it is often argued that evaluation is not utilized in a productive and meaningful way (Reeves, 2004). For example, principal evaluation on the one hand might be wholly positive or lack any constructive feedback for improvement. On the other hand, there may be overly negative results which might cause resentment in those addressed.

In order to be able to respond to the contemporary demands of evaluation, we will address the following topics:

1. Why should school principals be evaluated? In essence, in this section we are trying to explore the culture of educational evaluation (e.g. system monitoring, school inspection) as well as the core purpose of the evaluation of school principals.

2. How are school principals evaluated? What are the means through which evaluation is conducted and by whom? What are the major evaluation methods utilized?
3. Critical review of evaluation practices. What do we know about what works and what does not in the case of principal evaluation?
4. Challenges for Evaluation Practices and Future Research on Evaluation. How does evaluation relate to principals' career stages? We conclude with some propositions for future research.

WHY SHOULD SCHOOL PRINCIPALS BE EVALUATED?

At the beginning of the 21st century, we are experiencing a rapidly changing environment, characterized by complexity and uncertainty. The only stable factor is continuous change. We have witnessed numerous changes in every domain of the human enterprise, from technology to biogenetics as well as huge demographic shifts, state interdependence and globalization, increased competition pressures, knowledge expansion. These societal changes have transformed the school environment into a more dynamic and complex one than in the past (Crow, 2006).

These changes, inevitably, have had a significant impact on education. Increased demands for effectiveness and improved quality have resulted in increased monitoring of the education system at various levels. Evaluation is necessary so that the proper improvements may be made at various levels of the system. For example, the Program for International Student Assessment (PISA), the Third International Mathematics and Science Study (TIMSS) and other ongoing assessments focus on the school level (e.g. school inspection) and in the individual level (teacher evaluation, student achievement). As part of the self-evaluation process, a profile of the school and the whole education system is created by assessing the strengths and weaknesses of each school. Evaluation also provides answers for students and parents in relation to problems facing the individual school and the school system. In addition, evaluation creates reports about the educational system and through these reports holds the school accountable to the community from which it gains support. Finally, evaluation satisfies statutory, normative or other regulations and, as a consequence, improves the whole learning experience of the children at school.

The widespread evaluation culture (seen on almost every governance level) therefore impacts the role of principals and their place in the school environment.

Governments and local stakeholders exert ever greater pressures upon school leaders to perform. Communities question school programs, policies and procedures. Parents demand greater participation in school programs and in the day to day running of the schools. Legislators demand more widespread results and higher student achievement and performance standards. However, tensions result in circumstances where it is difficult to provide an effective response to societal expectations (Stevenson, 2006). Tensions can be understood as the result of the conflict between self-image and the expected social role. The factors contributing to these tensions include accountability demands, lack of sufficient resources, and the instability induced by an uncertain and complex environment.

In view of this complex, accountability- led school environment, where various pressures and tensions are evident, it is essential to be able to effectively evaluate the principal's performance. As Heck and Marcoulides (1996) report, there is a growing interest in accountability and principal evaluation. For example, in the United States, the No Child Left Behind (NCLB) legislation has raised the expectations for school units and principals and "has led to their greater public scrutiny" (Crow, 2006, p.310).

Emphasis on school administrator accountability has changed the ways and processes through which districts evaluate their principals (White, Crooks & Melton, 2002). Moreover, accountability systems require evaluation to be aligned with the standards-based educational environment in order to accurately evaluate principals' effectiveness (Davis & Garner, 2003).

In general, one could argue that the purpose of evaluation is also reflected in the dichotomy of formative/summative evaluation. Formative evaluation is related to the principal's improvement and self-growth while summative evaluation relates primarily to purposes of retention, tenure or merit pay (Marcoulides, Larsen & Heck, 1995). Formative evaluation focuses on process as well as outcomes while summative evaluation is primarily outcomes-based (White, Crooks & Melton, 2002). Formative evaluation serves as a tool for further development while summative evaluation is mainly a tool for accountability (Davis & Garner, 2003). In short, formative evaluation of principals has the person as the main focus whereas summative evaluation has the system as the main focus. Therefore, the purpose of evaluations is twofold: to help a school leader improve (through the provision of constructive feedback) and to help the system make summative decisions about vacancies or rewards that ought to be given for exceptional performance.

In a study carried out in Alberta, Canada, Thomas, Holdaway and Ward (2000) investigated the policies and practice of evaluating principals, including the purpose for their evaluation. Data collection included documents, questionnaires and structured interviews. In phase one of the study, 67 superintendents were asked to provide relevant documents while in phase 2 all superintendents and 100 principals were asked to complete a questionnaire. However, only 63 (out of the 67) superintendents and 62 principals (out of the 100 selected) returned the questionnaire. In phase three of the study, interviews with 10 superintendents and 10 principals were conducted. The results indicated that the most common purpose found in the documents was to "promote professional growth and improvement" (84 per cent). Superintendents also rated the above statement in the questionnaire (mean 2.83 on a 3-point scale) as the most important purpose while the principals considered most important the statement "assess the extent to which expectations are being met" (mean 2.34). However, it must be noted that principals ranked "promote professional growth and improvement" (mean 2.18) as fourth in importance.

The study of Davis and Hensley (2000), investigating the politics of principal evaluation, revealed that most principals perceived the formal evaluation process as more summative in nature than formative. Having this in mind, they did not find the evaluation process useful in shaping their professional development and generally distrusted the evaluators' intentions. On the other hand, superintendents found that the evaluation process provided useful information about principal performance.

The main conclusion which can be drawn from these studies is that there is a gap between official or stated and perceived purposes of evaluation. While superintendents declare a focus primarily on formative purposes of evaluation in line with official documents, principals perceive evaluation more as summative or as a way to respond to external accountability demands. Ongoing consultation with principals regarding evaluation policy and practice and frequent monitoring of how far the process is in line with pre-stated and agreed upon processes may increase the effectiveness of evaluation.

