Chapter 11
Principal Leadership and Challenges for Developing a School Culture of Evaluation

Maria Luz Romay, Constance Magee, and Charles L. Slater

Abstract The purpose of this chapter is to identify how school principals can develop a culture of evaluation that will contribute to improving the quality of learning processes. If this goal is achieved, teachers, administrators and staff will respond more effectively to the needs of students and society in general. The chapter provides an overview of how the role of school principal has changed substantially during the last decade, demanding more rigorous assessment practices and accountability. The authors also discuss the nature and purposes of evaluation, emphasising the importance of integrating evaluation, planning, and decision-making. They describe common problems and attitudes that may impact the effectiveness of evaluations; in contrast, the chapter proposes several conditions that will allow the development of a culture of evaluation in schools. In order to demonstrate how these criteria work, a case study illustrates how they were applied by a school principal to resolve specific evaluation issues. The authors recommend that effective evaluations require cooperation between the school administration and teachers with open communication and active participation. This will not be possible without financial support and adequate training. It is the authors belief that if school principals put these recommendations into practice, it will be possible to develop a culture of evaluation within each educational community.

Keywords Assessment and accountability • Core curriculum standards • Culture of evaluation • Formal evaluations • Ethical issues • Formative evaluation • Summative evaluations • Integration of planning • Decision-making • Methodological capability • Participative dialogue • Principal's role • Programme evaluation • Resistance to evaluation

M.L. Romay
School of Physical Therapy, University of the Incarnate Word, San Antonio, TX, USA
e-mail: romay@uiwtx.edu

C. Magee
Lindbergh Middle School, Long Beach, CA, USA
e-mail: CMagee@lbschools.net

C.L. Slater (✉)
College of Education, California State University, Long Beach, CA, USA
e-mail: Charles.Slater@csulb.edu

11.1   Introduction

When schools fail, the first person held responsible for the failure is the principal. Principals are coming under increasing scrutiny from the public and private sector to ensure that their schools are meeting the needs of all students. In the United States schools are measured by how well their students perform on yearly state tests.

Principals are critical to school success (Fullan, 2001, 2008a, 2008b; Leithwood, Lewis, Anderson, & Wahlstrom, 2004; Reeves, 2009; Whitaker, 2003). Marzano, Waters, and McNulty (2005) found that principal effectiveness has a direct impact on school progress and student achievement. It is also clear that the job of the principal has changed dramatically over the past decade. Good principals used to be those who took care of student discipline and efficiently managed the site. Today's principals must be agents of change, committed to continuous improvement. They must be masters of finance, human resources, instruction, data analysis, and politics, while balancing the needs of their students, parents, teachers, and district administrators (Wildy & Clarke, 2008; Wohlstetter, Datnow, & Park, 2008). It is no wonder that many view the principal's increased responsibilities as overwhelming and some question whether one person can effectively accomplish everything that is expected (Wildy & Clarke, 2008; Wohlstetter et al., 2008).

Even with all of these expectations, we must add one more. Principals need to be able to evaluate student achievement and determine whether it is increasing in the short-term and in the long-term. As instructional leaders, principals lead teachers in setting goals, planning, and evaluating (Schmoker, 1999). Principals do not need to be experts in evaluation, but they need to have a firm grasp of how it works and how it can be integrated into the school programme (Slater, McGhee, Nelson, & Meno, 2011).

This chapter reviews in its first section how the role of school principals has changed substantially in the United States with the passage of the No Child Left Behind Act in 2001 and the advent of the common core curriculum. These developments have impacted policies related to assessment and accountability. The following sections discuss the nature and main purposes of evaluation in education, emphasising the importance of integrating evaluation, planning, and decision-making processes. Understanding these theoretical principles and factors will enable school leaders to oversee evaluation efforts. Another section of the chapter describes the most common problems of evaluation, in particular potential educators' attitudes or responses when they are called to participate in an evaluation in order to improve the practice of evaluation. To improve the practice of evaluation, several conditions can facilitate the development of a culture of evaluation in schools. Finally, a case study illustrates and elaborates on these evaluation issues.

Different evaluation methods and techniques have been developed based on diverse theoretical models (Hill, 2009; Madaus, Scriven, & Stufflebeam, 1990). Currently a wide variety of resources are available regarding its different concepts and principles; several authors have offered critical perspectives on issues that evaluators encounter as they conduct assessments in diverse environments. The purpose of this article is to guide principals in the development of an evaluation culture.

11.2   Evaluation in Elementary and Middle Education the United States

Since education is not mentioned in the United States Constitution, it has been left to individual states to develop and fund public schools. The role of the Federal Government in education was quite small until the beginning of the twenty-first century, but when the No Child Left Behind Act (NCLB) became effective on January 8, 2002, it opened a new era in educational history and framed the debate about the future of public education (U.S. Department of Education, 2009a). It began as history making bipartisan legislation passed by Congress and signed by the President. The decision to target improvements for public schools led to a high-stakes accountability programme and labelled an increasing number of schools as failing each year. The goals of the legislation require that students from low income families, different racial groups, with disabilities, or who are learning English as a second language, must demonstrate proficiency in mathematics and language arts.

The NCLB legislation (2001) was initially supported as a way to help all groups of students increase academic proficiency. The NCLB legislation mandated that all subgroups meet the national proficiency standard of 100 % by 2014. African American, Latino, and Special Education students from low socioeconomic backgrounds are each looked at as individual groups.

Schools whose students did not meet federal targets are placed in Programme Improvement (PI) and must meet state targets for two successive years in order to exit from the programme. Failure to exit PI came with sanctions that increased in severity for each additional year that a school failed to meet the targets. All sanctions included removing the current principal unless the principal was new to the site. In some cases sanctions also included reconstitution of the teaching staff, closing the school, or re-opening the school as a charter. Programme Improvement schools also lost funding and were required to offer transfers to parents who requested a non-PI school.

Schools were held accountable through annual testing, academic progress, school report cards, and teacher qualifications. The four goals behind the legislation included: (1) assistance for economically disadvantaged students; (2) increasing the pool of highly qualified teachers; (3) increasing the literacy rate of students; and (4) holding schools accountable for the success or failure of their students (Munro, 2008). Schools that failed to meet Annual Yearly Progress (AYP) goals were placed in Programme Improvement. Parents could transfer their children out of low performing schools.

The NCLB required annual testing of at least 95 % of students at each school in Grades 3-8 in reading and mathematics. In addition to overall scores, data were compiled on students from low income families, students from different racial groups, those with disabilities and English language learners. The tests were aligned with state academic standards. Students as a whole and all student groups were required to make adequate yearly progress (AYP) (Slater et al., 2011).