In closing, it should be remembered that principals' evaluation practices can be used for different purposes. Firstly, they can serve during the selection process of qualified candidates for the principal's position. However, this issue is treated in detail in another chapter of this handbook. Secondly, principals can be evaluated while on the job in the form of a qualifications evaluation both in a summative and formative way.

HOW ARE SCHOOL PRINCIPALS EVALUATED?

Methods of Evaluation

What are the means through which evaluation is conducted and by whom? Different methods of evaluation have been utilized internationally for the evaluation of principals. Some of the most important ones will be presented in this section.

External evaluation at the systemic level can be exercised through:

1. Principal evaluation as part of the whole school evaluation,
2. Standards-based evaluation

External evaluation at the individual level can be exercised through:

1. Direct observation
2. Peer evaluation
3. Observation of simulated activities

Internal evaluation at the individual level can be exercised through:

1. Self-evaluation
2. Portfolio evaluation

Principal Evaluation as Part of the Whole School Evaluation

Evaluation can be focused on the whole school. Such evaluation is usually concerned more with academic results, ignoring other critical aspects of a school's life. Evaluation information will be used by policy-makers to make decisions about the allocation of resources to schools and whether to impose sanctions on low performing schools. In this case, the leader's performance is assessed alongside that of the school. Pashiardis and Orphanou (1999) suggest that evaluating schools and holding them accountable in relation to standardized scores might motivate principals and result in more enthusiastic and positive attitude towards educational outcomes. This was the main finding of a quantitative study on the perceptions of 115 elementary school teachers in Cyprus regarding their principals and their leadership styles. Cyprus as well as other countries do not have standardized exams and, therefore, principals do not really know if they are doing a good job, based on this criterion. It is interesting to note that in a qualitative study carried out by Pashiardis (1998), successful primary school principals indicated that they want to have standardized exams and to be compared with other schools.

Standards-Based Evaluation

It is important to note that accountability systems require evaluation to be aligned with the standards-based educational environment (if there are standards) in order to accurately evaluate principals' effectiveness (Davis & Garner, 2003). As Philips (2003) further reports, standards-based qualification systems are becoming increasingly popular in an effort to raise the skills of current and future employees. The Interstate Leaders' Licensure Consortium (ISLLC) Indicators (Council of Chief State School Officers, 1996) and the National Standards for Headteachers

(Department for Education and Skills, 2004) are quite popular in the US and the UK respectively. In the United States, a district or state may use a set of standards such as the ISLLC and decide upon levels of performance (Lashway, 2003). For instance, an evaluation academy design was developed by the Region VII Education Service Center in Kilgore, Texas, designed to provide standards-based evaluation profiles to participants (White, Crooks & Melton, 2002). The academy focused on making principals examine and clearly understand the school leadership standards in order to be able to effectively analyze their own performance.

Direct Observation

Marcoulides, Larsen and Heck (1995) maintain that different means of evaluation, including direct observation can be used to evaluate the leaders' progress. Direct observation involves the evaluation of leadership behavior in real school settings. In the state of Kentucky, principals are evaluated annually through observation of their on-the-job-performance and review of progress in achieving their professional growth targets (Browne-Ferrigno & Fusarelli, 2005). More specifically, in a study conducted by Thomas, Holdaway and Ward (2000) already mentioned, direct observation was regarded as important by five superintendents (of ten) and four principals (of ten). However, one superintendent remarked that this source of information is not useful in the case of paying a limited number of visits per year to each school.

Peer Evaluation

Peer evaluation constitutes a useful way of evaluating a principal's progress. This form of evaluation rests on the assumption that the effectiveness of a leader depends mainly on how others see him/her (Pashiardis, 2001). Principals acquire knowledge from many different sources (Sackney & Walker, 2006). Teachers, other school principals and their superintendents can take the role of a peer or a friend, evaluate the principal's progress and provide him/her with critical feedback. This was reported by Sackney and Walker (2006) in an examination of a number of Canadian studies showing how beginning principals learn to build capacity in learning communities. Discussing plans, problems and actions with critical friends, the principal can gain constructive support in becoming a reflective thinker and improving his/her performance (Quong, 2006). Critical friends bring an external perspective highlighting what may not be apparent to insiders (Earl & Katz, 2002). They facilitate reflection on school issues, remind the leaders of what they have achieved and to what extent, and help them move towards the next goals.

Evidence so far indicates that principals are quite favorably disposed towards the use of the peer evaluation method. They appreciate its value and contribution towards their professional development and progress. However, as Smith (2002) points out, successful principals take advantage of a critical friend if they are convinced that he or she is there to support and not to undermine the leader.

Observation of Simulated Activities

Evaluation can also take the form of observing a leader's simulated practice. Simulated practice can be identified in problem-based evaluation such as responding to case studies (Wong, 2004). Wong (2004), referring to this practice in schools in Hong Kong, points out that this is an important way of appraising principals' knowledge and skills. Porter et al. (2005) recommend the use of vignettes, a case-based evaluation in which principals are required to respond to a problem or dilemma of a scenario or short case. Rubrics are developed so as to evaluate leadership dimen-

sions based on the principals' responses. This is an opportunity to evaluate the leader's ability in applying knowledge and strategies to specific problems. However, evaluation through simulated practice may not relate directly to actual everyday performance. Yet, when used in combination with other evaluation methods validity can be greatly enhanced.

Self-Evaluation

Supervision of the principal's development can be self-monitored by the principals in order to judge their effectiveness in various roles such as personnel management and evaluation, conflict resolution, and public relations (Marcoulides, Larsen & Heck, 1995; Johnson & Rose, 1997). Vaughan (2003) proposes self-evaluation as a way to improve performance. The principal may be seen as a continuous learner who focuses not only on the tasks but also on the processes underlying the tasks. The principal becomes a "skilled self-organized" individual who engages in a "learning to learn dialogue" (Vaughan, 2003, p. 379). At the same time, it is important to highlight that the process of self-evaluation takes time, effort and patience (Johnson & Rose, 1997).

White, Crooks and Melton (2002) conducted a field study in a leadership assessment academy in Texas which focused on providing formative, self-assessment profiles to participants. Data were collected during ten three-day sessions over a period of four years through participant observation, document analysis, questionnaires and interviews as well as through case simulations. Five authentic assessment methods were used:

- *Business interviews*: The participant conducts an interview of a community businessperson so as to assess the principals' skills in applying professional knowledge to the business sector.
- *Case Simulation*: This is a technology-integrated simulation approach which examines complex leadership issues. The assessment provides information about participants' reflective thinking and decision-making skills.
- *Basic Leadership Inventory*: This method provides important information determined by the school faculty on the participants' leadership ability in areas which include leadership behavior, communication skills and organizational management.
- *In-basket writing*: This method focuses on assessing the participant in solving specific problems faced by the school leader. The participants review an in-basket case study and write a letter responding to the case.
- *Student-performance skills*: This object is used to assess the participants' skills in analyzing student achievement data sets and generating development plans for school academic success.

Commenting on the design dynamics of the academy, White, Crooks and Melton (2002) point out that the process of self-assessment increases responsibility for one's own performance and establishes a collaborative relationship between assessors and assessees. Also, the multiplicity of assessment methods and reflective lenses provides participants with multiple perspectives in analyzing their strengths and weaknesses and informing their professional development plans. On the whole, the academy gives special emphasis to assisting principals to cultivate reflective thinking and mental processes associated with effective school leadership.

In the context of self-evaluation practices, Robbins and Alvy (2003) further stress the importance of keeping a self-reflection journal and using it as a tool for evaluating one's own job and bringing greater insight to the experiences and problems faced at school. In the long-term, this can take the form of a portfolio.

Portfolio

A portfolio is a “conceptual container” in which principals place artifacts to document their accomplishments (Lashway, 2003). According to Johnson and Rose (1997), a professional portfolio could be used by administrators to demonstrate their leadership capabilities, to reflect on their learning and to set future goals. The principal identifies a goal and gathers evidence to show that there has been progress towards the desired outcome. Evidence might include journal entries, parent newsletters, staff meeting notes, photographs of classroom activities, records of dropout rates, test scores, handbooks and whatever else the principal considers important data (Lashway, 2003; Robbins & Alvy, 2003). The development of a portfolio is a valuable tool for reflecting on one’s professional growth during a particular school year or over one’s career (Robbins & Alvy, 2003).

Johnston and Thomas (2005) conducted qualitative research on the usefulness of a four-year portfolio evaluation system for beginning principals which was field tested in the state of Ohio. The Portfolio Assessment for School Leaders was designed by the Interstate School Leaders Licensure Consortium and the Educational Testing Service (ETS). It is important to note that the portfolio components were aligned with the ISLLC standards. The participants were 26 principals who completed the portfolio between September 1999 and January 2002. Data were collected through focus group interviews, phone interviews, surveys, the principals’ completed portfolios, annual regional evaluation reports and regional meeting minutes. Almost half of the principals (12 of 26) mentioned that they benefited from the portfolio process. They reflected on problems and developed new initiatives guided by their work. On the whole, the principals reported the existence of a “reciprocal process” between their portfolio work and their development as leaders. On the other hand, there were those principals (7 of 26) who did not find the portfolio meaningful to their practice and saw it only as extra work. In the middle, were the remainder of the principals (7 of 26) who valued the portfolio only at the end of the process. The research also revealed that those who valued the portfolio process most were part of a “supportive social network of professional development” (p. 383). This indicates that if portfolio evaluation is used as a means to professional development then it has the potential to be a learning tool and not just a tool for complying with policy demands. In other words, its real value lies in serving the formative purposes of evaluation.

Based on the above, one may conclude that different means or methods can be used to evaluate a principal’s performance. Each one is characterized by its own advantages or drawbacks. However, no single method seems to outweigh the others. In fact, a combination of methods responding to a specific purpose seems to be the ideal process.

Composition of the Evaluation Team

As far as the composition of the evaluation team is concerned, it is widely accepted that there should be multiple sources of information (Heck & Marcoulides, 1996; Johnson & Rose, 1997; Lashway, 2003). Teachers, students, parents and members of the general public depend on the principal for leadership and direction. Therefore, they should be given an equal opportunity to evaluate the principal’s performance. According to a study by the NAESP (Doud & Keller, cited in Lashway, 2003), there is an increasing tendency to involve parents, teachers and principals themselves in evaluation procedures (Lashway, 2003). A “360 degree” model is proposed by Porter, Murphy, Goldring, Elliott & Linn (2005) in which leaders, teachers and other stakeholders rate specific behaviors in terms of their frequency or importance. In a study conducted by Thomas, Holdaway and Ward (2000) in Alberta, Canada superintendents and principals stressed the need for more stakeholder involvement in evaluation processes.