Schools with a high concentration of students from poor families received Title I funds from the Federal Government (U.S. Department of Education, 2009b). Title I

schools that failed to meet targeted goals 2 years in a row must offer students a choice of other public schools to attend. After 3 years, students must be offered supplemental educational services. All students were required to reach a minimum level of proficiency by 2013 until the goal was revised. Moreover, states and districts completed a report including reporting on student achievement for all groups and schools. Additionally, all teachers must meet the definition of highly qualified by having a Bachelor's degree, state certification, and proof that they know the discipline. Schools are also expected to provide quality professional development experiences for teachers and paraprofessionals.

11.2.1   Assessment and Accountability

Student assessment in the US has become synonymous with accountability and high-stakes testing. Criterion-referenced assessments replaced norm-referenced tests that were used in many states. The states then measured the extent to which students were meeting state objectives.

In the first years of the legislation, school districts grappled for the first time with an examination of test results that were disaggregated by school group. Previously, a district might have good results overall and not notice or publicise lower results of minority students such as African Americans or Latinos. Achievement is now measured for all students in a school and disaggregated by ethnicity, gender, students in poverty, English language learners, and special programme students. Discussion at all levels has centred on the gap in achievement between the majority and minorities (Ladson Billings, 2006). The system for reporting data is completely transparent so that parents, teachers, citizens, or researchers can consult school and state websites to see complete test results as well as demographic data. In California, each school is compared to overall state results as well as to comparable schools with similar demographics.

Educators have become informed about individual student performance and the public has unprecedented access to data about schools. Many schools have developed careful plans to monitor students, assess, and plan based on test results.

11.2.2   Problems with Educational Accountability

Unfortunately, standardised testing for educational accountability has had several negative effects. The use of standardised tests has driven out more authentic means of instruction. The system has been limited to paper and pencil tests, and there is little room for assessment in which students demonstrate performance in real world settings.

Standardised testing also tends to limit teachers' focus on areas of the curriculum that are not tested such as science, social studies, the arts, health, second languages, and physical education. Testing only language arts and mathematics has resulted in a narrowing of the curriculum to emphasise just what is tested. Even within language arts and mathematics there is often a restriction to content and instruction related to the form of the test.

Students who are most likely to need help in passing the test are assigned to special test preparation classes that are separate from the regular curriculum and may emphasise test taking skills (McNeil, 2000a, 2000b). They may be taken out of music, art, or special education to focus on the state test. There is less opportunity for field trips, extended activities such as library research projects, scientific investigations, or arts performances.

The amount of additional time in test preparation is quite significant and while it takes away from the regular curriculum schedule, it may still not improve test scores, much less make long-term learning gains for students. In Texas superintendents reported requiring students to take practice tests, and in some cases, students were spending up to 35 days, or 7 weeks practicing for accountability system-related examinations (Nelson & McGhee, 2004; Nelson, McGhee, Reardon, Gonzales, & Kent, 2007).

Disaggregating data by income and ethnic group helped to focus attention on students who were not achieving. However, these students have not necessarily been receiving additional resources or an improved curriculum. Rather, they may be receiving a curriculum of test preparation. When compared to the National Assessment of Educational Progress Results, a number of studies have indicated very weak relationships, if any, between accountability testing and student achievement (Nichols, Glass, & Berliner, 2012).

In the worst cases, students who were not likely to pass the test were pressured to leave school. McNeil, Coppola, Radigan, and Heilig (2008) reported that Texas had publicly reported gains in test scores even as additional numbers of students were dropping out of school. Heilig and Darling-Hammond (2008) reported that some school districts tried to obtain higher test scores by testing fewer students at the elementary level and pushing out students at the high school level.

One way to combat some of these problems is to focus on growth targets instead of rigid Adequate Yearly Progress (AYP) percent targets. Individual targets should be calculated for each student and subgroup based on current achievement, rather than using a set percent for proficient or advanced proficient. It is unrealistic to expect that all students in all schools be 100 % proficient in both math and language arts. The system also did not indicate levels of growth, it only signified whether or not the school had made the percent target. Students who qualified for special education and students who were learning English were placed in specific programs, based in part on low test scores, to help them succeed academically. A growth model would more accurately evaluate the progress of the schools and pinpoint the students who are in need of additional services.

11.2.3   Common Core Curriculum Standards

Until recently, each state had different standards, and testing in one state was not necessarily comparable to another state. There was also great variation among school districts within a state. Some districts and schools followed state standards closely while others ignored them.

In 2012, the National Governors Association Centre for Best Practices (NGA Centre) and Council of Chief State School Officers (CCSSO) published a set of national standards that gained wide attention. In a period of only 2 years states began to adopt the new standards to replace their separate sets of standards (NGA & CCSSO, 2012). These standards are intended to emphasise the knowledge and skills that students need to succeed in college and careers, while emphasising complex thinking (Porter, McMaken, Hwang, & Yang, 2011). The Federal Government helped spur the rush to participate when it made participation in the Common Core Curriculum a requirement for states to get funding for Race to the Top grants (U.S. Department of Education, 2009b).

The Common Core Curriculum has pushed school districts toward common assessments as well. States were required to develop new standardised tests by 2014–15. To accomplish this work, states joined either the Partnership for Assessment of Readiness for College and Careers Assessment Consortium (PARCC, 2013) or the SMARTER Balance Consortium (Smarter Balance Consortium, 2013). Common curriculum and assessments bring questions about the nature and role of evaluation to the fore.

11.3   Nature and Role of Evaluation

Evaluation is a natural part of our everyday life: people make evaluations in the form of judgments determining whether something is good or bad, desirable or not. Evaluation seems to be fundamental in our developmental process, as we make decisions that allow us to become mature adults and to assume different responsibilities. Evaluations are also made at the personal or the professional level, and are influenced by personal expectations or preferences. Often those judgments are not made carefully and in an objective manner (Shawn & Greene, 2006).

Formally speaking it is important to acknowledge that evaluation is "a profession, a practice, and a discipline" (Mathison, 2005, p. 1). As the practice of evaluation evolved, it became increasingly professionalized; and it has become entrenched within educational systems in many countries. Applied to different educational problems or areas, evaluation implies an intentional process that responds to different needs of people, groups, or institutions (Martínez Slanova, 1980).

Thus, systematic and formal evaluations require explicit evidence and objective criteria for interpreting data (Kemmis, 1989). These types of evaluations are used to

analyse the status of any educational or social program, assess teacher performance, identify what have been the outcomes of learning processes, or to conduct large and complex institutional self-studies (Berk, 1999; Erwin, 1991; Glatthorn, Boschee, Whitehead, & Boschee, 2012; Guerra-Lopez, 2008; Kennedy, 2010; Peterson, 2009; Rueda, 2011). Scientific methods are applied in these cases making clear what sources were consulted before any judgments were made. Usually these evaluations are based on scientific principles that regulate social research. Formal evaluations should demonstrate that the evidence does not rely only in individual opinions, but that information is gathered collectively.