It should be borne in mind that role makes a difference in perceptions about performance (Heck & Marcoulides, 1996). This was one of the findings of a multilevel analysis evaluating the principal's performance. Principals and teachers of a sample of 56 schools responded to a questionnaire comprising 34 administrative actions. The results showed that principals rated themselves higher than teachers did, suggesting that school role makes a difference in evaluating leadership. However, if the views of the principal do not match the views of the staff then it seems that there may be problems and dysfunctionalities in the day-to-day operations of the school (Pashardis, Savvides & Tsiakkiros, 2005). The degree of congruence of the principal-teacher perceptions could be derived by analyzing the results from leadership effectiveness questionnaires, semi-structured interviews and shadowing. In the case that other stakeholders' views are investigated, the methodology used may need to be sensitive to the role of the person providing the evaluation. For example, the state of Kentucky has established School-Based Decision Making Councils (Browne-Ferrigno & Fusarelli, 2005). These Councils are entitled to select principals from a number of candidates recommended by superintendents. Also, in Hong Kong the School Management Committee (SMC) is responsible for hiring people to the post of principal. The government only sets minimum academic and professional qualifications. The selection criteria, procedures and panel membership are determined by the SMC. However, a close relationship with the sponsoring body which appoints the SMC will likely affect the success in securing a principalship position (Wong & Ng, 2003).

CRITICAL REVIEW OF EVALUATION PRACTICES

What do we know about what works and what doesn't in the case of principal evaluation? In this section we provide a critical review of the different evaluation methods as presented in the previous section. The evidence concerning what works and what doesn't is indeed scarce and inconsistent. Moreover, it is always contextual. In any case, we offer some of the major criticisms which can be found in the literature on principal evaluation. The criticisms mainly revolve around two major areas:

- Content/ Process/Output aspects of evaluation
- Methodological aspects of evaluation

First we proceed with a more in-depth examination of the content, procedural and instrumental aspects of evaluation.

Content/Process/Output Aspects of Evaluation

Firstly, standards-based systems offer a number of benefits for evaluating principals. Such specific systems can hold principals accountable through the use of indicators of school leadership (Kimball, 2006). Performance evaluation includes specific leadership indicators or different levels of performance as well as guidance on how to evaluate principals in relation to standards (Heck & Marcoulides, 1996). As a result, the evaluation data may become more accurate and more reliable.

On the other hand, there are those who strongly criticize the use of standards-based evaluation systems or indicate their inadequacies. More specifically, Reeves (2004) highlights that standards in many evaluation instruments are unclear, incoherent or unrealistic. He criticizes standards that do not explicitly state expectations and stresses that only scarce data exist on the

effectiveness of standards as tools for improving leadership evaluation procedures. Standards are also criticized for not being responsive to the context of the particular school in which a principal is placed. A sound evaluation system would be underpinned by social systems theory, which suggests that there is a close relationship of interdependence between the environment and the system. Consequently, standards should take into account the local and wider context, in short, all those factors which might influence the principal's functioning. This stance is taken by Heck and Marcoulides (1996) who maintain that performance evaluation should be based upon the local school setting. For example, both school level data and student achievement affect the way principals are perceived to be leading. Therefore, if comparisons are to be made about principal performance across schools, a weighting system may be needed to accommodate the influence of contextual factors.

Gronn and Ribbins (2003) suggest that the potential inexperience and unfamiliarity of school councils with selection criteria and evaluation processes might create a gap between national standards and local choices. Davis and Garner (2003) agree that "one size fits all" may not be appropriate for all cases and therefore may render evaluation a counterproductive process. As an alternative, each organization may establish its own leadership domains according to its own culture and needs (Reeves, 2004).

A criticism aimed specifically at the ISLLC standards in the United States is that they underemphasize certain aspects of effective leadership practices. These include the leaders' participation in curriculum development, the promotion of effective instructional and student evaluation practices, the recognition of individual and school achievements, and the leaders' response to the context-specific needs of teachers, students, and other stakeholders (Waters, Marzano & McNulty, 2003; Davis, Darling-Hammond, LaPointe & Meyerson, 2005). However, Murphy (2005) responds by claiming that leadership practice is complex and context-dependent and thus it is impossible to provide a full, definitive list of competencies. The standards were developed "to direct not determine action". Success criteria can be contextualized in accordance with school level data, existing school performance, and the principal's career stage (Murphy, 2005).

Leithwood, Aitken and Jantzi (2001) point out that process measures can detect unsatisfactory features of a school's functioning, allowing one to change these features for school improvement. These processes can be used to demonstrate a school's accountability. Educators have control over these processes in contrast to the standards against which outcomes are compared. Outcomes are influenced by many variables such as an individual's background and the wider community. In a study conducted in Alberta, Canada, 32 out of the 63 principals noted that school and community characteristics influenced principal evaluation, for instance "the nature of the school, the type of the students, the spiritual atmosphere in the school, and the size, culture, language, and unique needs of the community" (Thomas, Holdaway & Ward, 2000, p. 231). It is therefore important to acknowledge the particular context or situation in relation to leadership behavior.

Student achievement is also used to evaluate the principal's performance (Scheerens & Witziers, 2005; Reeves, 2004). Nevertheless, one-dimensional evaluations based only on test scores are usually flawed. More specifically, it is argued that leadership is a necessary but not sufficient condition for high academic achievement (Marcoulides, Larsen & Heck, 1995). Evaluation based upon student achievement may place too much emphasis on test scores when the principals may not have direct control over all the elements which contribute to these (Heck & Marcoulides, 1996; Kythreotis & Pashiardis, 2006; Usdan, McCloud & Podmostko, 2000).

Standardized tests cannot give a complete picture of learning (Johnson & Rose, 1997). Smith (2002) argues that standardized tests cannot be regarded as authentic evaluation since they cannot represent the complex, real world of school. Hoy and Miskel (2005) also contend that evaluating

organizational effectiveness only in terms of student achievement underestimates the importance of the social systems theory, which advocates that school organizations represent open systems with inputs, processes and outputs. School performance should therefore be evaluated in the whole context of a school's operation and not just student achievement (Kellaghan & Greany, 2001; Reeves, 2004). Similarly, Vaughan (2003) and Smith (2002) indicate that performance evaluation needs to address processes as well as outcomes.