These formal evaluations respond to different purposes. For example, they provide information to public audiences for accountability. They could also be useful for policy making, promoting knowledge through the development of theories, or enhancing specific practices. In each case the choices for the purpose of evaluation and how it is done influences its approach, and validates the process (Nevo, 1986).

Even though the distinction between informal and formal evaluations is important, one needs to recognise that often individuals involved in these processes interpret data in the context of their own practice and knowledge. In other words, informal and formal evaluations may be related in different ways. A formal evaluation could be proposed to offer more explicit and usable knowledge than what is presented informally about a specific situation. Both types of evaluations could be complementary, and could interact providing some reliable knowledge (Patton, 1990).

The root of the word "value" comes from the Latin "valere", meaning "to be worth or to work out the value of something" (Shawn & Greene, 2006, p. 6). Therefore the term itself could lead to measuring the quantitative value of something or estimating its worth. To understand this full meaning, one must accept or use quantitative and qualitative methods.

Most definitions of evaluation include at least one of the following elements: the assessment of worth or merit, its functions, roles, methods, and its purpose. Based on these distinctions we present three definitions that represent these diverse emphases:

> Evaluation is a type of inquiry undertaken to determine the merit and/or worth of some entity, in order to improve or refine what is evaluated, or to assess its impact. (Lincoln & Guba, 1981, p. 550)

> Evaluation refers to the process of determining the value of something, or the product of that process. It normally involves identification of a relevant standard, investigation of the performance of those who are evaluated, and integration or synthesis of the results achieved. (Scriven, 1991, p. 139)

It is not surprising that no single definition is universally accepted by evaluators today. Given the different perspectives and dynamic nature, evaluation as a discipline encompasses several theories, models, and methodologies. Shadish, Cook,

and Leviton (1991) in their meta-analysis describe three stages of the development of major evaluation theories: in the beginning, according to Madaus et al. (1990), theorists emphasised a search for truth, looking for solutions to social problems (Scriven, 1967). In a second stage evaluators developed studies aimed to produce politically and useful results based on detailed knowledge of how organisations operate [this stage may be represented by Cronbach (1982), Carol Weiss (1992) and Robert Stake, (1990)]. More recently evaluators have tried to integrate previous contributions insisting on organisational processes and decision-making with a more comprehensive approach [such as the work of Stufflebeam et al. (1971), and Rossi & Freeman, (1992)].

In light of the previous concepts and contributions of numerous authors, in this chapter we adopt a more recent and broad definition:

> Evaluation is an applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs, value, merit worth, significance, or quality of a program, policy, or plan related to educational processes. Conclusions made in evaluations encompass both an empirical aspect (if it is a case) and a normative aspect (judgment about value). It is the value feature that distinguishes evaluation from other types of inquiry. (Mathison, 2005, p. 139)

Generally evaluations in education serve a broad purpose, which is to assess the status and effectiveness of specific policies, programs, students' learning outcomes, or institutional development. According to Álvarez García (1997), the most commonly identified purposes and functions of evaluations are:

(a) **Accountability** – The intention is to demonstrate how far a programme has achieved its objectives, how well it has used its resources, and what has been its impact. This type of evaluation will mainly meet the needs of administrators, programme coordinators, or sponsors from diverse organisations. Often this purpose can be related to control or supervision. It is useful because it allows stakeholders to know what has happened to the resources devoted to specific projects or programs.

(b) **Increasing the efficiency of planning processes or policy making** – Evaluations could be proposed to justify a policy or programme analysing developmental stages to define the next steps in strategic planning processes (Álvarez García, 2008). This type of evaluation mainly meets the needs of planners and policy makers. They could follow a conventional planning process or focus more on innovation (Bridges & Groves, 2000).

(c) **Organisational improvement** – These evaluations allow institutions or schools to enhance or review their performance, structures, and procedures (Schmoker, 1999), in order to determine the level of their effectiveness or assess the strategies used. This kind of evaluation mainly meets the needs of principals or school administrators who want to identify opportunities for change. In today's educational reality research has proven that those evaluations should incorporate the teacher's own reflection on their teaching practice, in other words to include self-assessment practices (Romay & Crispin, 2000).

Table 11.1 Typical questions in evaluation phases

Evaluation phases	Most common questions
Preliminary proposals	What are the current priorities set by governmental agencies related to evaluating quality of school programs?
	Are there specific areas that need to be evaluated?
	What resources are available that can be used in evaluation efforts?
	What problems should be more urgently analysed or studied?
Initial stage	Have the objectives of the planning process been adequate to the needs of the target population?
	Are the goals and policies consistent with the needs of students and teachers?
	What is known about the problem that has been proposed for evaluation?
Processes analysis	Are all members of the institution/programme involved as they need to be?
	Are the existing programs achieving their goals? Are there other alternatives?
	Are the resource allocations transparent and known to those who manage the programme?
	What are the strengths and weaknesses of the programme?
	Have the standards set by leaders been achieved by students?
	How could delivery of the programme be improved?
Implementation/application of the evaluation results.	Have the results of the evaluation been clearly presented and well understood?
	What are the key points that require change or improvement?
	How much may the implementation of these changes cost?
	Who will oversee the implementation of the recommended changes in the study?

Adapted from Shawn & Green (2006) in Chapter 15, Tables 5.1 – 5.4

(d) **Knowledge production** – This type of evaluation is for groups or institutions that want to confirm specific assumptions and theories that they have applied in their practice (Chen, 1990), and determine what lessons can be learned for the future. These evaluations would be particularly important for leaders and policy makers who want to develop new projects or renew existing programs.

Depending on the evaluation's purpose and the stage of the process, one can identify typical questions as Table 11.1 shows.

There are other specific purposes of evaluations such as diagnostic studies, innovative projects, or support of particular objectives established by principals or administrators. In these cases, evaluations serve as strategies to facilitate the growth and learning of small groups, communities, or people. Scriven (1967) originally proposed two central functions of evaluations: formative or summative.

(a) Formative evaluation provides information to improve a product or process. For example, a formative evaluation of instructional materials would ideally be conducted prior to full-scale implementation (Flagg, 1990), or expert reviews of the content of a programme may provide useful information for modifying or revising selected strategies (Owen, 2006). Therefore, this type of evaluation is predominately used in educational and training settings; it often allows educators to discover issues related to organisational structures, confusions within the learning process, or a need for more illustrations and examples. It may reveal concerns that would lead to revised and improved teaching strategies.

(b) Summative evaluation provides short-term effectiveness or long-term impact information to decide whether or not to adopt a product or process. Summative evaluation can occur just after new materials, programs, or software are implemented in full or after they have been in place for a long period of time. It is important to specify what decisions will be made as a result of this type of evaluation, and then, develop a list of questions to be answered. Other times that summative evaluation could be appropriate are: when teachers or administrators would like to know if certain objectives have been met; or if an innovation was efficient in terms of time to completion or had any unexpected outcomes.

Álvarez García (1997) has proposed a list of elements that all evaluations should include:

1. Clear identification of the issues or needs to be studied, analysing whether there is room for change;
2. Contextual factors and resources that may influence the evaluation process;
3. Level of complexity of the study;
4. Analysis and interpretation of data;
5. Initial results and recommendations based on the information gathered;
6. Expected and non-expected results;
7. Positive and negative impact;
8. What resources can be used in the change process; and
9. Follow-up and implementation of recommendations.

An understanding of the broad purposes of evaluation suggests that it should be tied to systematic processes that determine the direction of schools, including planning and decision-making (Álvarez García, 2008). Stufflebeam et al. (1971) maintains that what is important is how evaluation is integrated with those processes (See Fig. 11.1).
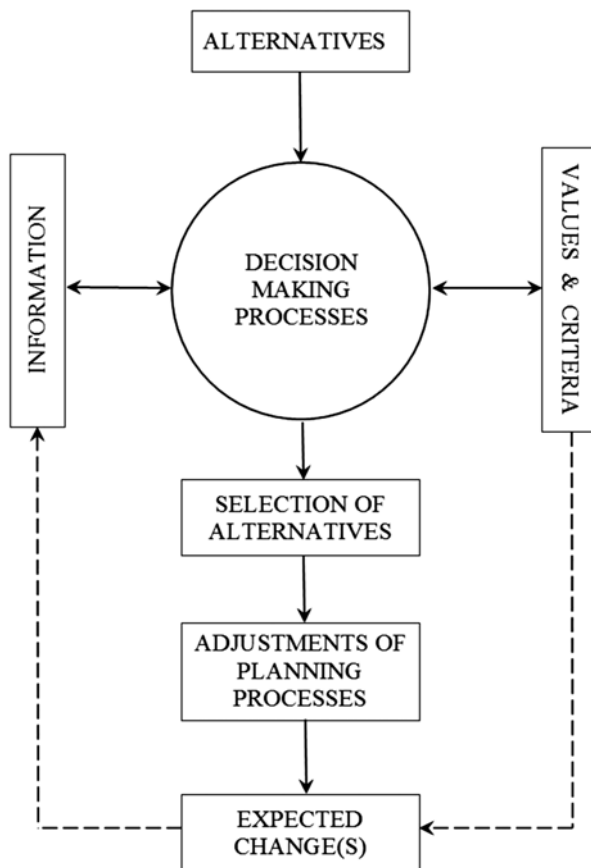
Fig. 11.1 Integration of Planning and Evaluation (Adapted from Stufflebeam et al. (1971)

11.4   Main Issues Affecting Evaluation Processes

Often evaluations face some of the following obstacles or challenges (Calonghi, Gianola, Groppo, Perucci, & Reguzzoni, 1991):

- Lack of clear ideas about evaluation;
- No clear identification of the issues to be evaluated;
- Misunderstanding of some aspects of the purpose and functions of evaluation;
- Confusion of the evaluation process with scientific research;
- Disarticulation of evaluation processes with planning, decision-making and other organisational processes;
- Inadequate methods or techniques applied;
- Not enough knowledge on how to gather valuable information (overlooking aspects of validity, reliability, usefulness);
- Incorrect interpretation or use of findings; and
- Conditions of the social context that make the evaluation not feasible.

In some circumstances defining evaluation criteria may involve negotiations at various levels and throughout the whole process, and it is difficult to get a consensus on relevant decisions. It is particularly problematic if the objectives and purposes of the evaluation are not clear at the outset, or when evaluators find ambiguity between declared and hidden objectives (Álvarez García, 1997), particularly if the organisation is large and complex. It is not uncommon that during evaluations participants feel stressed, even fearful, expending too much time in discussions that waste energy (Spaulding, 2008).

Evaluations proposed at the organisational level require a commitment to participate in interventions that may bring positive changes and define initially specific criteria for measuring success. Also it is important to be aware that in many organisations, members tend to place more value on an external evaluation than when it is conducted by internal resources, arguing that external evaluation is more objective and self-evaluation has the risk of being subjective. However, experienced evaluators recognise that internal evaluations are particularly valuable and truthful if they are conducted in alignment with expected standards. In fact, when members of an organisation are involved in an evaluation more directly, they will have more opportunities for learning and personal development. Of course, there must be controls to assure that administrators, teachers, or students adhere to ethical standards of evaluation.

11.5   Ethical Challenges of Evaluation

Unfortunately, cheating is commonplace in US schools: 56 % of middle school students and 70 % of high school students report having cheated (Decoo, 2002). The pervasiveness of cheating by students requires attention to ethical issues as well as organisational structures to minimise the incidence of cheating.

Cheating by students has been around for as long as schools have administered tests, but the turn of the century has brought a new kind of cheating, cheating by schools. The testing and accountability system that was implemented on a national level in the US was begun in the State of Texas. The Houston Independent School District became known for high-stakes testing that carried financial rewards and punishments for principals and teachers depending on how well students scored (Nelson, McGhee, Meno, & Slater, 2007; Slater et al., 2011).

There have been allegations of several types of cheating. The most straightforward example is when school employees change test results or give students advance information about what is on the test. Several Houston school results have been officially questioned and in 2006 a State audit cited 442 campuses for testing irregularities. In 2010, some Houston employees were reassigned after allegations of cheating (Radcliffe, 2010).

Another type of cheating in Houston is more indirect and is part of the way the system was designed in Texas. Students may show test score gains on the officially reported state measure, but fall far short on other standardised measures that are not

reported. Schemo and Fessenden (2003) reported that Houston school gains on the Stanford Achievement Test were far smaller than on the Texas State Test. While there was no wrong-doing that could be traced to any individual, the lack of correlation between the Texas State Test and other measures suggests toleration of systematic deception.

Linda McNeil (1986, 2000a, 2000b) at Rice University in Houston has been a persistent critic of the Houston testing system. She argues that school officials 'game' the system to make the results look good. They systematically exclude some students through manipulation of the rules, provide instruction only for students who are likely to show test score increases, and provide so much testing practice as to harm students' broader learning. Bohte and Meier (2000) have called this type of cheating goal displacement. The organisation operates to maximise incentive rewards based on published criteria while neglecting or even working against the broader intent of the policy.