Methodological Aspects of Evaluation

It can be observed that the summative purposes of evaluation attract more attention than the formative ones (White, Crooks & Melton, 2002; Usdan, McCloud & Podmostko, 2000) and therefore more criticism. According to White, Crooks and Melton (2002) principal evaluation is characterized by top-down processes for the purposes of reward or punishment rather than for professional development. The same stance is taken by Usdan, McCloud and Podmostko (2000) who maintain that principal evaluation rarely provides opportunities for reflection and professional growth. Nevertheless, effective evaluation also needs to address the professional growth of the administrator (Davis & Garner, 2003). In this context, both the evaluator and the evaluatee have a responsibility for reflection, analysis and improvement (Reeves, 2004). As Vaughan (2003) indicates, the principal may be seen as a continuous learner who focuses not only on the tasks but also on the processes underlying the tasks. The conditions and processes could more effectively be improved through the use of formative methods of evaluation. Thomas, Holdaway and Ward (2000) also report that the majority of principals and superintendents who responded to their questionnaire desire to move towards formative evaluation.

Reliability and validity of principal effectiveness measures are also critical (Heck & Marcoulides, 1996; Marcoulides, Larsen & Heck, 1995; Davis & Garner, 2003). Any evaluation model should be able to yield valid and reliable data about the principal's performance. Validity refers to the extent that the model measures what it is designed to measure. Reliability refers to the consistency of the results when performance is repeatedly measured. According to Reeves (2004), if five different superintendents evaluate a principal in five different ways, then such a process is not consistent and therefore unreliable. Reliability is suggested to be the key to fairness and therefore candidates must be treated in the same way.

Also contested is the degree to which the evaluation of leadership has been undermined by the paucity of valid tools for measuring leadership performance (Porter et al., 2005). Expert panels have sometimes been employed in order to review the specific instruments to be used for evaluation. Porter et al. (2005) propose a Leadership Evaluation System (LAS) utilizing a multicomponent system of evaluating leadership behaviors. These components include a set of behavior-rating scales, scoring rubrics, vignettes, logs, and collections of evidence. This combination of a number of evaluation methods may enhance the quality of information and the validity of evaluating the principal's performance. Validity increases by using multiple data sources and by examining the multilevel nature of organizations (Heck & Marcoulides, 1996). As far as reliability is concerned, it may be increased by carrying out repeated measurements over time and taking into account local school data.

Finally, an evaluation instrument requires field-testing (Reeves, 2004) against a wide range of leaders, beginning principals and veterans. During the field testing period, the use of at least two evaluators conducting independent evaluations of a particular leader allows comparison. If they come to different conclusions, the instrument appears to exhibit ambiguity and signifies that leadership domains and performance descriptions are not specific enough. During the field testing period, the leaders may also evaluate themselves and their results should be compared to

those of the evaluators. When there is agreement about a performance level, then the leadership domains and performance specifications can be judged reliable. The field test will provide feedback about the leadership evaluation process before it is implemented at the school level. In this way evaluation results will be more accurate, fairer and more useful to the evaluation team and the principal him/herself.

After having reviewed some of the major criticisms of principal evaluation, at this point, it is important to gain a deeper understanding of the career trajectories of teachers as they potentially move towards, into and through the principalship. This is important, as school leaders may need a differentiated approach to their evaluation, depending on their career position.

CHALLENGES FOR EVALUATION PRACTICES AND FUTURE RESEARCH ON EVALUATION

The main aim of this section is to explore what we should know and what we should do, depending on career point. We conclude with some propositions for future research.

Challenges for Evaluation Practices

The previous review of the literature indicated that the role of the principal has become complex and demanding. This is the result of: (a) external, societal, and economic changes, (b) the growing accountability movement, and (c) pressures for raising school performance. In this changing context of leadership practice, a shortage of qualified candidates for the principal's position is observed in numerous parts of the world (Stevenson, 2006; Davis et al., 2005; Usdan, McCloud & Podmostko, 2000). Thus, it is necessary to provide support both for the aspiring as well as for serving principals in order to attract them to the profession and help them to achieve increased effectiveness.

Leadership evaluation is a significant aspect of a principal's development and a currently evolving field of research (Porter et al., 2005). The improvement of the principal's performance is a central issue concerning educators and policy makers. One aspect of the issue which attracts much attention is the debate over the nature of the motivation related to a specific career stage. Motivation is defined as "an internal state that stimulates, directs and maintains behaviour" (Hoy & Miskel, 2005, p. 157). According to Vaughan (2003), performance-related pay has become increasingly popular. This means that high performance is rewarded with increased pay or extra money to spend for school purposes (Rasch, 2004; Heck & Marcoulides, 1996; Hoy & Miskel, 2005). For example, in Cincinnati, Ohio, the administrators' pay raises partly depend on whether they achieve high student performance. However, when evaluation is used to determine pay-for-performance, fear and competition emerge. It is arguable that this kind of reward system suppresses the employee's internal motives and morale (Vaughan, 2003). According to Earl and Katz (2002), extrinsic motivation results only in short-term, superficial outcomes. Alternatively, intrinsic motivation could more effectively motivate principals to improve their performance (Rasch, 2004; Earl & Katz, 2002). This is partly because sanctions or praises are not effective; it is in fact the activity itself which rewards the principal's performance to a great extent. Of course, depending on the career stage, pay-for-performance or other motivation techniques could affect him/her to a different extent; this will vary amongst individuals and amongst different cultures.

This is why, in part, principal evaluation schemes need to be closely aligned with the socialization phases or the various career stages of a principal from preparation prior to the principalship, to entry, development, achieving autonomy, and finally the moment of exit. Identification of

career stages produces profiles for principal's trajectory (Weindling & Dimmock, 2006). When an individual principal's behavior is compared to these collective profiles, the degree of deviation can be traced and interpreted accordingly. The knowledge of the stages through which principals move can significantly inform the practices of evaluation. Nevertheless, the achievement of the desired objectives for school improvement requires that realistic career frames are created for school leaders. Too many ambitious expectations may lead to the leaders' physical and professional burn out.

Learning to be a principal is a continuous developmental process of socialization (Weindling & Dimmock, 2006). According to Merton's socialization theory, it is constituted by two overlapping phases (Weindling & Dimmock, 2006, p. 334):

- Professional Socialization refers to learning what it means to be a principal, from personal experience and formal training prior to taking up the post.
- Organizational Socialization refers to the knowledge, values and behaviors which are required to perform a particular role within the organization.
- Hart distinguishes three categories of socialization (Cheung & Walker, 2006, p. 392):
- Encounter, adjustment, stabilization.
- Anticipation, accommodation, role management.
- Confrontation, clarity, location.

Weindling adapted the categories to describe a three-stage model of socialization (Cheung & Walker, 2006). The first stage includes the processes of encounter, anticipation or confrontation. In this stage, a new principal learns to make sense of the new environment. The second stage consists of adjustment, accommodation and clarity. In this stage, the principal tries to "fit in", creates interpersonal relations with school stakeholders and learns to manage ambiguous situations. The third stage consists of stabilization of the principal in the new position.

In addition, the National Foundation for Educational Research (NFER) Secondary Heads Project conducted by Weindling and Earley (1987, cited in Weindling & Dimmock, 2006, pp. 335–337) indicated that no matter how good the preparation of the principals is, they will have to go through a transition phase during which there is an attempt to make sense of the particular organizational situation (Weindling & Dimmock, 2006). Weindling used the results of the NFER study and other authors in order to create a model of stages of transition through the principalship (headship) (Weindling & Dimmock, 2006, pp. 335–337):

Stage 0 Preparation prior to headship.

Stage 1 Entry and encounter (first months). This is the beginning of organizational socialization at which the new head attempts to make sense of the new situation.

Stage 2 Taking hold (three to 12 months). The new head attempts to challenge the status quo and decide on the new priorities.

Stage 3 Reshaping (second year). This is the period of implementing major changes in the school.

Stage 4 Refinement (years three to four). Previous changes are refined.

Stage 5 Consolidation (years five to seven). Consolidation of previous changes.

Stage 6 Plateau (years eight and onwards). After a seven years period the heads have initiated most of the changes they wanted and they have reached a plateau.

The NFER data show that principals were less likely to reach a plateau if they had moved to a second headship (Weindling & Dimmock, 2006). The latter would move to Stage one and

then move at different rates through the rest of the stages. The NFER data suggest that primary principals can move through the stages more quickly due to the smaller size of primary schools and their less hierarchical structures.

As principals move through the various stages of their career, their evaluation needs change considerably. Therefore, evaluation should be differentiated in such a way as to respond to the needs and career profiles of the specific phase the principal is experiencing. Evaluation should distinguish between the pre-principalship stage, the initial stage, the seasoned stage, and the exiting principalship stage. Pre-principalship evaluation should take into account the Professional Socialization stage as described by Merton, which is similar to Stage 0 of Weindling. Initial Principalship evaluation should address the challenges faced by beginning principals including the Organizational Socialization stage as described by Merton and Stages 2 and 3 described by Weindling as well as the latter's adapted model of encounter, anticipation or confrontation. Seasoned principalship evaluation should be informed by Weindling's Stages 4 and 5 as well as the model of adjustment, accommodation and clarity. The exiting principalship evaluation should take into account Stage 6 as well as Weindling's recommendation on the stabilization, role management and location model.

Recommendations for Future Research

As Lashway (2003) points out, research on principal evaluation is still inadequate. Effective leadership evaluation has been notably absent. Most leadership evaluations may be "infrequent, late, unhelpful, and largely a source of administrative bother" (Reeves, 2004, p. x). Taking into account these observations as well as the findings from the literature review, we offer a number of recommendations relating to future research on principal evaluation. The use of a differentiated form of evaluation should, in every case, help principals cope with tensions between personal perceptions and outside demands (Cheung & Walker, 2006). Hence, principals can become more successful in building adaptive confidence in themselves and other members of the learning community as they enter uncertain situations, learn from their mistakes and adapt as they move through their career (Sackney & Walker, 2006).

As far as leadership competencies are concerned, there are specific trends traced while examining the standards adopted by a number of countries. These involve creating a vision, distributing leadership, acting in an ethical manner, using information technologies, and collaborating with the wider community. According to Hage and Powers (cited in Crow, 2006), the evaluation criteria for principals in the post-industrial era should emphasize creativity and innovation in contrast to the criteria of efficiency and quantity of work which characterized the rationality of the industrial society (Crow, 2006). In the new knowledge society the principals have to create school environments which respond to the demands of life-long learning (Crow, 2006). They have to respond to the expanding technology needs of teachers and students and support an inclusive learning environment. Nevertheless, the success criteria should be contextualized in accordance with influencing factors such as the school level, the existing school performance, and the principal's career stage (Murphy, 2005). Thus, it becomes an indispensable need for policymakers and public education officials to collect information regarding the supply, effectiveness and changing roles, as well as career needs, of principals (Usdan, McCloud & Podmostko, 2000). Therefore, more research is needed in the area of career progression of principals and their needs as they evolve.

It is also important to include dispositions as well as behavioral standards for evaluation (Hill, 2005; Cheng, 2002; Murphy, 2005). This is a necessary provision since leadership practice and evaluation is influenced by values and beliefs. More comparative research is needed in this area in order to establish the universal and the contextual.