The largest case of cheating to date took place in Atlanta where 178 principals and teachers were charged with cheating by artificially raising test scores to meet district targets (Winerip, 2011). The superintendent would regularly gather all staff in the Georgia Dome at the beginning of the school year and have school personnel sit in the order of their school test scores. The highest performing schools would sit at the front, and the lowest performing would sit at the back. The superintendent had been named 'superintendent of the year' and was recognised by the Secretary of Education. She collected $600,000 in bonuses over 10 years in addition to her $400,000 annual salary. She said, "Where people consciously chose to cheat … the moral responsibility must be with them."

One of the central issues is who bears responsibility; the school officials who designed and implemented the system, or those who did the actual erasing of scores. In a moral wrong, someone loses and someone gains. Teachers risked being marginalised if they did not participate in "erasure parties". Principals might even lose their jobs if they did not show score increases.

To what extent should the superintendent be held responsible for the cheating? Heads of organisations are quick to take credit for accomplishments but slow to acknowledge a role in failure. Quick and Normore (2004) argue that moral leadership rests with the institution's leader. Not only should the leader act according to a personal code of ethics but he/she must also understand concepts of systems thinking to determine how relationships, support structures, and decisions made by school leaders impact the entire school.

Beyond the school level, we could also look at the accountability system itself. Some organisational structures are much more likely to elicit cheating than others. Several positive cultural qualities can reduce the likelihood of cheating. If teachers are motivated by internal rewards such as satisfaction with class activities and their own professional development are less likely to cheat than those who work for external rewards of salary and bonuses. If they perceive the demands of the school and the district as legitimate, they are more likely to buy into the system of testing. If there are caring relationships and tolerance for error or acceptance of mistakes, teachers are more likely to report results honestly.

Those responsible for designing the system need to take into account the features of the system that can encourage or discourage cheating. A positive culture is crucial to create an ethical environment against cheating, but there also need to be systems in place to guard against cheating. A few incidents of cheating can spread and undermine a positive culture.

Students suffer the most when test results are falsified because they gain their own concepts of truth at least partially from their experiences in school. The message from the 178 teachers and principals in Atlanta was that it is all right to cheat in order to avoid punishment and gain what you want. The truth of the curriculum becomes subject to convenience. We change the facts to fit our beliefs.

11.6   Different Attitudes toward Evaluations

Based on the above difficulties, it is common to find different responses toward formal evaluation processes. Eventually some people will try to ignore what the evaluation may demand from them. These attitudes are self-protective mechanisms that often block the purpose of the evaluation. These responses are similar to what people do when they have to go through a tax audit. Education and communication is required for personnel at all levels of the organisation to convince people of the intended benefits of the proposed evaluation before they are able to modify their assumptions or correct misapprehensions about evaluation.

There are four attitudes that participants involved in evaluation processes could take:

1. **Rejection or resistance**. Often people respond to serious systematic evaluative efforts evading participation, fearing that it will bring more control from management, or negative results. In these cases evaluations are seen then as oppressive actions.
2. **Indifference**. This attitude is a result of misunderstanding the nature of evaluation, or lack of information about the objectives of the project. The attitude could be related to deficiencies of management, or the belief that nothing useful can come from it. In this case people just tolerate what is going on without responding honestly and thinking that there is no other option than to acquiesce.
3. **Passive agreement** due to pragmatic reasons. Another response could be to follow instructions from management but not show commitment to the results of the evaluation. This attitude is common if the evaluation has a conventional approach because it is perceived as part of a routine or a required investigation connected with organisational planning procedures.
4. A positive **collaboration and participation** with a critical perspective. Evaluation projects that generate these attitudes are generally well communicated from the beginning of the process. People realize the importance of getting useful information to improve a problem that has been identified. It is relevant because participants acknowledge opportunities provided by the evaluation for

group development; people who have this attitude will accept more easily the role of values within the evaluation process and change will be welcome (Wholey & Newcomer, 1994).

11.7   Conditions for Developing a Culture of Evaluation

A culture of evaluation requires direct involvement of the principal in a cooperative relationship with teachers. The role of the principal is to work with teachers to present clear ideas about the role and function of evaluation, identify specific needs or problems to be evaluated, select adequate methodology and techniques, and consider contextual factors that may influence the evaluation process. This cooperative relationship is characterised by several conditions that imply actions and beliefs on the part of both the principal and teachers. The following conditions are explained in general and then applied to a case study of a school that developed a culture of evaluation: political support, technical knowledge, administrative feasibility, methodological feasibility, ability to follow up, and participative dialogue.

1. **Political Support**: Teachers must not only be willing to carry out evaluation tasks but ideally, they will embrace the philosophy of evaluation; in other words, they are willing to do it. Principals set up a clear organisational structure and designate responsibilities. They lead different constituencies or groups that might be involved in an evaluation process so that they accept all that the evaluation process implies.
2. **Technical Knowledge**: The principal is responsible for conducting the process and providing training for evaluation projects. A clear purpose and approach will lead to a good understanding of all steps that need to be followed. Teachers will gain a sense that they know how to do it.
3. **Administrative feasibility**: Evaluation projects often require complex and challenging actions. The principal's role is to manage and create a positive atmosphere and space to obtain resources and gain access to information. Before the evaluation starts, the principal may need to analyse and negotiate conditions that arise from power struggles or obstacles to the evaluation. Teachers will feel capable to do it.
4. **Methodological capability**: Teachers will have adequate skills to conduct the process and design instruments. The principal will lead a team to assume the coordination of the evaluation process and assure that they have specific training on the methods that they intend to use and enough statistical knowledge to interpret the results. Teachers will have a sense of competence.
5. **Ability to follow up**: Teachers will know how to use the findings to make use of the information to improve their practice and design new systems. The principal will keep in mind the objectives of improving student achievement, facilitating the quality of instruction, and making ethical decisions. Teachers will be able to apply their knowledge.

6. **Participative dialogue**: The principal will promote two-way communication with teachers in a transparent process. They will receive information and critical comments will be welcomed and encouraged. The principal will integrate the results and actions that may follow within the organisational planning processes and which are supported by management (Álvarez García, I & Romay, 2013). Teachers will feel that their voices are heard.

11.7.1 The Case of Leonard Middle School

The conditions for developing a culture of evaluation can be illustrated by a case study from an urban middle school in Southern California that we will call Leonard Middle School. It is a diverse school with an enrolment of 706 students: 22 % are African American and 64 % are Latino; 88 % of the students are eligible for free or reduced lunch; 22 % are English Language Learners.

California uses the Academic Performance Index (API) as a measure of accountability to determine the extent to which schools are meeting state and federal standards and testing requirements. The scores at Leonard had been decreasing and in 2011, the school API score was down to 702. A year later in 2012 there was a remarkable increase to 750, a 48 point gain. It met its goals school-wide and for all student subgroups. However, it did not meet the federal requirement for Adequate Yearly Progress (AYP) in mathematics and continues to be in Programme Improvement (PI) status.