Reliability and validity of principal effectiveness measures are also central. As a result, evaluation should examine the appropriate skills for the principal's role, be flexible enough to allow for variations in how the role is perceived, and include multiple data sources (Heck & Marcoulides, 1996). Multilevel data analysis can be effectively used. However, as Heck and Marcoulides (1996, p. 12) highlight, "any evaluation model that tries to capture all of the subtleties of the role and operationalize all of the day-to-day activities of the principal is doomed to failure". Browne-Ferrigno and Fusarelli (2005) believe that research needs to address the issue of whether high-stakes accountability has made the job too demanding. Longitudinal and cross-case inquiries are needed on the implementation of standards for the evaluation of principals as well as their influence on student achievement.

In summary, evaluation is not just the design of a plan or process; it involves the creation of a new culture. It is, in fact, "the result of gradual steps that serve to build trust in the process and confidence in the ability to succeed in it" (Hill, 2005, p. 29). It is therefore crucial to try to build such a culture of evaluation for the sake of the educational system as a whole. As Usdan, McCloud and Podmostko (2000) indicate, state and local school systems, higher education, businesses and principals themselves will need to work together to find ways to support the profession and to ensure enhanced quality and improved results.

In conclusion, we would like to stress once more that there is an urgent need for education systems worldwide to revisit the issue of principal evaluation. This position is based on research evidence showing that the principal's role is indeed crucial for students' improved academic achievement. Therefore, there is a need for new evaluation systems which are based on meritocracy. Such an evaluation needs to be systematic and must examine who produces excellent results and why. These new systems need to provide us with the ability to differentiate between excellent and mediocre principals and therefore reward those who perform best and, at the same time, provide the necessary professional growth activities for those who do not perform as well. We need a variety of arguably revolutionary principal evaluation systems to help transform educational systems into modern and open ones able to react effectively to the challenges posed to education around the globe. These new evaluation schemes need to provide answers to crucial questions, such as: how do we reward and differentiate better principals from not so good ones? And how do we connect these rewards with enhanced student learning?

Finally, there is a need to review and refine the existing systems of evaluation. The issue of the quality, functionality and effectiveness of any evaluation system should be the concern of all stakeholders, especially when important decisions are made regarding the professional future of principals. Consequently, evaluation itself should become the object of evaluation. This concept of meta-evaluation, first introduced by Scriven (1995), involves the attempt to control and understand the way it functions, with the objective of identifying the problems which might occur and taking corrective action for their early treatment. According to Pashiardis, Savvides and Tsiakkiros (2005), the implementation of meta-evaluation procedures is necessary in order to secure that:

- the wider objectives and the more specific goals of the evaluation are achieved;
- the proposed evaluation procedures are correctly implemented;
- the evaluation is valid and reliable.

Through such meta-evaluation it is hoped that those who undertake the implementation of principal evaluation will realize that they are expected to exhibit enhanced responsibility knowing that their work itself will be evaluated.

REFERENCES

- Browne-Ferrigno, T. & Fusarelli, B. C. J. (2005). The Kentucky Principalship: Model of School Leadership Reconfigured by ISLLC Standards and Reform Policy Implementation. *Leadership and Policy in Schools*, 4, 127–156.
- Cheng, Y. C. (2002). *The Changing Context of School Leadership: Implications for Paradigm Shift*. In K. Leithwood & P. Hallinger (Eds.). *Second International Handbook of Educational Leadership and Administration* (pp. 485–518). Dordrecht: Kluwer Academic Publishers.
- Cheung, R. B. & Walker, A. (2006). Inner Worlds and Outer Limits: The Formation of Beginning School Principals in Hong-Kong. *Journal of Educational Administration*, 44(4), 389–407.
- Council of Chief State School Officers (1996). *Interstate School Leaders Licensure Consortium: Standards for School Leaders*. Washington, DC: Council of Chief State School Officers.
- Crow, G. M. (2006). Complexity and the Beginning Principal in the United States: Perspectives on Socialization. *Journal of Educational Administration*, 44(4) 310–325.
- Davis, E. E. & Garner, E. (2003). *Current Practices of Evaluating Superintendents and Principals in a Standards-Based Environment*. Technical Report 2002–2005. Pocatello, ID: ICEE.
- Davis, S., Darling-Hammond, L., LaPointe, M. & Meyerson, D. (2005). *School Leadership Study - Developing Successful Principals. Review of Research*. Stanford, CA: SELI.
- Davis, S. H. & Hensley, P. H. (2000). The Politics of Principal Evaluation. *Journal of Personnel Evaluation in Education*, 13(4), 383–403.
- Department for Education and Skills (2004). *National Standards for Headteachers*.
- Earl, L. & Katz, S. (2002). *Leading Schools in a Data-Rich World*. In K. Leithwood & P. Hallinger (Eds.). *Second International Handbook of Educational Leadership and Administration* (pp. 485–518). Dordrecht: Kluwer Academic.
- Gronn, P. & Ribbins, P. (2003). Evolving Formations: The Making of Secondary School Principals on Selected Small Islands. In P. Pashiardis & P. Ribbins (Eds.). *International Studies in Educational Administration. Special Issue on the Making of Secondary School Principals on Selected Small Islands*, 31(2), 76–94.
- Heck, R. H. & Marcoulides, G. A. (1996). The Assessment of Principal Performance: A. Multilevel Evaluation Approach. *Journal of Personnel Evaluation in Education*, 10, 11–28.
- Hill, J. S. (2005). Developing a Culture of Assessment: Insights from Theory and Experience. *Journal of Political Science Education*, 1, 29–37.
- Hoy, K. & Miskel, I. (2005). *Educational Administration. Theory, Research and Practice*. New York: McGraw Hill.
- Johnson, N. J. & Rose, L. M. (1997) *Portfolios. Clarifying, Constructing and Enhancing*. Lancaster, PA: Technomic Publishing Company.
- Johnston, M. & Thomas, M. (2005). Riding the wave of administrator accountability: a portfolio approach. *Journal of Educational Administration*, 43(4), 368–386.
- Kellaghan, T. & Greany, V. (2001). *Using assessment to improve the quality of education*. Paris: International Institute for Educational Planning (UNESCO).
- Kimball, S. M. (2006). *Case Study of the Initiation of Standards-Based Principal Performance Evaluation in Washoe County School District*. Consortium for Policy research in Education, Wisconsin Center for Education Research, University of Wisconsin-Madison.
- Kythreotis, A. & Pashiardis, P. (2006). Exploring Leadership Role in School Effectiveness and the Validation of Models of Principals' Effects on Students' Achievement. CCEAM Conference "Recreating Linkages between Theory and Praxis in Educational Leadership". Nicosia, Cyprus, 12–17 October, 2006.
- Lashway, L. (2003). Improving Principal Evaluation. Clearinghouse on Educational Policy Management. *Eric Digest* 172.
- Leithwood, K., Aitken, R. & Jantzi, D. (2001). *Making Schools Smarter. A System for Monitoring School and District Progress*. Thousand Oaks, CA: Corwin Press.
- Marcoulides, G. A., Larsen, T. J. & Heck, R. H. (1995). Examining the Generalizability of a Leadership

- Model: Issues for Assessing Administrator Performance. *International Journal of Educational Management*, 9(6), 4–9.
- Murphy, J. (2005). Unpacking the Foundations of ISLLC Standards and Addressing Concerns in the Academic Community. *Educational Administration Quarterly*, 41(1), 154–191.
- Pashiardis, P. (1998). Researching the Characteristics of Effective Primary School Principals in Cyprus. A Qualitative Approach. *Educational Management and Administration*, 26(2), 117–130.
- Pashiardis, P. (2001). Secondary Principals in Cyprus: The Views of the Principal Versus the Views of the Teachers- A Case Study. *International Studies in Educational Administration*, 29(3), 11–23.
- Pashiardis, P. & Orphanou, S. (1999). An Insight Into Elementary Principalship in Cyprus: The Teachers' Perspective. *International Journal of Educational Management*, 13(5), 241–251.
- Pashiardis, P., Savvides I. & Tsiakkiros, A. (2005). Meta-Evaluation of the Proposed Systems. In P. Pashiardis, I. Savvides & A. Tsiakkiros (Ed.). *The Evaluation of Teachers' Work: from Theory to Praxis* (pp. 343–349). Athens: Ellin Publishing House (in Greek).
- Philips, D. (2003). Lessons from New Zealand's National Qualifications Framework. *Journal of Education and Work*, 16(3), 289–304.
- Porter, A., Murphy, J., Goldring, E., Elliott, S. & Linn, R. L. (2005). A Proposal to Develop and Test an Education Leadership Performance Assessment. Report Submitted to the Wallace Foundation by Vanderbilt University.
- Quong, T. (2006). Asking the Hard Questions: Being a Beginning Principal in Australia. *Journal of Educational Administration*, 44(4), 376–388.
- Rasch, L. (2004). Employee Performance Appraisal and the 95/5 Rule. *Community College Journal of Research and Practice*, 28, 407–414.
- Reeves, D. B. (2004). *Assessing Educational Leaders. Evaluating Performance for Improved Individual and Organizational Results*. Thousand Oaks, CA: Corwin Press.
- Robbins, P. & Alvy, H.B. (2003). *The Principal's Companion*. Thousand Oaks, CA: Corwin Press.
- Sackney, L. & Walker, K. (2006). Canadian Perspectives on Beginning Principals: Their Role in Building Capacity for Learning Communities. *Journal of Educational Administration*, 44(4), 341–358.
- Scheerens J. & Witziers, B. (2005). *Educational Leadership and Student Performance*. University of Twente: Department of Educational Organisation and Management.
- Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation*, 21, 111–129.
- Smith, W.J. (2002). *School Leadership and Self-Assessment: Guiding the Agenda for Change*. In K. Leithwood & P. Hallinger (Ed.). *Second International Handbook of Educational Leadership and Administration* (pp. 485–518). Dordrecht: Kluwer Academic.
- Stevenson, H. (2006). Moving Towards, Into and Through Principalship: Developing a Framework for Researching the Career Trajectories of School Leaders. *Journal of Educational Administration*, 44(4), 408–420.
- Thomas, D. W., Holdaway, E. A. & Ward, K. L. (2000). Policies and Practices Involved in the Evaluation of School Principals. *Journal of Personnel Evaluation in Education*, 14(3), 215–240.
- Usdan, M., McCloud, B. & Podmostko, M. (2000). *Leadership for Student Learning: Reinventing the Principalship. School Leadership for the 21st Century Initiative. A Report of the Task Force on the Principalship*. Washington, DC: Institute for Educational Leadership.
- Vaughan, S. (2003). Performance: Self as the Principal Evaluator. *Human Resource Development International*, 6(3), 371–385.
- Weindling, D. & Dimmock, C. (2006). Sitting in the “Hot Seat”. New Headteachers in the UK. *Journal of Educational Administration*, 44(4), 326–340.
- White, D. R., Crooks, S. M. & Melton, J. K. (2002). Design Dynamics of a Leadership Assessment Academy: Principal Self-Assessment Using Research and Technology. *Journal of Personnel Evaluation in Education*, 16(1), 45–61.
- Wong, P. (2004). The Professional Development of School Principals: Insights from Evaluating a Programme in Hong-Kong. *School leadership and Management*, 24(2), 139–162.
- Wong, K. & Ng, H. (2003). On Hong Kong: The Making of Secondary School Principals. In P. Pashiardis

& P. Ribbins (Ed.). *International Studies in Educational Administration, Special Issue on the Making of Secondary School Principals on Selected Small*, 31(2), 35–53.

Waters, T. Marzano, R. J. & McNulty, B. (2003). *Balanced Leadership: What 30 Years of Research Tells Us About the Effect of Leadership on Student Achievement*. Denver, CO: Mid-continent Research for Education and Learning.