This extraordinary gain in student achievement was made only 1 year after a new principal was appointed. Ronda Madison was an experienced teacher and assistant principal in the school district and had just spent 5 years as principal of Wentworth, a similar middle school, where she was also able to change the achievement pattern. The API went from 609 to 729 API. The decile ranking at the state level went from 2 to 3 and the ranking among schools with similar demographics went from 7 to 9. Her record of turning around Wentworth school led the superintendent to appoint her at Leonard with the hope that she could do the same thing there.

Madison's philosophy was expressed in her doctoral dissertation in which she said that she started at her first school, Wentworth, by getting to know the staff, students, and parents as quickly as possible. Teachers had asked for many changes and improvements to student discipline at their Change of Principal Workshop. This workshop is conducted by the school district whenever there is a change of principals. Administrators from the district use surveys and interviews to prepare a report of what teachers feel needs to be changed and what needs to be kept as is. The report summarised a school meeting where the staff discussed, openly and honestly, what they wanted to change and keep, for example a specific 'dos and don'ts' list for the new principal. Teachers wanted to change the procedures at Wentworth Middle School.

During her first few months, Madison worked on the list of changes the teachers created at the Change of Principal Workshop. They discussed behaviour standards and created a system of rewards and consequences to help motivate students to improve behaviour. The teachers expressed concern that the students were "running the school" and the teachers did not feel supported by the prior administration. The teachers said they felt blamed for student actions. During her first year as principal, Madison focused on changing student behaviour, modelling expectations, and creating a scholarly climate.

At Leonard, Madison began with a similar report from the Change of Principal Workshop that asked for greater attention to student discipline. She began making organisational changes in the summer before teachers returned to school. By the time they arrived new systems were in place. The changes were widely accepted and allowed Madison to proceed to the next step, developing a culture of evaluation.

She formed a leadership team and delegated discipline and other time-consuming duties that did not directly affect classroom instruction. Visiting classrooms and giving teachers timely and direct feedback was a priority. The more visits and notes she left teachers, the more instruction improved. Visits to classrooms were a critical part of making sure teachers were collaborating, implementing professional development strategies, and working to improve student achievement.

Madison also used teacher data meetings to track student progress. She met with teachers by department to look at the data. These data meetings were scheduled at least once a quarter, and in some cases, they met each month. These meetings gave teachers a chance to share successful practice, while the principal had the opportunity to hear first-hand what teachers were doing to help students learn and help teachers use their data to inform classroom interventions. The data questions that she used were:

1. Tell me about your student results.
2. Where did you see the most improvement? What strategies did you use?
3. Who is continuing to struggle? What is your plan to help these students?
4. Comparing your class results to your English Language Learners (ELLs) and African American (AA) subgroups, what do you see?
5. Who are your students who scored Far Below Basic (FBB), what is the story for each of them?
6. What content are you planning to revisit and why?
7. Let's look at the test and your most-missed items. Show me the ones where most students missed (50 % or more). What did they have problems with? Why?
8. Did you try any new strategies that you would like to share?
9. How do you motivate your classes to improve? How do you display the data in your room?
10. What is your goal for the next assessment?

At the end of the first semester (end of January), she began to plan for the next year. She talked with teachers about evaluating who should be teaching certain classes or grade levels, and they explored changing master schedules to improve

Table 11.2 How the principal looked at school's data

Date	Data & purpose	Frequency
September	Academic Performance Index 9 (API)	Yearly
	Look at results, big picture	
	These data come from California Department of Education (CDE) website – look at subgroups and trends over time, graph API growth and compare to other schools	
September	Grade level data by department	Yearly
	Faculty meeting looking at California Standards Test (CST) data by department	
	Make a list of positive accomplishments, areas that need to change, and prescriptions for change (plus, delta, RX)	
September	CST-Longitudinal data by teacher	Yearly
	Individual	
	Count the number of improved students and the number who decreased (+−count) and plan for improvement based on results	
Monthly	Classroom data compared to CST and subgroup	Monthly
	Look at student level data, share stories of students and strategies that work	
Quarter data	Quarter exams compared to CST, and subgroup	Quarterly
	Colour graph of class results compared to CST and also separated by subgroups	
Quarter	Comparing Quarter Data across school and teacher, and compare to trends from previous year	Quarterly
	Look at quarter trends, predicting CST	
Mid-year	Similar Schools Rank – CDE website	Yearly
	Compares like schools across the state on 10 point scale	
June	End of Course assessments (EOC) data and CST predictions	Yearly
	Teachers share results and predict CST results based on data trends	

opportunities to learn. These types of long-term planning behaviours signalled that she had enough information to begin changing the instructional programs of the school. Selecting the best teacher for a class or grade level can significantly change the climate and productivity of the school. She weighed the pros and cons of each change as she contemplated how to improve Leonard for the following year. The annual calendar to review data is shown in Table 11.2.

Finally, Madison reviewed summative data to make other decisions regarding student placement, interventions, and resources for the following year. She began setting goals for the next year based on summative results. This cycle repeated itself each year and became part of the school culture.

At Leonard, the principal met with teachers monthly to look at formative assessments. The early gain in API could be attributed to the monthly department data meetings and her visits to classrooms and feedback to teachers each week. She followed a model that suggested a sequence of change starting with discipline, transitioning to classroom instruction, and finally looking at school systems.

Madison's actions at Leonard Middle School and the reaction of the teachers illustrate the conditions necessary for creating a culture of evaluation and suggest additional conditions that are desirable to support principal leadership.

11.7.2   Conditions for Creating a Culture of Evaluation at Leonard School

1. In this case the political support necessary for a culture of evaluation was multi-layered. The data system in the school was required by the federal government after the passage of the NCLB Act that mandated testing and accountability across the grades. The State of California extended accountability and mandated the California Standards Test (CST). The Long Beach School District put into place common quarterly assessments in Mathematics, Science, History, and Language Arts and designed the Change of Principal Workshop. Finally, it was the principal who brought a philosophy of using data to improve instruction and implemented regular classroom observations and data meetings.

2. The principal was attuned to the teachers' need for technical knowledge. She introduced the assessment process in small steps and set up a data wall in her office. In the first year, she required teachers to post their results on a quarterly basis and established a norm of transparency where all teachers could see all class data. In the second year she increased the posting of data to twice quarterly. The district established a new data system called LROIX that allowed teachers to see data across schools. Teachers were able to make comparisons with similar schools and look at each other's data.

3. The principal arranged for **administrative feasibility** by setting aside time for teachers to meet and discuss planning and assessment. Sixth grade teachers were reluctant to discuss data as a group, and the principal responded by mandating a time when teams came together in a common location, the library.

4. The principal became the chief instructor to create **methodological capacity** among teachers. She met with each teacher once a month to review test results on an overall basis. These data ranked students from Advanced to Far Below Basic. She and the teacher looked at specific areas in which students had difficulties. Then she made the teachers responsible for developing plans to address the deficiencies by providing a data form that they could use and suggesting ways that they could display data in the classroom.

5. The principal attended to **follow-up** to make sure that the results of data analysis were being used in the classroom. She also used information from objective data to explore more subjective data. The stories of successful students and those who were struggling were shared and examined in light of school and community factors. Professional development was planned to address common concerns that arose out of the process.

6. The principal communicated in a transparent and timely manner to create **participative dialogue** by establishing a timeline for assessment activities. She met regularly with teachers to work on assessment plans and made sure that teacher concerns were addressed by asking teachers to evaluate each session according positive aspects of the process and areas needing improvement (plus/delta).

The development of a culture of evaluation at Leonard illustrates the main points of this article. The principal paid attention to both formative and summative evaluation processes, making sure that teachers could use data to improve instruction and monitoring the overall progress of students. The key was to integrate evaluation, planning, and decision-making.

The principal addressed teacher attitudes toward evaluation by taking a proactive approach and first understanding their concerns about student discipline, and then putting into place systems that require attention to data. Her work met the conditions necessary to create a culture of evaluation, but it would not have been possible without complementary evaluation systems at the national, state, and district level.

11.8   Conclusions and Recommendations

Evaluation and assessment form the critical strategy for accountability to improve school performance. This chapter has described why the management and use of evaluations, particularly for educational leaders, are not easy, but they are crucial for improving the quality of learning processes.

The globalisation of contemporary society and the need for democratic knowledge require that education more than ever before has to be an integral part of social development and culture. Effective assessment processes include a conscious effort to create and maintain what we are calling a "culture of evaluation".

The role of the principal is to ensure that there is the political will on the part of different actors in the process, and that they are able to learn continually how to conduct an evaluation process with rigour and objectivity. The principal should find strategies to ensure readiness for participation, paying attention to different reactions or personal interests that might be affected. The best way is to communicate objectives clearly and avoid punishment. From the managerial point of view, evaluation efforts always require good organisation skills to assure the implementation of coordinated action.

The most critical conditions of any assessment process are the timeliness and usefulness of results. As this chapter highlights, it is not enough to develop a good design, utilising sound methods or gathering enough data; the evaluation results must probe for validity and indicate how the information obtained can be applied for improvement. An evaluation project can be valuable beyond the school site. Clear processes can be replicated by other schools to enhance knowledge of both the evaluation process and successful practices with students.

There are several recommendations for any school that is undertaking an assessment of students. First, teamwork is essential. Effective evaluations require cooperation between the administration and teachers with open communication and active participation. Second, state and national standards must be adhered to as required by law but they must be developed in a way that is appropriate for the social and cultural context. Third, the justification for assessment must always be related to the improvement of the quality of education. Fourth, assessment cannot be carried out without adequate financial support. Fifth, continuous training is necessary for all staff, and support from specialized personnel is critical to support their efforts.

The principal has the responsibility for the development of a culture of evaluation, but issues of growth, equity, interdependence, and auto-determination go beyond the principal's control. School districts often grow in size with new students to serve and at the same time, districts change demographically often with greater diversity from students of colour, immigrants, and families in poverty. The school is also part of a larger system and is dependent upon enlightened polices on the state and national level. Depending on the system, the principal will have more or less autonomy to carry out an evaluation.

Promoting and developing a culture of evaluation in schools goes beyond technical requirements or traditional functions. The principal does not control many of the large variables and will thus need courage to innovate and advocate for constant improvement. Authentic leadership requires risk, persistence, and dedication to create a culture of evaluation.

References

Álvarez García, I. (1997). Theory of evaluation: Some practical examples (Teoría de evaluación: Algunos ejemplos prácticos). In *Cuadernos de Filosofía y Letras (10).* UNAM, 1985. Re-edited by UNESCO/CREFAL.

Álvarez García, I. (Ed.). (2008). *Planning and development of educational and social Projects. (Planificación y desarrollo de proyectos sociales y educativos)* (9th ed.). México: Limusa.

Álvarez García, I., & Romay, M. L. (2013). *Challenges for a culture of evaluation in educational institutions (Desafíos para una cultura de evaluación en las Instituciones Educativas).* Mexico: Limusa-Noriega (Eds.).

Berk, R. A. (1999). *Thinking about program evaluation*. Thousand Oaks, CA: Sage Publications.

Bohte, J., & Meier, K. J. (2000). Goal displacement: Assessing the motivation for organizational cheating. *Public Administration Review, 60*, 173–182.

Bridges, E. M., & Groves, R. (2000). The micro and macro politics of personnel evaluation: A framework. *Journal of Personnel Evaluation in Education, 12*, 321–337.

Calonghi, L., Gianola, L. P., Groppo, M., Perucci, G., & Reguzzoni, M. (1991). *The evaluation issue (El problema de la evaluación)*. Madrid: ITER Editors.

Chen, H. T. (1990). *Theory-driven evaluation*. Newbury Park, CA: Sage.

Common Core State Standards Initiative. (2010a). *Common Core State Standards for English Language arts & literacy in history/social studies, science, and technical subjects.* Retrieved from http://www.orestandards.org/assets/CCSSI_ELA%20Standards.pdf

Common Core State Standards Initiative. (2010b). *Common Core State Standards for mathematics.* Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Decoo, W. (2002). *Crisis on Campus: Confronting Academic Misconduct*. Cambridge, MA: MIT Press.

Erwin, D. T. (1991). *Assessing student learning and development: A Guide to the principles, goals and methods of determining college outcomes*. San Francisco: Jossey-Bass.

Flagg, B. N. (1990). *Formative evaluation for educational technologies*. Hillsdale, NJ: Erlbaum Associates.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco: Jossey-Bass.

Fullan, M. (2008a). *The six secrets to change*. San Francisco: Jossey-Bass.

Fullan, M. (2008b). *What's worth fighting for in the principalship?* San Francisco: Jossey-Bass.

Glatthorn, A., Boschee, F., Whitehead, B. M., & Boschee, B. F. (2012). *Curriculum leadership: Strategies for development and implementation* (3rd ed.). Los Angeles: Sage.

Guerra-Lopez, I. (2008). *Performance evaluation: Approaches for improving program and organizational performance*. San Francisco: Jossey-Bass.

Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*(2), 75–110.

Hill, H. C. (2009). Evaluating value-added models: A measurement perspective. *Journal of Policy Analysis and Management, 28*, 102–109.

Kemmis, S. (1989). Seven principles for programs evaluation in curriculum development and innovation. (pp. 117-140). In E. R. House (Ed.), *New directions in educational evaluation*. Lewes, VA: Falmer Press.

Kennedy, M. M. (2010). *Teacher assessment and the quest for teacher quality: A handbook*. San Francisco: Jossey-Bass.

Ladson Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. Schools. *Educational Researcher, 35*(7), 3–12.

Leithwood, K., Lewis, K. S., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning*. Minneapolis, MN: University of Minnesota, Center of Applied Research and Educational Improvement.

Lincoln, I. S., & Guba, E. G. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.

Madaus, G., Scriven, M., & Stufflebeam, D. (1990). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff.

Martínez Slanova, E. (1980). *Principles of evaluation (Los principios de la evaluación)*. Madrid: ICE/UPM. Retrieved from http://www.uhu.es/cine.educacion/didactica/0092principiosevaluacion.htm

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results.* Alexandria, VA: Association for Supervision and Curriculum Development & Aurora: Mid-continent Research for Education Learning.

Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.

McNeil, L. (1986). *Contradictions of control: School structure and school knowledge*. London: Routledge.

McNeil, L. (2000a). *Contradictions of school reform: The educational cost of standardized testing*. London: Routledge.

McNeil, L. (2000b). Creating new inequalities: Contradictions of reform. *Phi Delta Kappan, 81*(10), 729–734.

McNeil, L. M., Coppola, E., Radigan, J., & Heilig, J. V. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives, 16*(3), 1–45.

Munro, J. H. (2008). *Educational leadership*. Boston: McGraw Hill Higher Education.

National Governors' Association Center for Best Practices (NGA Center) & Council of Chief State School Officers (CCSSO). (2012). *Common core state standards initiative*. Retrieved from http://www.corestandards.org/

Nelson, S., & McGhee, M. (2004). *Time-off task: How test preparation is siphoning instructional time for students of colour and students of poverty*. Paper presented at the annual meeting of the University Council for Educational Administration, Kansas City, MO.

Nelson, S., McGhee, M., Meno, L., & Slater, C. L. (2007). Fulfilling the promise of educational accountability: A unique perspective. *Phi Delta Kappan, 88*(9), 702–709.

Nelson, S., McGhee, M., Reardon, R., Gonzales, K., & Kent, C. (2007). *Supplanting teaching with testing: Does it raise test scores?* Paper presented at the annual meeting of the University Council for Educational Administration, Alexandria, VA.

Nevo, D. (1986). The conceptualization of educational evaluation. (pp. 15-30). In E. R. House (Ed.), *New Directions in educational evaluation*. Lewes, DE: Falmer.

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2012). High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives, 20*(20). Retrieved from http://epaa.asu.edu/ojs/article/view/1048

Owen, J. M. (2006). *Program evaluation: Formative approach* (3rd ed.). New York: The Guilford.

PARCC. (2013). *Partnership for assessment of readiness for College and careers*. Retrieved from http://www.parcconline.org/

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park: SAGE Publications, Inc.

Peterson, K. D. (2009). *Teacher evaluation: Comprehensive guide*. Thousand Oaks, CA: Sage.

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher, 40*(3), 103–116.

Quick, P. M., & Normore, A. H. (2004). Moral leadership in the 21st century: Everyone is watching: Especially the students. *The Educational Forum, 68*(4), 336–347.

Radcliffe, J. (2010, March 4). Cheating allegations tighten TAKS test security: State switches exam essay topic, sends monitors to two HISD schools. *Houston Chronicle.*

Reeves, D. (2009). *Leading change in your school*. Alexandria, VA: ASCD.

Romay, M. L., & Crispin, B. M. L. (2000). Self-evaluation and faculty development. In *Faculty evaluation (Evaluacion de la Docencia)* (Chapter 14, pp. 341–364). México: Paidós.

Rossi, P. H., & Freeman, M. (1992). *Evaluation: A systematic approach* (3rd ed.). Thousand Oaks, CA: Sage.

Rueda, R. (2011). *The three dimensions of improving students' performance*. New York: Columbia University Press.

Schemo, D. J., & Fessenden, F. (2003, December 03). A miracle revisited: Measuring success; gains in Houston Schools: How real are they? *New York Times*. http://www.nytimes.com/2003/12/03/us/a-miracle-revisited-measuring-success-gains-in-houston-schools-how-real-are-they.html

Schmoker, M. (1999). *Results: The key to continuous improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Chicago: Rand McNally.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.

Shawn, I., & Greene, J. (2006). *Handbook of evaluation: Policies, programs and practices*. Thousand Oaks, CA: Sage.

Slater, C. L., McGhee, M., Nelson, S., & Meno, L. (2011). Lessons learned: The promise and possibility of educational accountability in the United States. In C. F. Webber & J. Lupart (Eds.), *Leading student assessment* (pp. 41–58). Dordrecht, The Netherlands: Springer.

Smarter Balanced Assessment Consortium (2013) retrieve from http://www.smarterbalanced.org/

Spaulding, D. T. (2008). *Program Evaluation in practice: Core concepts and examples for discussion and analysis*. San Francisco: Jossey-Bass.

Stake, R. E. (1990). Responsive evaluation. In H. G. Walberg & G. D. Haertel (Eds.), *The International encyclopaedia of educational evaluation* (pp. 75–77). Oxford, NY: Pergamon Press.

Stufflebeam, D. L., et al. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.

U.S. Department of Education. (2009a). *Title I — Improving the academic achievement of the disadvantaged*. Retrieved from http://www.ed.gov/policy/elsec/leg/esea02/pg1.htm

U.S. Department of Education. (2009b). *President Obama, U.S. Secretary of Education Duncan announces national competition to advance school reform*. Retrieved from http://www2.ed.gov/news/pressreleases/2009/07/07242009.html

Weiss, C. H. (1992). *Evaluation research: Methods to determine efficiency of programs (Investigación Evaluativa: Métodos para determinar la eficiencia de los programas)*. México: Trillas.

Whitaker, T. (2003). *What great principals do differently?* Larchmont, VA: Eye on Education.

Wholey, J., & Newcomer, K. (1994). *Handbook of practical program evaluation*. San Francisco: Jossey-Bass.

Wildy, H., & Clarke, S. (2008). Charting an arid landscape: The preparation of novice primary principals in Western Australia. *School Leadership and Management, 28*(5), 469–487.

Winerip, M. (2011, July 18). Cracking a system in which test scores were for changing. *New York Times*. http://www.nytimes.com/2011/07/18/education/18oneducation.html?_r=0

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement, 19*(3), 239–259.